applied sciences

MDPI

Article

Machine Learning Regressors to Estimate Continuous Oxygen Uptakes ($\dot{V}O_2$)

Daeeon Hong [1] **and Sukkyu Sun** [2,*]

[1] Department of Interdisciplinary Program in Bioengineering, Seoul National University, 1, Gwanak-ro, Gwanak-gu, Seoul 08826, Republic of Korea; shong65@snu.ac.kr
[2] Department of AI Software Convergence, Dongguk University, 30, Pildong-ro 1-gil, Jung-gu, Seoul 04620, Republic of Korea
* Correspondence: sukkyu.sun@dgu.ac.kr; Tel.: +82-2-2260-3342

Abstract: Oxygen consumption ($\dot{V}O_2$) estimation is vital for evaluating aerobic performance and cardiovascular fitness. This study explores various regression models to develop a real-time $\dot{V}O_2$ and $\dot{V}O_{2max}$ estimation model. Utilizing a dataset from PhysioNet, encompassing cardiorespiratory measurements from 992 treadmill tests conducted at the University of Malaga's Exercise Physiology and Human Performance Lab from 2008 to 2018, participants aged 10 to 63, including amateur and professional athletes, underwent breath-by-breath monitoring of physiological parameters. The study underlines the efficacy of regressor models in handling complex datasets and developing a robust real-time $\dot{V}O_2$ estimation model. After adjusting parameters to $\dot{V}O_2$ in "mL/kg/min" from "mL/min", and selecting 'Age', 'Weight', 'Height', 'HR', 'Sex', and 'Time' as parameters for $\dot{V}O_2$ estimation, XGBoost emerged as the optimal choice. Validation using a test dataset of 132 participants yielded the following results for Mean Absolute Error (MAE), Mean Squared Error (MSE), Root Mean Squared Error (RMSE), R-squared (R^2), Root Mean Squared Logarithmic Error (RMSLE), and Mean Absolute Percentage Error (MAPE) metrics: MAE of 0.1793, MSE of 0.1460, RMSE of 0.3821, R^2 of 0.9991, RMSLE of 0.0140, and MAPE of 0.0066. This study demonstrates the effectiveness of various regressor models in developing a continuous $\dot{V}O_{2max}$ estimation model that has promising performance metrics.

Keywords: machine learning; maximal oxygen consumption ($\dot{V}O_{2max}$); estimation

check for updates

Citation: Hong, D.; Sun, S. Machine Learning Regressors to Estimate Continuous Oxygen Uptakes ($\dot{V}O_2$). *Appl. Sci.* **2024**, *14*, 7888. https://doi.org/10.3390/app14177888

Academic Editor: Keun Ho Ryu

Received: 25 June 2024
Revised: 2 September 2024
Accepted: 3 September 2024
Published: 5 September 2024

1. Introduction

Oxygen consumption ($\dot{V}O_2$) estimation is a vital feature in exercise physiology and human performance assessment that offers significant insights into an individual's physiological response to exercise [1]. While traditional studies have focused on maximal oxygen consumption ($\dot{V}O_{2max}$) to measure aerobic capacity at peak exertion, the potential for wearable device applications in continuous $\dot{V}O_2$ estimation is substantial [2–4]. Continuous VO_2 monitoring, explicit in various studies [5–8], provides real-time data on oxygen consumption, vital for matching total body oxygen consumption and delivery. This capability ensures optimal oxygen utilization, potentially improving overall performance and health outcomes. Additionally, it allows for the assessment of physical activity levels, personalized fitness tracking using heart rate-based algorithms, comparison of the effects of different exercises on oxygen uptake, and facilitation of personalized data-driven interventions for optimized health outcomes.

These devices facilitate continuous monitoring of $\dot{V}O_2$ levels, enabling ongoing adjustments in training intensity, duration, and recovery strategies. They optimize performance outcomes across various athletic endeavors. Significantly, wearable devices extend the utility of $\dot{V}O_2$ estimation beyond exercise physiology to include personalized training optimization, fatigue management, and health risk assessment during physical activities [9,10].

Machine learning strategies are increasingly recognized for their efficiency in $\dot{V}O_2$ estimation [2,11,12]. They allow the investigation of complex relationships between physiological parameters and $\dot{V}O_2$ levels. This offers insights into individual exercise capacities and performance optimization strategies. By leveraging machine learning algorithms, this study aims to enhance the methodology for continuous $\dot{V}O_2$ and $\dot{V}O_{2max}$ estimation for wearable devices, broadening our understanding of its applications in exercise physiology and human performance assessment.

The field has primarily focused on incorporating $\dot{V}O_{2max}$ into machine learning models using various approaches, such as Graph Neural Networks (GNN) [13] and regression models [14]. These studies typically utilize demographic factors like age, weight, and height to predict $\dot{V}O_{2max}$ [12,15]. However, our research takes a novel approach by considering additional physiological parameters, including heart rate, age, sex, weight, height, and categorically mapped $\dot{V}O_2$ level. This comprehensive approach draws inspiration from the works by Drinkwater et al. [16], Jones et al. [17], Hansen et al. [18], and Edvardsen et al. [19], which emphasize the multifactorial nature of $\dot{V}O_{2max}$ estimation. The study integrates these physiological variables into regression models to develop a continuous $\dot{V}O_2$ and $\dot{V}O_{2max}$ estimation model. The studies utilizing PhysioNet [20] for $\dot{V}O_{2max}$ estimation encounter limitations impacting the accuracy and reliability of their findings. This includes challenges in measuring accurate $\dot{V}O_{2max}$ values differences between predicted and actual values. Specific equations and methods, such as the American College of Sports Medicine (ACSM) running equation [21], raise concerns about generalizability and methodological issues [22,23]. Lastly, there are challenges specific to certain populations, such as patients with chronic conditions like chronic obstructive pulmonary disease (COPD) [24]. This approach represents a significant departure from traditional methods and holds promise for additional precise and personalized $\dot{V}O_2$ predictions, with potential applications in sports science, healthcare, and fitness monitoring.

