npj | microgravity

Article

Published in cooperation with the Biodesign Institute at Arizona State University, with the support of NASA

https://doi.org/10.1038/s41526-025-00542-4

AI-based prediction of VO_2 max from 24-h Holter ECG recording

Sarah Solbiati[1,5], Maria Chiara Fiorentino[2,5], Riccardo Bendandi[1], Sara Moccia[3,6] & Enrico Gianluca Caiani[1,4,6] ✉

Cardiovascular (CV) deconditioning is a consequence of spaceflight, characterized by functional and structural changes in the heart and blood vessels due to prolonged exposure to microgravity. These adaptations lead to a noticeable reduction in exercise capacity, particularly evident in alterations of maximum aerobic power (VO_2 max [mL/kg/min]). This work explored the hypothesis that microgravity-induced CV deconditioning could be anticipated through information derived from a non-exercise longitudinal 24-h Holter ECG recording, thereby offering a means for its regular monitoring in space. Pooled data acquired at baseline and at the last day of six Head-Down Tilt bed rest campaigns of different duration were utilized to test this hypothesis. Machine Learning models were developed using multiple features to predict VO_2 max after deconditioning. Obtained results showed good prediction accuracy (RMSE < 5 mL/kg/min) demonstrating the potential of this approach to identify VO_2 max deterioration during bed rest, with possible applications in real world space environments and clinical settings.

Cardiovascular (CV) deconditioning is a well-recognized consequence of microgravity exposure during spaceflight. It is characterized by altered CV dynamics, remodeling, and autonomic nervous system adaptations[1,2], increasing the risk of adverse CV outcomes during and after spaceflight and planetary operations. It manifests as reduced exercise capacity and orthostatic tolerance[3,4], potentially impairing performance in demanding tasks such as extravehicular activities or emergency situations[5]. This also poses additional risks during partial gravity restoration after a long-duration space travel, such as planetary landing, where great gravitational forces could be reached[3]. Predicting the onset and severity of CV deconditioning is crucial for evaluating mission feasibility[6] and defining risk thresholds for physically demanding tasks to ensure astronaut safety and mission success[3]. However, the large inter-individual variability in CV and musculoskeletal response to spaceflight[7] highlights the need for personalized and predictive risk assessment methods.

A key indicator of aerobic capacity and CV health is maximal oxygen consumption (VO_2 max), representing the highest rate of oxygen uptake during maximal exertion[8]. VO_2 max is typically measured through an incremental exercise on a treadmill or a cycle ergometer, with the intensity gradually increasing until the participant reaches the maximum effort or exhaustion[9]. The maintenance of aerobic fitness during spaceflight depends on both pre-flight and in-flight factors, including nutrition and exercise countermeasures[10,11]. Observations from 9- to 17-day Shuttle and Spacelab showed VO2 max reductions from 10.4% to 22%[12,13]. Interestingly, during longer missions (168.6 ± 19.2 days), the decline appeared non-linear with mission duration[14], reaching a plateau within several week[15].

Conducting a VO_2 max assessment during spaceflight is costly and time consuming. The current in-flight Periodic Fitness Evaluation (PFE) on the International Space Station uses a submaximal four-stage (5 min each) cycling test at 25%, 50%, 75%, and 25% of pre-flight VO_2 max, performed on NASA's Cycle Ergometer with Vibration Isolation System, typically on flight day 15 and every 30 days thereafter[16,17]. VO_2 max changes are then estimated, rather than directly measured, as linear extrapolation of heart rate-oxygen consumption relationships, with limitations in its individual accuracy[18].

Recently, various machine learning (ML) models have been developed to predict VO_2 max from exercise, non-exercise, or hybrid tests[19,20]. Exercise-based models, whether maximal or submaximal, are accurate but time- and resource-demanding[21], while non-exercise models rely on questionnaires assessing perceived fitness. ML algorithms such as Multiple Linear Regression (MLR), Support Vector Machines (SVM), Multi-Layer Perceptron (MLP), and Artificial Neural Networks (ANN) have been applied[20], and wearable-sensor approaches have also been explored using ECG, seismocardiogram, or blood pressure data[22].

[1]Politecnico di Milano, Department of Electronics, Information and Bioengineering, Milan, Italy. [2]Information Engineering, Università Politecnica delle Marche, Ancona, Italy. [3]Department of Innovative Technologies in Medicine and Dentistry, Università degli Studi "G. d'Annunzio" Chieti-Pescara, Chieti, Italy. [4]IRCCS Istituto Auxologico Italiano, Milan, Italy. [5]These authors contributed equally: Sarah Solbiati, Maria Chiara Fiorentino.[6]These authors jointly supervised this work: Sara Moccia, Enrico Gianluca Caiani. ✉e-mail: enrico.caiani@polimi.it

Table 1 | Performance metrics obtaine×d in the tested different regression models

Regressor	SVR	KNR	DTR	RFR	ABR
MAE (mL/kg/min)	5.34 [4.63; 5.90]	5.63 [3.90; 5.94]	5.97 [5.41; 6.65]	4.58 [3.97; 4.79][a,b]	4.15 [3.43; 4.51][a,b,c]
MAPE (%)	14.62 [12.71; 15.73]	16.11 [10.79; 17.10]	16.19 [15.37; 17.33]	12.47 [10.86; 14.52][a,b]	11.33 [9.68; 13.39][a,b]
SMAPE(%)	7.08 [5.97; 7.89]	7.58 [5.20; 7.82]	8.11 [7.59; 8.91]	6.08 [5.31; 6.90][a,b]	5.44 [4.66; 6.37][a,b]
RMSE(mL/kg/min)	6.71 [6.08; 7.98]	6.88 [5.42; 7.42]	8.05 [6.77; 8.62]	5.72 [4.70; 5.99][a,b,c]	4.88 [3.97; 5.58][a,b,c]