2. Materials and Methods

2.1. Data Characteristics

The open dataset from PhysioNet comprises cardiorespiratory measurements obtained during maximal graded exercise tests (GETs). It offers useful insights into the dynamic physiological responses to exercise. These measurements, encompassing parameters such as heart rate, $\dot{V}O_2$, carbon dioxide production ($\dot{V}CO_2$), respiration rate, and pulmonary ventilation, are crucial for calculating various cardiorespiratory indices utilized in sports science and medicine. Key aspects of oxygen consumption and heart rate dynamics during GETs, including the rate of increase at exercise onset, maximal values, changes at ventilatory thresholds, and dynamics during recovery, are of particular interest.

The measurements were performed between 2008 and 2018, with athletes undergoing maximal GETs on a treadmill connected to a gas analyzer system. The respiratory parameters, including oxygen consumption and pulmonary ventilation, were measured breath-by-breath using a CPX MedGraphics gas analyzer system (Medical Graphics, Saint Paul, MN, USA) connected to a PowerJog J series treadmill (Metagenics Fitness Inc., West Vancouver, BC, Canada), while heart rate was monitored with a Mortara 12-lead ECG device (Milwaukee, WI, USA). The dataset includes participant information, such as age, weight, height, humidity, temperature, and cardiorespiratory measurements, collected during each effort test. In particular, the dataset is characterized by its extensive longitudinal format that contains one line for each breath measurement across 992 effort tests. It facilitates in-depth analyses of cardiorespiratory dynamics during exercise. It enables the development of advanced predictive models and analytical techniques to enhance our understanding of human performance and physiological responses to exercise.

2.2. Data Portion

After transitioning parameters to represent $\dot{V}O_2$ in "mL/kg/min" rather than "ml/min", this adjustment was aligned with the standard rating criteria for $\dot{V}O_2$ levels in wearable devices [25,26]. Additionally, the methodological approach incorporates the utilization of $\dot{V}O_{2max}$, where $\dot{V}O_2$ levels are categorized into 'Poor', 'Fair', 'Good', 'Excellent', and 'Superior' based on age group [25]. This additional categorization parameter enriches the model training process by providing a more subtle representation of performance of $\dot{V}O_2$ levels across different age groups. The age groups below 20 years and 80 or above were set as an exclusion criterion with a given standard rating (n = 154). This decision provides that the analysis focuses on the target population and avoids potential biases or confounding factors associated with extreme age groups or individuals with exceptionally high $\dot{V}O_2$ levels.

Additionally, to address participant overlap and enhance the reliability of the findings, a 5-fold cross-validation approach was deployed. A total of 857 participants were initially considered for the study. After excluding 154 participants due to age-related criteria and removing 37 participants based on Z-score anomaly detection, the final dataset comprised 666 participants. This dataset was then stratified into 80% (n = 532) for model training and 20% (n = 134) for testing. This data partitioning strategy enables a comprehensive evaluation of the model's performance across diverse age demographics while minimizing the risk of overfitting and ensuring generalizability to unseen data.

2.3. Modeling

Various regression models, including Extreme Gradient Boosting (XGBoost), Light Gradient Boosting Machine, Gradient Boosting Regressor, Linear Regression, and AdaBoost Regressor, are trained on the training set. The best-performing model is selected based on the Mean Absolute Percentage Error (MAPE) and further optimized via hyperparameter tuning. The tuned model's performance is evaluated using the test set. The accuracy of the model's predictions is assessed using the R-squared (R^2) metric that is computed to estimate the proportion of variance in the target variable VO_2 (mL/kg/min) explained by the model. An important step in the preprocessing pipeline involves dividing $\dot{V}O_2$ by weight (in kg) to normalize the $\dot{V}O_2$ values to observe standard ratings. This normalization enables fair comparisons across individuals of varying weights and allows the model to capture the relative oxygen consumption per unit of body mass. The features, including age, weight, height, HR, sex, and time, were selected based on their known physiological influence on oxygen consumption and performance in previous $\dot{V}O_2$ estimation studies (Table 1). These variables are critical in capturing the variability in metabolic and cardiorespiratory responses during exercise. By obtaining $\dot{V}O_2$ (mL/kg/min), representing oxygen consumption per kilogram of body weight per minute, the model's predictions are independent of body weight variations, enhancing interpretability and generalizability across diverse populations.

Table 1. Various $\dot{V}O_{2max}$ equations with participant numbers with age group.

	Participants	Age	Equation
Edvardsen et al. (2013) [19]	n = 759 (394 M/365 F)	20–85	Female: $\dot{V}O_{2max}$ (L.min^{-1}) = 3.31 − 0.022 year $\dot{V}O_{2max}$ (mL.kg^{-1}.min^{-1}) = 48.2 − 0.32 year Male: $\dot{V}O_{2max}$ (L.min^{-1}) = 4.97 − 0.033 year $\dot{V}O_{2max}$ (mL.kg^{-1}.min^{-1}) = 60.9 − 0.43 year

Table 1. *Cont.*

	Participants	Age	Equation
Jones et al. (1985) [17]	n = 100 (50 M/50 F)	15–71	Female: $\dot{V}O_{2max}$ (L.min^{-1}) = −0.624 sex + 0.046 height − 0.021 age − 4.31 Male: $\dot{V}O_{2max}$ (L.min^{-1}) = −0.492 sex + 0.032 height − 0.024 age + 0.019 weight − 3.71
Hansen et at. (1984) [18]	n = 77 (77 M)	37–74	Male: $\dot{V}O_{2max}$ (mL.min^{-1}) = weight × (50.75 − 0.372 age)
Drinkwater et al. (1975) [16]	n = 109 (109 F)	10–68	Female: $\dot{V}O_{2max}$ (L.min^{-1}) = 2.46 − 0.016 age $\dot{V}O_{2max}$ (mL.kg^{-1}.min^{-1}) = 83.663 − 4.114 age + 0.127 age^2 − 0.0012 age $\dot{V}O_{2max}$ (mL.kg^{-1}.min^{-1}) = 71.237 − 3.524 age + 0.104 age^2 − 0.0010 age $\dot{V}O_{2max}$ (mL.kgLBM^{-1}.min^{-1}) = 90.684 − 3.808 age + 0.118 age^2 − 0.0011 age $\dot{V}O_{2max}$ (mL.kgLBM^{-1}.min^{-1}) = 88.99 − 4.459 age + 0.140 age^2 − 0.0014 age