MAE mean absolute error, *MAPE* mean absolute percentage error, *SMAPE* symmetric mean absolute percentage error, *RMSE* root mean squared error, *SVR* support vector regressor, *KNR* K-nearest Neighbor Regressor, *DTR* decision trees regressor, *RFR* random forest regressor, *ABR* adaboost regressor.
[a]vs KNR.
[b]vs DTR.
[c]vs SVR (Friedman, $p < 0.05$, post-hoc Wilcoxon signed-rank test with Hochberg correction).

However, few studies have estimated VO_2 max from resting ECG-derived heart rate variability (HRV) features[23], which could offer a simpler, exercise-free, and cost-effective alternative, and in this perspective Holter 24-h ECG recordings could represent a valuable tool.

Building on our previous evidence linking microgravity-induced deconditioning to cardiac electrical activity changes[24–26], this study hypothesized that features derived from 24-h ECG recordings could predict microgravity-induced VO_2 max changes compared to baseline. Our aim was therefore to develop and compare ML models to predict the VO_2 max from non-exercise 24-h ECG recordings obtained in the context of head-down (-6°) tilt (HDT) bed rest studies, a validated ground-based analog of microgravity inducing cardiac deconditioning. Specifically, the VO_2 max prediction after HDT was based on BMI, age, and ECG-derived features measured at baseline (i.e., before HDT started) and at the last day of HDT, together with the VO_2 max measured at baseline.

Results

Predictive models comparison

Results are reported as median (25^{th} percentile; 75^{th} percentile). Among the different tested pipelines, the AdaBoost Regressor (ABR) and Random Forest Regressor (RFR) showed lower VO_2 max prediction errors, with ABR providing the lowest Mean Absolute Percentage Error (MAPE), Symmetric MAPE (SMAPE), Mean Absolute Error (MAE) and Root Mean Squared Error (RMSE) (Table 1). Accordingly, ABR was therefore selected for further analyses. In particular, it resulted in a SEE of 5.24 [4.40; 6.11] mL/kg/min (SEE adjusted[23] = 8.24 [6.85; 9.82] mL/kg/ min), and an r^2 of 0.57 [0.51; 0.65]. Interestingly, as reported in Table 2, prediction performance progressively worsened with increasing HDT duration (from 5 to 60 days) across all metrics, regardless of the campaign facility.

Feature importance and classification performance

Feature importance analysis on the ABR results (Fig. 1) indicated that baseline VO_2 max value was the most influential predictor along with the delta features relevant to the QT interval and T wave morphology, representing changes in parameters between post- and pre-HDT recordings. The proposed a-posteriori classification (Table 3) of the subjects based on VO_2 max classes, as defined in ref. 27, also achieved good performance: binary classification achieved 78% accuracy, >96% recall, and >77% precision, while three-class classification reached 71% accuracy, >75% recall, and >60% precision.

Discussion

This study demonstrates the possibility to successfully predict VO_2 max, after cardiovascular deconditioning induced by HDT bed rest, using ML models trained with features derived from non-exercise longitudinal (at baseline and at the last day of HDT) 24-hour Holter ECG recordings, BMI, age and baseline VO_2 max, holding potential application of this approach in both real world and space environments. Specifically, the model is fed with multiple ECG-based features including trends, nonlinear and geometrical indices, frequency- and time-domain HR and QT variability, and circadian parameters, and achieves VO_2 max prediction accuracy comparable to exercise-based methods. To our knowledge, this is the first application of a non-exercise approach potentially suitable for monitoring aerobic fitness in space. A standardized dataset we collected over 15 years from six ESA HDT campaigns of varying duration and countermeasures was used for its development, thus capturing different degrees of deconditioning.

Table 2 | Evaluation metrics of the AdaBoost Regressor on VO_2 max predictions grouped by HDT campaign

Metrics	DLR 5-day	MEDES 5-day	DLR 21-day	MEDES 21-day	DLR 60-day	MEDES 60-day
MAE (mL/kg/min)	3.45	2.90	3.97	4.16	5.62	4.62
MAPE (%)	7.38	7.38	9.02	13.28	19.06	16.80
SMAPE (%)	7.79	7.51	9.09	12.57	16.32	15.02
RMSE (mL/kg/min)	4.23	3.34	5.26	4.98	7.57	5.37

MAE mean absolute error, *MAPE* mean absolute percentage error, *SMAPE* symmetric mean absolute percentage error, *RMSE* root mean squared error.