Various regression models were compared to assess performance. However, XGBoost was chosen based on the comprehensive comparison of metrics, such as MAE, MSE, RMSE, R^2, RMSLE, MAPE, and training time in seconds [15]. Moreover, the various configurations operated to enrich model performance for the training process. The multicollinearity operated to eliminate highly correlated features. Additionally, 5-fold cross-validation was incorporated to assess the stability and generalization capability of the models (Figure 1). This method ensures that the model is robust and performs well on unseen data, helping to avoid overfitting and ensure that predictions generalize well to the broader population. Further, assigning a session ID ensured reproducibility of experiments, facilitating tracking and comparison of results. The tuned XGBoost model was optimized with a random search with the MAPE. It involves evaluating the model's performance on both training and testing datasets. Yet, predictions on the test data to estimate $\dot{V}O_2$ mL/kg/min values demonstrate the suitability of XGBoost as the most appropriate model for predicting $\dot{V}O_2$ mL/kg/min values. XGBoost has been widely used across different research domains, showcasing its versatility and effectiveness. Notable applications include predicting stock prices in financial markets [27], proactive damage estimation in infrastructure management [28], early earthquake magnitude prediction in earthquake research [29], forecasting air quality in environmental science [30], improving intrusion detection in cybersecurity [31], predicting treatment responsiveness in healthcare [32], and providing accurate forecasts of virus spread during the COVID-19 pandemic for public health management [33].

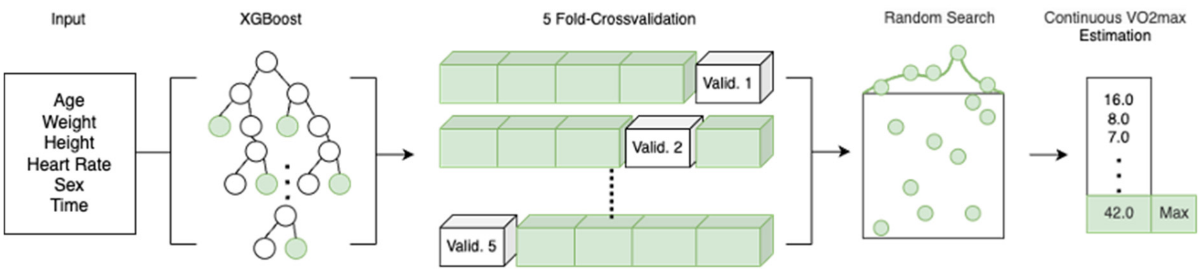

Figure 1. This flowchart shows the input parameters and the modeling approach. It was finalized to approach continuous $\dot{V}O_{2max}$ estimation.

XGBoost is an acclaimed ensemble learning technique utilized for both classification and regression tasks. It leverages multiple decision trees combined, exhibiting outstanding computational efficiency and speed compared to traditional gradient-boosting methods. Notable features include robust performance and speed, facilitated by parallel processing for handling large datasets, automatic tree pruning to prevent overfitting, and strong regularization capabilities. Additionally, XGBoost (version 2.0.3) implements L1 (Lasso) and L2 (Ridge) regularization techniques to manage model complexity, enabling effective prevention of excessive tree depth and reduction of overfitting. Furthermore, it offers pruning functionality to manage model complexity and enhance generalization, along with automatic handling of missing data, distinguishing itself from algorithms requiring manual intervention. Before model training, missing data were addressed using XGBoost's built-in handling techniques, which automatically manage incomplete datasets. Additionally, anomalies detected in the data, such as extreme outliers, were filtered out using Z-score thresholds to ensure model reliability and accuracy. Lastly, the model provides flexibility by allowing custom optimization objectives and evaluation criteria, catering to diverse business needs and research questions. The model hyperparameters include a learning rate of 0.2. It allows the model to adapt its parameters during training. The maximum depth of each tree in the ensemble is set to 7, controlling the level of complexity. The model builds multiple trees to improve predictive accuracy with 220 estimators. The regularization parameters, L1 and L2 regularization, are set to 10 and 4, respectively, to prevent overfitting by penalizing coefficients. Other settings, such as random state, subsample, and verbosity, are also specified to ensure reproducibility, control sampling, and manage output verbosity during training.

The code is available at https://github.com/SeanPresent/VO2max_Estimation, accessed on 25 August 2024.

3. Results

3.1. Baseline Participant Characteristics

Based on the analysis of participant characteristics, our study comprised a total of 532 unique individuals for the training dataset. The gender distribution revealed that the majority of participants were male, accounting for approximately 87.76%, while females comprised the remaining 12.24%. Both genders exhibited similar patterns when examining the age distribution, with males averaging around 32.45 years (±9.08 SD) and females around 30.11 years (±7.74 SD). The participants' height and weight further reflected gender disparities, with males presenting an average height of approximately 176.72 cm (±6.73 SD) and weight of 76.24 kg (±9.95 SD), and females showing averages of 166.11 cm (±8.17 SD) and 61.67 kg (±10.98 SD), respectively. Moreover, the investigation into $\dot{V}O_{2max}$ and min values by gender unveiled significant differences, with males showing higher values compared to females, with $\dot{V}O_{2max}$ at 2421.38 mL/min (±985.61 SD) and $\dot{V}O_{2min}$ at 32.09 mL/kg/min (±13.10 SD) for males, and 1670.74 mL/min (±693.45 SD) and 27.59 mL/kg/min (±10.85 SD) for females, respectively. These findings collectively provide a comprehensive understanding of participant demographics and physiological attributions.

3.2. $\dot{V}O_{2max}$ Parameter Composition

The study conducted feature selection using the correlation matrix for affirmation. The Spearman correlation matrix is a statistical tool used to quantify the strength and direction of relationships between variables in a dataset. It consists of a square matrix where each cell represents the correlation coefficient between two variables. A correlation coefficient close to +1 indicates a strong positive linear relationship, while a coefficient close to −1 indicates a strong negative linear relationship. By examining the correlation matrix, the study can identify which variables correlated with the target variable and select those as features for our model (Figure 1). This process helps to understand the interdependencies between variables and choose the most relevant ones for predicting the target variable accurately.

A GET involves incremental increases in treadmill speed until exhaustion. It offers a practical means to approximate $\dot{V}O_{2max}$. In the study, we explore several established equations presented by Edvardsen et al., Jones et al., Hansen et al., and Drinkwater [16–19] (Table 1). Each is suitable for specific exercise protocols and incorporates model parameters.

The equations incorporate parameters such as 'Age', 'Weight', 'Height', 'HR', 'Sex', and 'Time' deployed to estimate $\dot{V}O_{2max}$ from GET data. Therefore, those parameters are selected to allow machine learning to assess the $\dot{V}O_{2max}$ and continuous $\dot{V}O_2$ estimation.

A weak negative relationship exists between age and $\dot{V}O_2$. It is shown by a correlation coefficient of -0.11. Moreover, weight shows a weak positive correlation with $\dot{V}O_{2max}$ of 0.25. Similarly, height has a weak positive correlation of 0.29 with $\dot{V}O_2$. The heart rate (HR) has a positive correlation with $\dot{V}O_2$ of 0.78. The relationship between sex and $\dot{V}O_2$ shows a weak negative correlation of -0.22. Lastly, the correlation between time and $\dot{V}O_{2max}$ is positive at 0.48.

3.3. Continuous $\dot{V}O_2$ Estimation

In the research paper's Section 3, various regression models were evaluated for their performance in predicting the target variable, $\dot{V}O_2$ mL/kg/min. The table (Table 2) presents a comprehensive comparison of each model's performance metrics, including MAE, MSE, RMSE, R^2, RMSLE, MAPE, and execution time (TT in seconds). Among the models assessed, XGBoost exhibited the most promising results, achieving an MAE of 0.1834, MSE of 0.0640, RMSE of 0.2529, and an impressive R-squared value of 0.9996. Importantly, XGBoost exceeded other models in terms of predictive precision and model fit. Further analysis involved fine-tuning the XGBoost model using 5-fold cross-validation and optimizing for MAPE.

Table 2. Comparison of various machine learning regressors on the training set, and the performance of XGBoost in cross-validation and on the test set.

Training Model (n = 532)	**MAE**	**MSE**	**RMSE**	**R^2**	**RMSLE**	**MAPE**	**TT (s)**
Extreme Gradient Boosting	0.1834	0.0640	0.2529	0.9996	0.0146	0.0077	0.4600
Light Gradient Boosting Machine	0.2419	0.1274	0.3570	0.9992	0.0169	0.0090	1.0520
Gradient Boosting Regressor	0.6158	0.6804	0.8248	0.9959	0.0282	0.0210	3.4760
Linear Regression	0.8622	1.4428	1.2011	0.9913	0.0929	0.0407	5.3260
AdaBoost Regressor	2.2091	7.4254	2.7245	0.9551	0.1206	0.0944	2.4900
Cross Validation	MAE	MSE	RMSE	R^2	RMSLE	MAPE	
1	0.1297	0.0377	0.1942	0.9998	0.0137	0.0058	
2	0.1289	0.0379	0.1947	0.9998	0.0142	0.0057	
3	0.1240	0.0335	0.1829	0.9998	0.0129	0.0055	
4	0.1265	0.0349	0.1869	0.9998	0.0136	0.0056	
5	0.1265	0.0373	0.1933	0.9998	0.0124	0.0055	
Mean	0.1271	0.0363	0.1904	0.9998	0.0134	0.0056	
Standard deviation	0.0020	0.0018	0.0047	0.0000	0.0006	0.0001	
Test set (n = 154)	MAE	MSE	RMSE	R^2	RMSLE	MAPE	
Extreme Gradient Boosting	0.1793	0.1460	0.3821	0.9991	0.0140	0.0066	

This process resulted in enhanced performance metrics, with the tuned XGBoost model demonstrating a mean MAE of 0.1217 and a mean MAPE of 0.0056 on the validation set. These findings underscore the model's robustness and its ability to generalize well to unseen data. Subsequently, the tuned XGBoost model was evaluated on an independent test dataset, where it maintained its MAE of 0.1793, MSE of 0.1460, and an R^2 value of 0.9991, indicating its effectiveness in predicting the $\dot{V}O_2$ for the mL/kg/min variable. Across the board, the results validate the efficacy of the XGBoost model in predicting oxygen consumption levels and highlight its potential utility in practical applications requiring precise estimation of $\dot{V}O_2$ in mL/kg/min.

3.4. $\dot{V}O_{2max}$ Estimation

In this experiment, the model's predictions showed a high correlation with the actual values. Specifically, the R^2 value, which represents the correlation between the predicted $\dot{V}O_2$ mL/kg/min and the actual values, was found to be 0.9683 (Figure 2). This indicates that the model has generalized well to the data, suggesting its capability to accurately predict health indicators based on the given input variables. In estimating the $\dot{V}O_{2max}$, the results revealed an MAE of 0.7109, MSE of 2.5859, and RMSE of 1.6081. Additionally, the RMSLE was calculated to be 0.0277, and the MAPE was 1.2971. The R^2 value of the model stood at 0.9683, indicating a high level of explanatory power in predicting $\dot{V}O_{2max}$ (Figure 3). Finally, a comparison of $\dot{V}O_2$ with continuous $\dot{V}O_2$ estimation resulted in Figure 4. To provide a comprehensive evaluation of the model's performance, we selected three participants based on the accuracy of their $\dot{V}O_2$ predictions: the participant with the smallest prediction error ("best"), the participant with the largest prediction error ("worst"), and the participant whose prediction error was closest to the median ("other"). The absolute prediction error for each participant was calculated by taking the mean of the absolute differences between the measured $\dot{V}O_2$ and the predicted $\dot{V}O_2$ values. The "best" participant was identified as the one with the lowest mean absolute error, while the "worst" participant had the highest. The "other" participant was chosen by finding the individual whose error was closest to the median error across all participants.