Among the tested ML techniques, ABR achieved the best overall performance, with a median RMSE of 4.87 mL/kg/min, further reduced when considering the 5-day HDT campaigns, representing approximately 40% of the total study population. These results confirm that an ensemble approach such as ABR, based on combining multiple "weak" learners, could effectively result in an improved prediction accuracy[28]. These results are comparable to ML models that include exercise data. For example, combining maximal and submaximal treadmill data (HR, grade, speed) from 100 healthy subjects with two questionnaires (Q-PFA and Physical Activity Readiness), it yielded an RMSE of 2.92 mL/kg/min using SVM[29]. Similarly, in a study[30] conducted on 98 subjects, adding treadmill parameters to demographic and physiological variables improved RMSE from 8.3 to 4.51 mL/kg/min using Generalized Regression Neural Networks. Materko et al.[23]. predicted VO_2 max using HRV features from 10-minute supine resting ECG in 70 young, physically active male students (age 19–29), achieving a SEE_{adj} of 4.40 mL/kg/min (i.e., approximately half compared to our model) with 76% explained variance, hence within a very narrow VO_2 max range (39.5–43.7 ml/kg/min, 95% confidence interval, all subjects in Class 3 as defined in ref. 27), thus limiting its generalizability. In contrast, our study considered a larger (148 male subjects) and more heterogeneous cohort (age range 20–45 years old) covering the full VO_2 max spectrum (all Classes from 1 to 7 as proposed in ref. 27, see Fig. 2 and Supplementary Table 1). This population exhibited high inter- and intra-subject variability, altered circadian rhythms, and progressive CV adaptations, inherently increasing prediction uncertainty. Furthermore, the SEE_{adj} has been defined as a less suitable metric than SEE[31], and the SEE values we obtained (5.24 mL/kg/min) remain comparable to both exercise-based (2.8–7.10 mL/kg/min)[31–33] and non-exercise (3.36 mL/kg/min) models[34]. As additional comparison, consumer wearables relying on resting HRV features show higher errors (up to ±30.5 mL/kg/min[35]) in estimating VO_2 max, confirming the robustness of our approach.

Binary classification confirmed the expected shift in the number of subjects from normal to sub-normal VO_2 max after bed rest, also reflected in

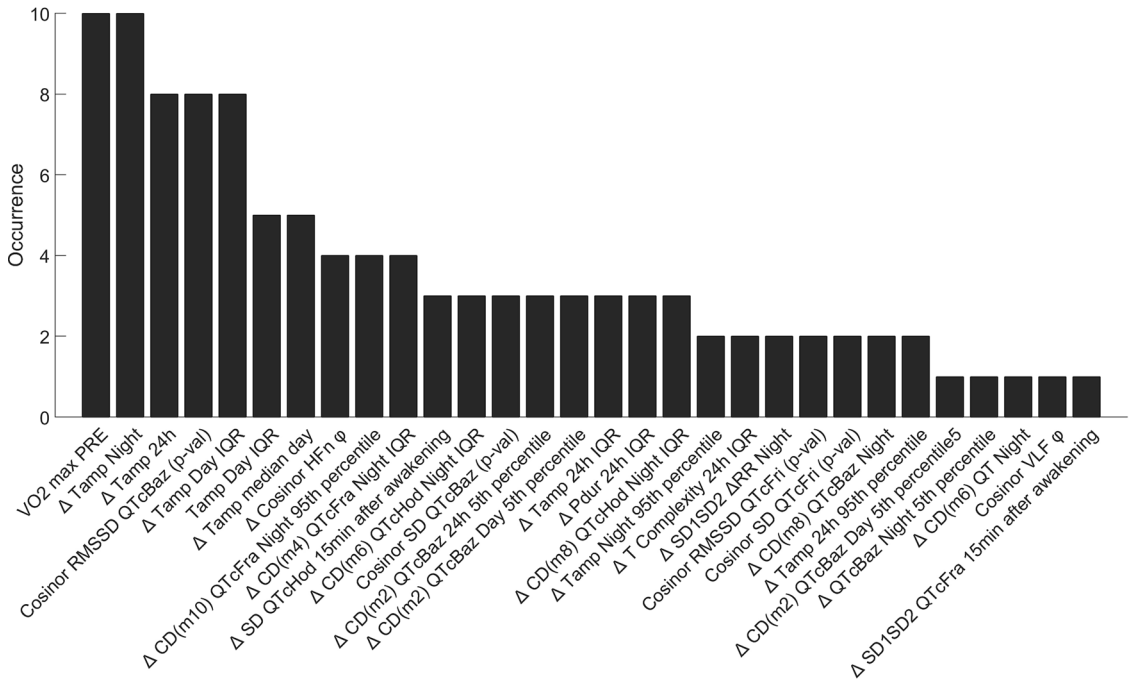

Fig. 1 | Features importance. Histogram displaying the frequency of the most common features extracted from the ECG signals for VO_2 max prediction in the AdaBoost Regressor. The features are ranked by their occurrence in the feature selection process, with the value of baseline VO_2 max, and the T amplitude at night and in the 24-h, in leading positions.

Table 3 | Average out-of-sample performances across the 10 test folds of a-posteriori subjects' classifications using ABR

Metric	Classification Method	Classes (mL/kg/min)	Result
Accuracy	Binary	Classes A1, B1	0.78 [0.74; 0.92]
	Three-class	Classes A2,B2, C2	0.71 [0.66; 0.79]
Recall	Binary	Class A1	0.96 [0.80; 1.00]
		Class B1	0.97 [0.87; 1.00]
	Three-class	Class A2	0.96 [0.57; 1.00]
		Class B2	0.89 [0.64; 0.97]
		Class C2	0.75 [0.00; 1.00]
Precision	Binary	Class A1	0.77 [0.45; 0.98]
		Class B1	0.96 [0.80; 1.00]
	Three-class	Class A2	0.77 [0.45; 0.98]
		Class B2	0.73 [0.60; 0.78]
		Class C2	0.60 [0.00; 0.66]

Classification labels correspond to VO_2 max classes in mL/kg/min, as defined in ref. 27. Binary classification: A1, "very poor" and "poor" classes; B1, "fair" to "excellent" classes. Three-class classification: A2, "very poor" and "poor"; B2, "fair" and "average"; C2, "good" to "excellent".

the three-class analysis with fewer "good/excellent" subjects. Applying these a-posteriori classifications to ABR predictions showed high precision (0.96) and recall (0.96) in classifying subjects in class B1 (i.e., "fair" to "excellent"). These results confirm a potential application of such predictive model in astronauts during spaceflight: a B1 classification could be interpreted as a reliable estimate for a normal VO_2 max level, thus suggesting skipping the routine in-flight PFE protocol. Conversely, A1 subjects would still require assessment to confirm sub-normal VO_2 max.

Interestingly, the observed reduced prediction accuracy in longer (60-day) HDT campaigns may be partly due to the larger temporal gap (4–5 days) between the Holter ECG acquisition at the end of bed rest and the VO_2 max assessment after bed rest, compared to shorter campaigns (1–2 days). This delay could have created a mismatch between ECG-derived features feeding the model and the measured VO_2 max.

Feature importance analysis evidenced that Δ features compared to the baseline value, together with basal VO_2 max, had a predominant role in predicting post-HDT VO_2 max. Among these, repolarization-related features such as T-wave amplitude and QT-corrected correlation dimension, and the zero-amplitude test p-value for the Cosinor analysis performed on QT-related parameters reflecting circadian rhythm strength, emerged as key predictors.

The T-wave amplitude has been shown to correlate with various variables measured on a standard stress test in healthy individuals, and thus it has been proposed as a marker of physical fitness[36]. Altered T-wave morphology and complexity (e.g., QRS-T angle and gradient) have been linked to increased risk of life-threatening arrhythmia[37]. In control groups of bed rest studies, we previously observed how 5-day HDT caused reduced T-wave amplitude/area, likely due to hypovolemia and autonomic changes[38], while 60-day HDT induced night-time T-wave alternans, suggesting repolarization instability[24].

T wave changes are sensitive to potassium levels, with hyperkalemia producing peaked T-waves and hypokalemia causing T-wave depression or inversion[39,40]. Circadian rhythms, strongly regulating potassium homeostasis including urinary excretion[41], can be disrupted by bed rest or microgravity[42,43], thus altering electrolytes, hydration, and calcium/potassium balance, making potassium a key element in space nutrition[44]. As we previously shown[26], in subjects undergoing HDT as control group the amplitude of day-night RR and QTend oscillations decreased from 5-days to 60-days HDT, with no recovery and anticipated acrophase. Conversely, when gravity was restored, differences between the last day of the HDT and R + 0 in RR, QT, and their circadianity were significant for all HDT durations, and even more markedly in the 60-days HDT campaigns, with QTc even exceeding the 450 ms normality threshold[45].

The QT-corrected correlation dimension features, measuring the complexity of autonomic modulation of ventricular repolarization through nonlinear assessment of QT dispersion, also seemed to play a significant role in fitness level prediction. This aligns with evidence of an inverse relationship between HRV and QT dispersion, suggesting that reduced variability may reflect autonomic imbalance[46,47].

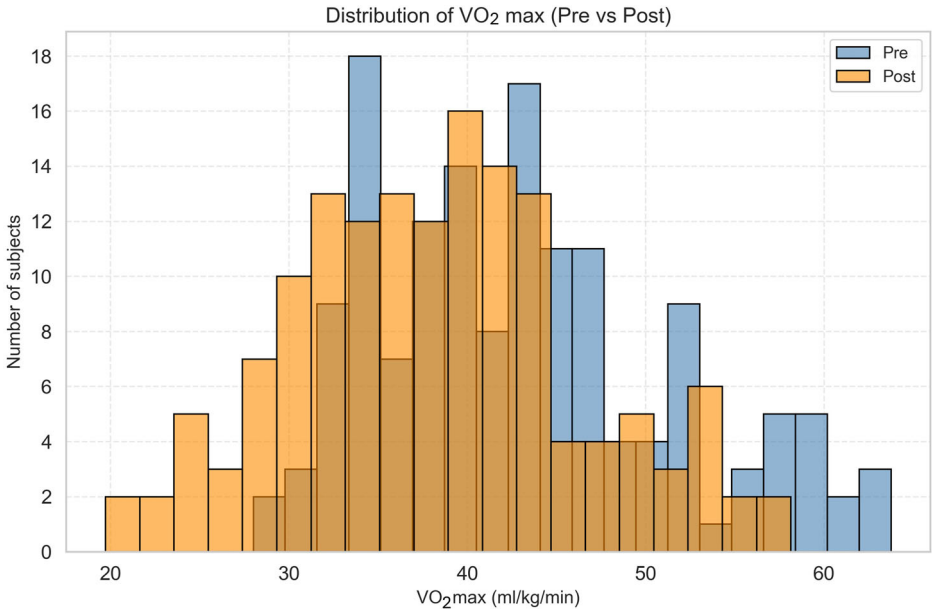

Fig. 2 | Distribution of VO_2 max. Distribution of VO_2 max values measured at BDC and at the end of HDT bed rest in the whole study population.