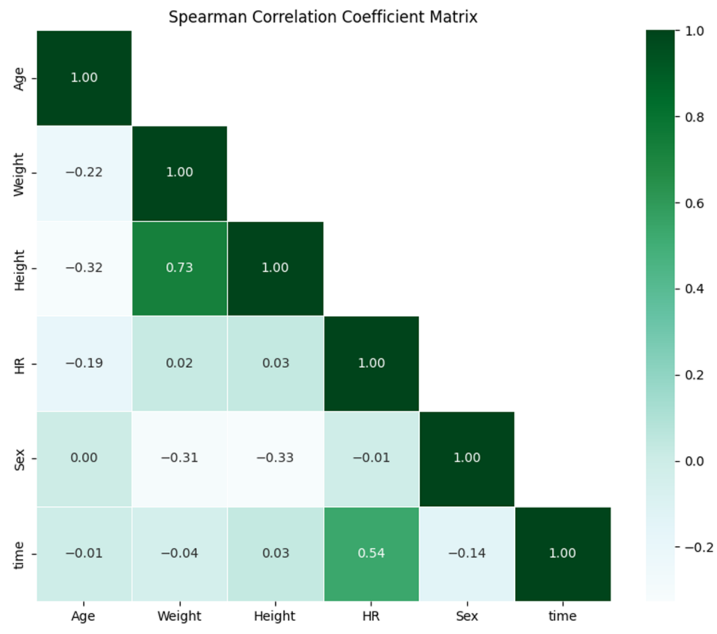

Figure 2. The image represents a Spearman correlation matrix for various fitness indicators of age, weight, height, heart rate (HR), sex, $\dot{V}O_2$, and a categorized measure of $\dot{V}O_2$. The Spearman correlation coefficient measures the potency and direction of the linear relationship between two variables. The coefficient values range between −1 and 1, where 1 indicates a perfect positive linear relationship, −1 indicates a perfect negative linear relationship, and values near 0 indicate no linear relationship.

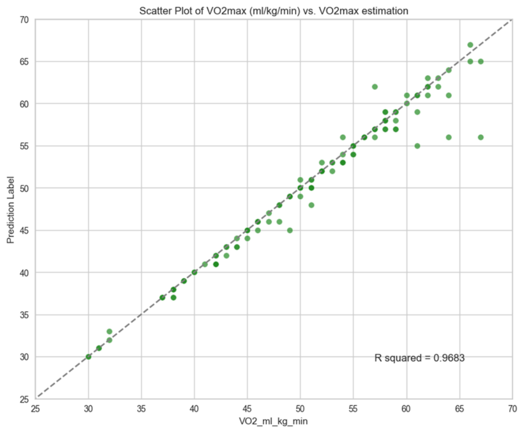

Figure 3. This is indicating $\dot{V}O_{2max}$ evaluated with R^2. The scatter plot illustrates the relationship between the predicted $\dot{V}O_{2max}$ (mL/kg/min) and the actual $\dot{V}O_{2max}$ values. The R^2 value of 0.9683 further quantifies the model's explanatory power, showing a correlation between the predicted and actual $\dot{V}O_{2max}$ values.