Although the relationship between T wave morphology and VO_2 max is not fully established, this work supports its possible correlation with the previously observed HDT-induced changes in ventricular repolarization and underlines the potential role of circadian rhythm entrainment in maintaining physiological and fitness homeostasis. Nevertheless, predicting VO_2 max values appeared more challenging after longer HDT. Previous studies on the control group of the same population showed that, after initial abrupt changes in cardiac electrical features, a new homeostasis is reached with HDT prolongation[26]. This suggests that the progressive, although non-linear, decline in exercise capacity cannot be fully explained by cardiac electrical variability alone. The interplay of the alterations, or adaptations, of cardiac electrical and structural adaptations[48–50], vascular remodeling[51], fluid redistribution and plasma volume loss[52,53], in turn connected with hypovolemia[52] and hypokalaemia[53], could play a significant role in improving the prediction of in-flight astronauts fitness.

An important limitation of this study is the inclusion of only male participants in the training and test of our model, dictated by data availability, which limits the generalizability of our findings. Gender differences in cardiac function and exercise capacity suggest that the mechanisms underlying VO_2 max may vary between males and females[54], and in turn CV responses to deconditioning and VO_2 max dynamics could differ between genders due to physiological and hormonal factors. Future research aiming to include both male and female participants would be beneficial to explore potential sex-specific differences in ECG predictors of aerobic capacity, thus improving a broader applicability of this approach.

In conclusion, this study demonstrates that VO_2 max deterioration due to cardiovascular deconditioning induced by HDT bed rest can be reliably predicted from exercise-free longitudinal 24-h Holter ECG recordings using ML approaches. The proposed test-agnostic approach achieved accuracy comparable to exercise-based prediction models, while avoiding the logistical and physiological constraints of in-flight testing. Such results could make it suitable for both astronaut monitoring during long-term missions as well as for terrestrial medicine, in contexts in which exercise protocols cannot be performed (bedridden patients, under physical rehabilitation, in intensive care). In future spaceflight applications, VO_2 max prediction from longitudinal 24-h ECG recordings could help anticipating fitness decline, personalizing exercise countermeasures, and reducing the need for routine PFE testing. Further validation in real space conditions, integration with wearable ECGs, and inclusion of additional physiological markers will help optimize this approach.

Methods

Study population and design

The data used in this study were collected during six HDT bed rest campaigns conducted between 2010 and 2018, as part of the European Space Agency (ESA) bed rest strategy. These included two short-duration (5 days), two mid-duration (21 days), and two long-duration (60 days) studies, conducted at the Institut de Médecine et de Physiologie Spatiales (MEDES) in Toulouse (France), or at the German Aerospace Center (Deutsches Zentrum für Luft- und Raumfahrt e.V, DLR) in Cologne (Germany). Each campaign, and relevant experiments conducted within, received approval by the respective Ethical Committee for Human Research at each of the hosting institutions. In all the six HDT bed rest campaigns, an only-male healthy population was recruited. Inclusion criteria comprised physically and mentally healthy subjects aged 20–45 years, able and willing to participate in the entire study, and to be randomly assigned to either the treatment or control group. All participants required to successfully pass psychological and medical screening, provide written informed consent, and be covered by social insurance. The complete list of inclusion and exclusion criteria as defined in ESA guidelines[55] is reported in Supplementary Table 2. Short- and mid-duration HDT campaigns were randomized cross-over studies, with each subject completing HDT two (DLR 21-day) or three times (DLR 5-day; MEDES 5- and 21-day), once as control and the others with applied countermeasures. Long-duration campaigns were multi-group, with each subject undergoing HDT once, assigned randomly to control or countermeasure.

For this analysis, each participant in an HDT session, regardless of control or countermeasure allocation, was treated as a unique subject (i.e., a subject repeating three times the HDT will be considered as three distinct subjects). Each study included a baseline data collection (BCD) phase during acclimatization, a HDT bed rest period, and a recovery phase. Wake and sleep times were fixed for all campaigns. For a detailed description of the design relevant to each campaign, please refer to the Supplementary Material provided in ref. 26. Participant's age and body mass index (BMI) are summarized in Table 4, along with the number of subjects that were excluded due to missing or low-quality data. In addition, the distribution of the VO_2 max values in the population at both baseline and after HDT bed rest is presented in Fig. 2.

For each campaign and repetition, subjects completed a VO_2 max test during BDC and one during recovery. VO_2 max was assessed using an incremental cycle ergometer test, starting with a 5-minute baseline, followed

Table 4 | Summary of details relevant to the retrospectively studied population pooled from different ESA HDT bed rest campaigns

Campaign (Acronym)	Year	Total population	Excluded Subjects	Age (years)	BMI (kg/m^2)
MEDES 5-day (BRAG)	2010	36 (12 ×3)	4	32.8 [32;38] range: 21–41	23.7 [23.1; 25.1] range: 19.5–27
DLR 5-day (SAG)	2010-2011	30 (10 ×3)	1	29.6 [25.3;31.3] range: 23-44	24.4 [23.4;25.1] range: 22.5-26.8
MEDES 21-day (MNX)	2012-2013	36 (12 ×3)	7	35.8 [29.2;41.9] range: 20-45	22.5 [20.8;23.8] range: 19–27.9
DLR 21-day (MEP)	2011-2012	20 (10 ×2)	1	31.5 [28.3;36.2] range: 23–43	23.7 [22.2;25.0] range: 21.1–25.9
MEDES 60-day (Cocktail)	2017-2018	20	2	34.2 [27.8;41] range: 20–45	23.8 [22.6;25.3] range: 21.1–27.7
DLR 60-day (RSL)	2015-2016	23	2	29.1 [25;33.8] range: 21–42	23.5 [22.1;25.1] range: 19.6–26.3

Age and BMI are reported as median [25th percentile; 75th percentile], as well as with ranges.