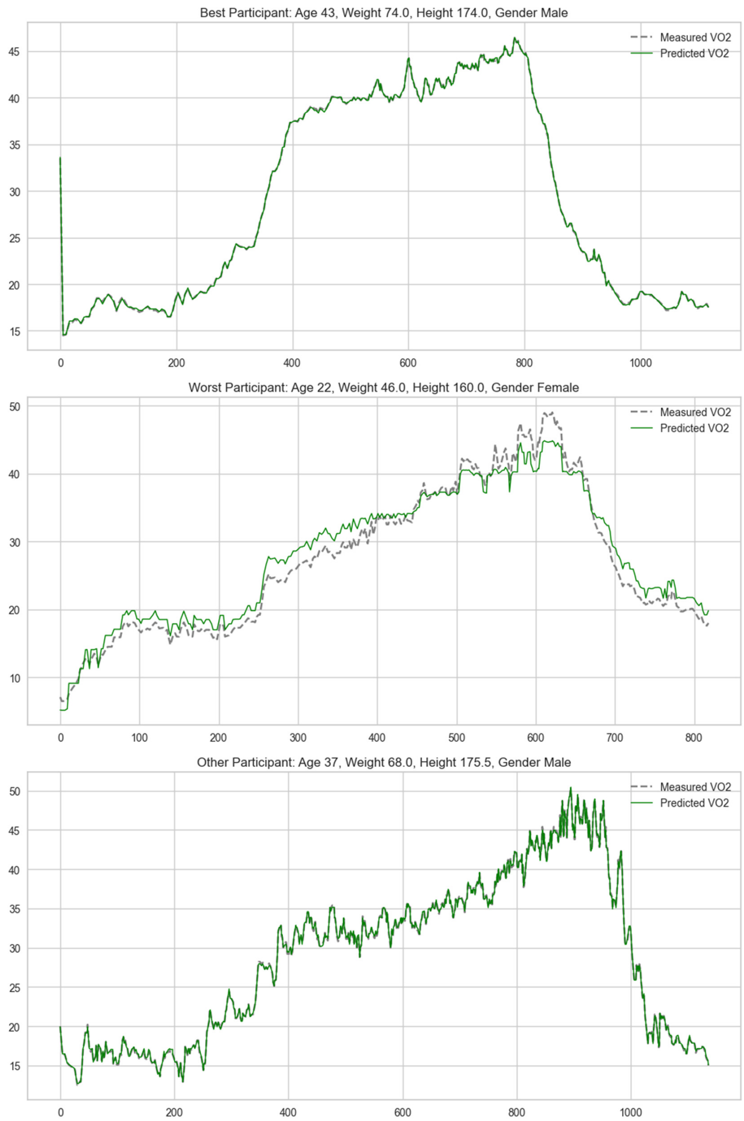

Figure 4. Comparison of $\dot{V}O_2$ with continuous $\dot{V}O_2$ estimation on the top figure, and $\dot{V}O_{2max}$ estimation comparison by the best, worst, and other on the bottom figure.

4. Discussion

In our study, we initially examined the baseline characteristics of participants, revealing a cohort of 532 unique individuals, predominantly male at 87.76%, with females comprising 12.24%. Analysis of the age distribution showed similar patterns across genders, with males averaging 32.72 years (±8.86 SD) and females 29.94 years (±7.44 SD). Disparities in height and weight were evident, with males exhibiting higher averages of 176.62 cm (±6.53 SD) and 76.40 kg (±10.07 SD) compared to females at 165.88 cm (±7.55 SD) and 60.81 kg (±10.02 SD), respectively. Moreover, $\dot{V}O_{2max}$ and min values showcased significant gender differences, with males displaying higher values: $\dot{V}O_{2max}$ at 2421.38 mL/min (±985.61 SD) and $\dot{V}O_2$ min at 31.94 mL/kg/min (±13.05 SD), compared to females at 1670.74 mL/min (±693.45 SD) and 27.59 mL/kg/min (±10.85 SD), respectively.

Moving on to the continuous estimation of $\dot{V}O_2$, various regression models were evaluated, among which XGBoost demonstrated superior performance, with an MAE of 0.1834, MSE of 0.0640, RMSE of 0.2529, and a significant R-squared value of 0.9996. Fine-tuning of the XGBoost model further improved its performance, achieving a mean MAE of 0.1217 (±0.0020) and a mean MAPE of 0.0056 (±0.0001) on validation. The model maintained its efficacy on an independent test dataset, retaining an MAE of 0.1793, MSE of 0.1460, and an R-squared value of 0.9991. These results underscore the robustness of the XGBoost model in accurately predicting $\dot{V}O_2$ levels. Furthermore, the estimation of $\dot{V}O_{2max}$ yielded promising results, with a high correlation reflected in an R^2 value of 0.9691. The model exhibited an MAE of 0.7109, MSE of 2.5859, RMSE of 1.6081, RMSLE of 0.0277, and MAPE of 1.2971, indicating its effectiveness in predicting $\dot{V}O_{2max}$. Overall, these findings emphasize the model's potential utility in estimating physiological parameters, with implications for health monitoring and fitness assessment.

Our study also provides an insightful comparison between the machine learning-based approach and traditional $\dot{V}O_{2max}$ estimation equations, such as those presented by Edvardsen et al., Jones et al., Hansen et al., and Drinkwater [16–19] (Table 1). Unlike the conventional equations, which rely heavily on demographic factors, like age, weight, and height, our machine learning models incorporate real-time data, allowing for a more accurate and individualized prediction of $\dot{V}O_{2max}$. The improvements in the MAE, RMSE, and other performance metrics highlight the enhanced predictive power of AI algorithms, particularly when fine-tuned through methods like hyperparameter optimization in XGBoost. These advancements are critical for applications requiring continuous monitoring and real-time feedback, which are incomprehensible with static equations.

The Spearman correlation coefficient matrix provides insights into $\dot{V}O_2$ and its relationship with other parameters. Although we followed the equations to select parameters, height and weight showed a relatively high correlation, along with a smaller positive correlation with age. This suggests that taller and heavier individuals tend to have higher $\dot{V}O_{2max}$ values. The scatter plot displaying the relationship between actual $\dot{V}O_{2max}$ (mL/kg/min) and its estimation demonstrates strong predictive accuracy, as indicated by a high R-squared value of 0.9691. This strong linear relationship confirms the reliability of the continuous $\dot{V}O_2$ estimation methods used in this study.

Compared to the study from Rosol, M. et al. [15], our study differs in model selection, input variables, and target population. We use machine learning models like XGBoost and LightGBM for real-time continuous $\dot{V}O_2$ estimation, relying on variables such as age, weight, height, heart rate, sex, and time—readily available from wearables. In contrast, the study focuses on submaximal exercise stages, which may limit real-time application. Our dataset, primarily composed of athletes from PhysioNet, offers a more homogeneous sample but limits generalizability to broader populations. The previous study likely includes a more diverse population, potentially increasing variability but broadening the applicability of its findings. However, we use 5-fold cross-validation to ensure robustness, while the Rosol, M. et al. study may follow a different validation approach. Both studies use metrics like MAE and RMSE, allowing for direct comparison. However, our focus on minimizing real-time prediction errors better suits continuous monitoring applications. Hyperparameter tuning played a critical role in our model's optimization, particularly with XGBoost, using random search to refine key parameters. The Rosol, M. et al. study may adopt a different strategy for model optimization, highlighting varying approaches to improving performance. In terms of practical application, our study is designed for real-time $\dot{V}O_2$ estimation using wearable devices, whereas the study focuses on retrospective analysis. Both studies acknowledge limitations; our study highlights dataset imbalance and suggests more diverse samples for future research, while the Rosol, M. et al. study likely addresses the limitations of submaximal stages and population diversity.

The first significant limitation of this study is that the model relies primarily on data from athletes, which may not represent the general population or those with sedentary lifestyles. Additionally, the lack of data from individuals with chronic health conditions, such as cardiovascular diseases or COPD, restricts its applicability. Furthermore, the model's accuracy depends on the precision and consistency of data from wearable devices, which can be affected by sensor performance and user adherence, potentially impacting real-world reliability. By employing adaptive learning systems, wearables can be generalized across diverse populations, including non-athletic individuals and those with chronic conditions, addressing the generalizability and accuracy issues noted in traditional $\dot{V}O_{2max}$ estimation methods.

Second, while our study provides valuable insights into $\dot{V}O_2$ estimation, it is important to recognize the influence of gender imbalance in our sample. Specifically, 87.76% of the participants were male, and only 12.24% were female. This imbalance may limit the generalizability of our findings, particularly concerning female populations, as differences in physiological responses between genders could affect $\dot{V}O_2$ estimation.

The integration of such technologies not only addresses the specific limitations underlined by PhysioNet but also highlights a significant opportunity to revolutionize exercise monitoring and optimization. This approach ensures that individuals receive the most beneficial and safe exercise prescriptions that are tailored to their unique physiological profiles. By enhancing the capability of wearables to provide real-time feedback and personalized training regimens, this research contributes to the growing field of personalized medicine and digital health, demonstrating the extensive potential of wearable technology in improving health management and fitness strategies across various populations. In comparison to other studies, our findings underscore the significance of continuous $\dot{V}O_2$ estimation in enhancing personalized training and health management. While traditional methods have not entailed providing real-time feedback, wearable devices offer a promising solution by continuously monitoring $\dot{V}O_2$ levels during physical activity. This capability not only facilitates more accurate performance assessment but also enables training regimens based on individual responses. Therefore, our research contributes to the growing body of evidence supporting the integration of wearable technology in fitness and healthcare domains, highlighting its potential to revolutionize exercise monitoring and optimization.

Author Contributions: Conceptualization, D.H. and S.S.; methodology, S.S.; software, D.H.; validation, S.S.; formal analysis, D.H.; investigation, S.S.; resources, D.H.; data curation, D.H.; writing—original draft preparation, D.H.; writing—review and editing, D.H. and S.S.; visualization, D.H.; supervision, S.S.; project administration, S.S. All authors have read and agreed to the published version of the manuscript.

Funding: This work was supported by the Institute of Information & Communications Technology Planning & Evaluation (IITP) under an Artificial Intelligence Convergence Innovation Human Resources Development (IITP-2024-RS-2023-00254592) grant funded by the Korea government (MSIT).

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The data are available at the following web-site address: https://physionet.org/content/treadmill-exercise-cardioresp/1.0.1/#files-panel, accessed on 25 August 2024.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Davis, J.M.; Carlstedt, C.J.; Chen, S.C.; Carmichael, M.D.; Murphy, E.A. The Dietary Flavonoid Quercetin Increases VO_2max and Endurance Capacity. *Int. J. Sport. Nutr. Exerc. Metab.* **2010**, *20*, 56–62. [CrossRef]
2. Abut, F.; Akay, M.F. Machine Learning and Statistical Methods for the Prediction of Maximal Oxygen Uptake: Recent Advances. *Med. Devices Evid. Res.* **2015**, *8*, 369–379. [CrossRef]
3. Jalili, M.; Nazem, F.; Sazvar, A.; Ranjbar, K. Prediction of Maximal Oxygen Uptake by Six-Minute Walk Test and Body Mass Index in Healthy Boys. *J. Pediatr.* **2018**, *200*, 155–159. [CrossRef]
4. Abut, F.; Akay, M.F.; George, J.D. Developing New VO_2max Prediction Models from Maximal, Submaximal and Questionnaire Variables Using Support Vector Machines Combined with Feature Selection. *Comput. Biol. Med.* **2016**, *79*, 182–192. [CrossRef]
5. Squara, P. Matching Total Body Oxygen Consumption and Delivery: A Crucial Objective? *Intensiv. Care Med.* **2004**, *30*, 2170–2179. [CrossRef]
6. Manzoni, C.; Carrard, A.; Fontana, E.; Lemay, M.; Bertschi, M.; Delgado-Gonzalo, R. Towards VO_2 Monitoring: Validation of a Heart Rate Based Algorithm. In Proceedings of the 2017 39th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC), Jeju, Republic of Korea, 11–15 July 2017; Volume 2017.
7. Aouf, M.K.; Seoudy, T.M.; El-Hawsh, G.M. Certain classes of analytic functions defined by integral operators. *Electron. J. Math. Anal. Appl.* **2015**, *3*, 245–249. [CrossRef]
8. Aleppo, G.; Webb, K.M. Continuous Glucose Monitoring Integration in Clinical Practice: A Stepped Guide to Data Review and Interpretation. *Diabetes Technol. Soc.* **2019**, *13*, 664–673. [CrossRef]
9. Vijayan, V.; Connolly, J.; Condell, J.; McKelvey, N.; Gardiner, P. Review of Wearable Devices and Data Collection Considerations for Connected Health. *Sensors* **2021**, *21*, 5589. [CrossRef] [PubMed]
10. Carrier, B.; Helm, M.M.; Cruz, K.; Barrios, B.; Navalta, J.W. Validation of Aerobic Capacity (VO_2max) and Lactate Threshold in Wearable Technology for Athletic Populations. *Technologies* **2023**, *11*, 71. [CrossRef]
11. Ashfaq, A.; Cronin, N.J.; Müller, P. Recent Advances in Machine Learning for Maximal Oxygen Uptake (VO_2Max) Prediction: A Review. *Inform. Med. Unlocked* **2022**, *28*, 100863. [CrossRef]
12. Alzamer, H.; AbuHmed, T.; Hamad, K. A Short Review on the Machine Learning-Guided Oxygen Uptake Prediction for Sport Science Applications. *Electronics* **2021**, *10*, 1956. [CrossRef]
13. Henriques, J.; Carvalho, P.; Rocha, T.; Paredes, S.; Cabiddu, R.; Trimer, R.; Mendes, R.G.; Borghi-Silva, A.; Kaminsky, L.A.; Ashley, E.A.; et al. A Non-Exercise Based V02max Prediction Using FRIEND Dataset with a Neural Network. In Proceedings of the 2017 39th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC), Jeju, Republic of Korea, 11–15 July 2017; Volume 2017.
14. Balakarthikeyan, V.; Jais, R.; Vijayarangan, S.; Premkumar, P.S.; Sivaprakasam, M. Heart Rate Variability Based Estimation of Maximal Oxygen Uptake in Athletes Using Supervised Regression Models. *Sensors* **2023**, *23*, 3251. [CrossRef] [PubMed]
15. Rosoł, M.; Petelczyc, M.; Gąsior, J.S.; Młyńczak, M. Prediction of Peak Oxygen Consumption Using Cardiorespiratory Parameters from Warmup and Submaximal Stage of Treadmill Cardiopulmonary Exercise Test. *PLoS ONE* **2024**, *19*, e0291706. [CrossRef] [PubMed]
16. Drinkwater, B.L.; Horvath, S.M.; Wells, C.L. Aerobic Power of Females, Ages 10 to 68. *J. Gerontol.* **1975**, *30*, 385–394. [CrossRef]
17. Jones, N.L.; Makrides, L.; Hitchcock, C.L.; Chypchar, T.; McCartney, N. Normal Standards for an Incremental Progressive Cycle Ergometer Test. *Am. Rev. Respir. Dis.* **1985**, *131*, 700–708.
18. Hansen, J.E.; Sue, D.Y.; Wasserman, K. Predicted Values for Clinical Exercise Testing. *Am. Rev. Respir. Dis.* **1984**, *129*, S49–S55. [CrossRef]
19. Edvardsen, E.; Edvardsen, E.; Hansen, B.H.; Holme, I.; Dyrstad, S.M.; Anderssen, S.A. Reference Values for Cardiorespiratory Response and Fitness on the Treadmill in a 20- to 85-Year-Old Population. *Chest* **2013**, *144*, 241–248. [CrossRef]
20. Mongin, D.; García Romero, J.; Alvero Cruz, J.R. Treadmill Maximal Exercise Tests from the Exercise Physiology and Human Performance Lab of the University of Malaga. *Age* **2021**, *27*, 36–32. [CrossRef]
21. Glass, S.; Dwyer, G.B. *ACSM'S Metabolic Calculations Handbook*; Lippincott Williams & Wilkins: Philadelphia, PA, USA, 2007.
22. Koutlianos, N.; Dimitros, E.; Metaxas, T.; Cansiz, M.; Deligiannis, A.; Kouidi, E. Indirect Estimation of VO_2max in Athletes by ACSM's Equation: Valid or Not? *Hippokratia* **2013**, *17*, 136.
23. Buttar, K.K.; Saboo, N.; Kacker, S. Measured and Predicted Maximal Oxygen Consumption (VO_2max) in Healthy Young Adults: A Cross-Sectional Study. *J. Health Sci. Med. Res.* **2023**, *41*, 2022896. [CrossRef]
24. Fregonezi, G.; Resqueti, V.; Vigil, L.; Calaf, N.; Casan, P. Maximal Oxygen Uptake Cannot Be Estimated from Resting Lung Function and Submaximal Exercise in Patients with Chronic Obstructive Pulmonary Disease. *J. Cardiopulm. Rehabil. Prev.* **2012**, *32*, 219–225. [CrossRef] [PubMed]