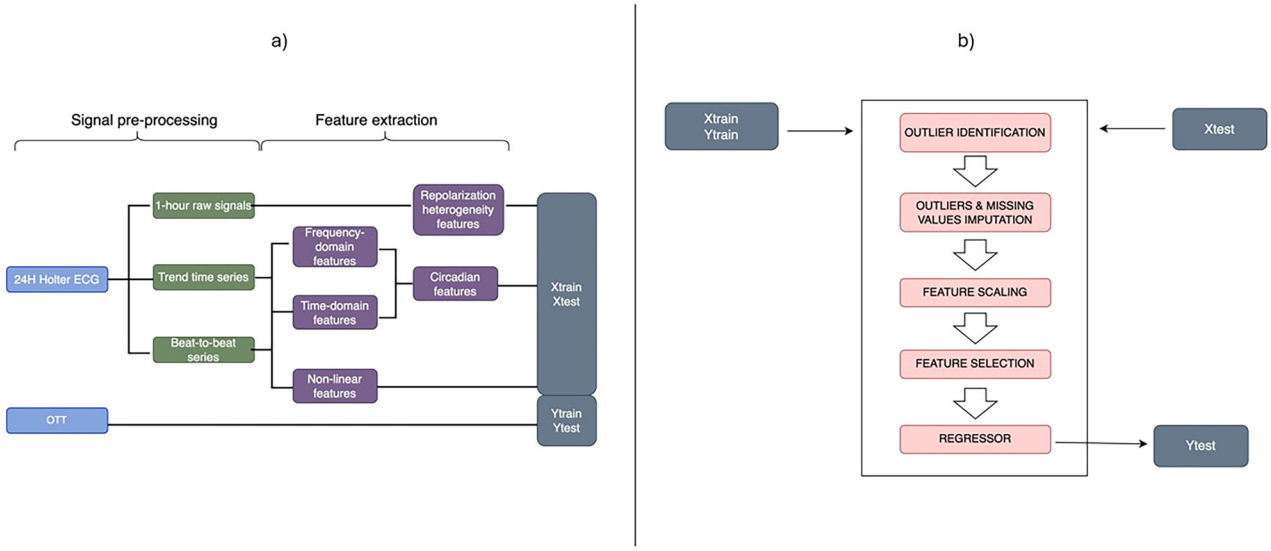

Fig. 3 | Features extraction workflow. a Workflow diagram of the regression framework for estimating VO_2 max using machine learning algorithms. The process begins with the collection of 24-h Holter ECG and VO_2 max reference data. Signals undergo processing and features extraction. These data are then split into training (Xtrain, Ytrain) and test (Xtest, Ytest) sets; (**b**) The pre-processing phase includes outlier identification, imputation of missing values and outliers, feature scaling, and feature selection. Finally, a regressor is applied to predict VO_2 max, with performance evaluation conducted using the test set.

by workload increases of 25% every 5 min, and then 25 W/min until exhaustion[55]. Additionally, 12-lead 24-h Holter ECG recordings (fs = 1000 Hz, H12+ Mortara Instrument Inc.) were recorded for each subject during BDC as well as short before the conclusion of the HDT. The VO_2 max test was never part of the 24-h Holter ECG monitoring period. In particular, during the BDC phase, the VO_2 max test was performed from 1 to 3 days after the Holter recordings, except in the DLR 5-day campaign where the VO_2 max test was performed 2 days before the Holter. During the recovery phase, the VO_2 max test was performed in different days, according to the bed rest campaign: on the same day of bed rest discontinuation (R + 0) for MEDES 5-day campaign, or one day after (R + 1) otherwise. The corresponding Holter ECG recordings were collected on the last day of HDT for the 5- and 21-day campaigns, or on HDT57-58 for the 60-day campaigns.

A total of 148 Holter 24-h ECG recordings at BCD and at the end of the HDT were available for model development, together with age, BMI, and the respective results of the VO_2 max test in terms of mL/kg/min, provided by ESA as part of the Core bed rest data. The distribution of the subjects in respect to their cardiovascular fitness classes based on their age and VO_2 max measured at both BDC and POST, following the classification norms as presented in ref. 27, is reported in Supplementary Table 1.

Signal pre-processing and features dataset creation

The workflow of the method developed in this study, presented hereafter, is schematized in Fig. 3(a: signal pre-processing and features extraction; b: regression framework).

Beat-to-beat series from the 24-h ECG signal were extracted using the proprietary Mortara Instrument research software SuperECG[56], including RR, QT and QT corrected (using Bazett[57], Fridericia[58], Framingham[59] and Hodges[60] formulae) intervals, T wave maximum downstroke velocity (T_{slope}), mean T wave amplitude along the first principal component axis (T_{amp}) and T wave symmetry (T_{symm}), computed as:

$$T_{symm} = \left(T_p - T_u\right) / \left(T_d - T_p\right) \quad (1)$$

where T_u is the point of maximum upward velocity on the T wave, T_p is the T wave peak, T_d is the point of maximum downward velocity on the T wave.

In addition, trend series of PR interval, P wave duration, QRS duration and T wave complexity (computed as the ratio of the second eigenvalue to the first) were computed over 5-min periods with 4 min overlap. Beat-by-beat and trend series were processed to remove outliers as described in ref. 26.

Frequency-domain, time-domain, geometrical, and non-linear HR and QT variability parameters, were additionally computed[61] (see ref. 62 for further details).