25. VENU^{TM} Owner's Manual. Available online: https://www8.garmin.com/manuals/webhelp/venu/EN-US/Venu_OM_EN-US.pdf (accessed on 10 May 2024).
26. Track Your Cardio Fitness Levels. Available online: https://support.apple.com/en-us/108790 (accessed on 26 May 2024).
27. Yu, S.; Tian, L.; Liu, Y.; Guo, Y. LSTM-XGBoost Application of the Model to the Prediction of Stock Price. In *Artificial Intelligence and Security, Proceedings of the 7th International Conference, ICAIS 2021, Dublin, Ireland, 19–23 July 2021*; Sun, X., Zhang, X., Xia, Z., Bertino, E., Eds.; Lecture Notes in Computer Science; Springer: Cham, Switzerland, 2021; Volume 12736. [CrossRef]
28. Lim, S.; Chi, S. Xgboost Application on Bridge Management Systems for Proactive Damage Estimation. *Adv. Eng. Inform.* **2019**, *41*, 100922. [CrossRef]
29. Joshi, A.; Vishnu, C.; Mohan, C.K.; Raman, B. Application of XGBoost Model for Early Prediction of Earthquake Magnitude from Waveform Data. *J. Earth Syst. Sci.* **2023**, *133*, 1–18. [CrossRef]
30. Lei, T.M.T.; Ng, S.C.W.; Siu, S.W.I. Application of ANN, XGBoost, and Other ML Methods to Forecast Air Quality in Macau. *Sustainability* **2023**, *15*, 5341. [CrossRef]
31. Zivkovic, M.; Tair, M.; Venkatachalam, K.; Bacanin, N.; Hubalovsky, S.; Trojovský, P. Novel Hybrid Firefly Algorithm: An Application to Enhance XGBoost Tuning for Intrusion Detection Classification. *PeerJ Comput. Sci.* **2022**, *8*, e956. [CrossRef]
32. Park, J.Y.; Kim, J.H.; Ryu, D.-H.; Choi, H.Y. Factors Related to Steroid Treatment Responsiveness in Thyroid Eye Disease Patients and Application of SHAP for Feature Analysis with XGBoost. *Front. Endocrinol.* **2023**, *14*, 1079628. [CrossRef]
33. Fang, Z.; Yang, S.; Lv, C.; An, S.; Wu, W. Application of a Data-Driven XGBoost Model for the Prediction of COVID-19 in the USA: A Time-Series Study. *BMJ Open* **2022**, *12*, e056685. [CrossRef]