Subsequently, circadian rhythm's features were derived from beat-by-beat, trend, and hourly variability parameters using the Cosinor Analysis[63], resulting in the Midline Estimating Statistic Of Rhythm (MESOR), oscillation amplitude (OA) and the acrophase (ϕ). A zero-amplitude test ($p < 0.05$) was also applied to confirm rhythmicity.

Finally, repolarization heterogeneity parameters and their relationship with the RR interval were computed over the night period, as described in ref. 38. Specifically, the RT_{end} and RT_p intervals, the amplitude and area of the T wave, and QRS gradient and angle, reflecting the difference in depolarization/repolarization propagation directions[64], were calculated.

The computed parameters were used to build the features dataset for VO_2 max prediction. Specifically, for beat-by-beat, trend and hourly-computed parameters, the median value, inter-quartile range, and the 5th and 95th percentiles were computed over day, night, 24-h, and for the first 15 min after awakening. Features were extracted from each subject both at BDC and at HDT_{last}, and their percent difference (Δ%) was computed as:

$$\Delta\% = \left(\left(HDT_{last} - BDC\right)/BDC\right) \times 100 \quad (2)$$

and considered as a feature. Age, BMI, and baseline VO_2 max were also added, yielding a final dataset of 148 samples with 1086 features each.

VO_2 max prediction approach

Starting from the extracted features, several pre-processing steps were performed, including outliers identification, missing values imputation, feature scaling and feature selection, coupled with a VO_2 max regression model. Given the possibility of residual missing data after outlier removal, the K-Nearest Neighbor (KNN) imputation method was applied. Both centering and scaling were conducted individually for each feature. To manage the high dimensionality relative to sample size, cross-validated Lasso regression (LARS algorithm with L1 regularization) was employed, ensuring robustness against overfitting[65].

On the resulting dataset, the following regression models were evaluated and compared: Support Vector (SVR), K-nearest Neighbor (KNR), Decision Trees (DTR), and Random Forest (RFR) and AdaBoost (ABR) ensemble methods, considered for their ability to combine predictions of several base estimators[28], thus improving generalization and robustness over a single estimator. The regressor achieving the highest performance (i.e., ABR, see Table 1) during comparative evaluation was selected for subsequent analyses, including feature importance assessment. Finally, two a-posteriori classification schemes based on VO_2 max for each subject were applied based on the classes described in ref. 27: (i) binary classification (Class A1: "very poor/poor" vs. Class B1: "fair–excellent") and (ii) three-class classification (Class A2: "very poor/poor", Class B2: "fair/average", Class C2: "good–excellent").

Experimental procedure

To train and evaluate the regression models, nested Cross-validation (CrV) was applied. The dataset was divided into 10 folds for the purpose of outer loop CrV. In each outer loop, the training data were further split for an inner loop, where a grid search (3-fold) optimized the hyperparameters. The selected model was then tested on the corresponding outer fold. Model performance during the grid search was evaluated using the Mean Absolute Percentage Error (MAPE):

$$MAPE\left(y,\hat{y}\right) = \left(1/n\right)\sum_{i=1}^{n}\left|\left(y_i - \hat{y}_i\right)/\left(y_i\right)\right| \times 100 \quad (3)$$

where n is the number of samples, $\hat{y}_i$ is the predicted value of the i-th sample and y_i is the corresponding true value.

To enhance the representation of the minority class, offline data augmentation was carried out by introducing Gaussian noise to the original samples, ensuring a more balanced dataset. The list of the optimized hyperparameters during the validation strategy, KNN imputation and feature selection is summarized in the Supplementary Table 3.

Evaluation metrics and statistical analysis

The performances of the different proposed pipelines on unseen data were evaluated in terms of Mean Absolute Error (MAE), Root Mean Squared Error (RMSE), Mean Absolute Percentage Error (MAPE), and Symmetric Mean Absolute Percentage Error (SMAPE). MAE, SMAPE, and RMSE are defined as follows:

$$MAE\left(y,\hat{y}\right) = \left(1/n\right)\sum_{i=1}^{n}\left|\left(y_i - \hat{y}_i\right)\right| \quad (4)$$

$$SMAPE\left(y,\hat{y}\right) = \left(\frac{1}{n}\right)\sum_{i=1}^{n}\left|\left(y_i - \hat{y}_i\right)\right|/\left(\left|y_i\right| + \left|\hat{y}_i\right|\right) \times 100 \quad (5)$$

$$RMSE\left(y,\hat{y}\right) = \sqrt{\left(1/n\right)\sum_{i=1}^{n}\left|\left(y_i - \hat{y}_i\right)\right|^2} \quad (6)$$

Statistical analysis was performed to compare the results from the different pipelines (Friedman, $p < 0.05$, post-hoc Wilcoxon signed-rank test with Hochberg correction). Additionally, for the best pipeline, the MAE, MAPE, SMAPE, and RMSE were computed for each HDT campaign separately, to assess the potential impact of the HDT duration and of the HDT facility on the prediction performance.

For the model that yielded the best prediction results, the global Standard Error of the Estimate (SEE) was calculated as follows, to allow comparison with previous work:

$$SEE\left(y,\hat{y}\right) = \sqrt{\left(\sum_{i=1}^{n}\left|\left(y_i - \hat{y}_i\right)\right|^2\right)/\left(n-2\right)} \quad (7)$$

$$SEE_{adj}\left(y,\hat{y}\right) = \sqrt{\left(\sum_{i=1}^{n}\left|\left(y_i - \hat{y}_i\right)\right|^2\right)/\left(n-2\right)} \times \left(1/r^2\right) \quad (8)$$

where r^2 is the coefficient of determination. For the a-posteriori classification, accuracy, recall, and precision were computed for each class across the 10 cross-validation test folds.

Data availability

The datasets generated and analysed during the current study are not publicly available due to agreements with the European Space Agency (ESA), as they were collected during ESA campaigns. However, they are available from the corresponding author upon reasonable request, subject to ESA's data access policies.

Code availability

The underlying code used to analyse the data is not publicly available but may be made available to qualified researchers on reasonable request from the corresponding author.

Received: 13 August 2025; Accepted: 5 November 2025;
Published online: 22 November 2025

References

1. Goldberger, A. L. et al. Heart rate dynamics during long-term space flight: report on MIR cosmonauts. *Am. Heart J.* **128**, 202–204 (1994).
2. Anzai, T., Frey, M. A. & Nogami, A. Cardiac arrhythmias during long duration spaceflights. *J. Arrhythmia* **30**, 139–149 (2014).

3. Gallo, C., Ridolfi, L. & Scarsoglio, S. Cardiovascular deconditioning during long-term spaceflight through multiscale modeling. *npj Microgravity* **6**, 27 (2020).
4. Demontis, G. C. et al. Human pathophysiological adaptations to the space environment. *Front. Physiol.* **8**, 547 (2017).
5. G. Clement, Fundamentals of space medicine, Vol. 23, Springer Science & Business Media, 2011. https://doi.org/10.1007/978-1-4419-9905-4
6. Gunga, H. C., von Ahlefeld, V. W., Coriolano, H. J., Werner, A., Hoffmann, U. *Cardiovascular System, Red Blood Cells, and Oxygen Transport in Microgravity* (Springer, 2016).
7. Scott, J. P., Kramer, A., Petersen, N. & Green, D. A. The role of long term head-down bed rest in understanding inter-individual variation in response to the spaceflight environment: a perspective review. *Front. Physiol.* **12**, 614619 (2021).
8. Hill, A.V., Long, C.N.H., Lupton, H. Muscular exercise, lactic acid, and the supply and utilization of oxygen.In *Proceedings of the Royal Society of London. Series B, Containing Papers of a Biological Character*, **96**, 438−475 (1924).
9. Liguori, G., Feito, Y., Fountaine, C. & Roy, B. American college of sports medicine. *ACSM's Guidelines for Exercise Testing and Prescription* https://cir.nii.ac.jp/crid/1130006883078813721 (2020).
10. Convertino, V. A. Planning strategies for development of effective exercise and nutrition countermeasures for long-duration space flight. *Nutrition* **18**, 880–888 (2002).
11. English, K. L. et al. High intensity training during spaceflight: results from the NASA sprint study. *npj Microgravity* **6**, 21 (2020).
12. Trappe, T. et al. Cardiorespiratory responses to physical work during and following 17 days of bed rest and spaceflight. *J. Appl. Physiol.* **100**, 951–957 (2006).
13. Levine, B. D. et al. Maximal exercise performance after adaptation to microgravity. *J. Appl. Physiol.* **81**, 686–694 (1996).
14. Ade, C. J., Broxterman, R. M., Moore, A. & Barstow, T. J. Decreases in maximal oxygen uptake following long-duration spaceflight: role of convective and diffusive O2 transport mechanisms. *J. Appl. Physiol.* **122**, 968–975 (2017).
15. Ferretti, G. & Capelli, C. Maximal O2 consumption: effects of gravity withdrawal and resumption. *Respir. Physiol. Neurobiol.* **169**, S50–S54 (2009).
16. Moore, A. D., Lee, S. M., Stenger, M. B. & Platts, S. H. Cardiovascular exercise in the US space program: past, present and future. *Acta Astronautica* **66**, 974–988 (2010).
17. Moore, A. D. et al. Peak exercise oxygen uptake during and following long-duration spaceflight. *J. Appl. Physiol.* **117**, 231–238 (2014).
18. Greenisen, M. C., Fortney, S. M., Lee, S., Moore, A. D., Barrows, L. H. Variability of prediction of maximal oxygen consumption on the cycle ergometer using standard equations, *Tech. Rep.* (1993) No S-737.
19. Abut, F., Akay, M. F. Machine learning and statistical methods for the prediction of maximal oxygen uptake: recent advances. *Med. Dev. Evidence Res.* 369–379 (2015).
20. Ashfaq, A., Cronin, N. & Müller, P. Recent advances in machine learning for maximal oxygen uptake (VO2 max) prediction: a review. *Inform. Med. Unlocked* **28**, 100863 (2022).
21. Alzamer, H., Abuhmed, T. & Hamad, K. A short review on the machine learning-guided oxygen uptake prediction for sport science applications. *Electronics* **10**, 1956 (2021).
22. Shandhi, M. M. H. et al. Estimation of instantaneous oxygen uptake during exercise and daily activities using a wearable cardio-electromechanical and environmental sensor. *IEEE J. Biomed. Health Inform.* **25**, 634–646 (2020).
23. Materko, W. et al. Maximum oxygen uptake prediction model based on heart rate variability parameters for young healthy adult males at rest. *Open Access Biostat. Bioinform.* **2**, 1–7 (2018).
24. Martin-Yebra, A. et al. Assessment of ventricular repolarization instability in terms of T-wave alternans induced by head-down bed-rest immobilization. *Physiol. Meas.* **40**, 104001 (2019).
25. Martin-Yebra, A. et al. Evaluation of T-wave alternans activity under stress conditions after 5 d and 21 d of sedentary head-down bed rest. *Physiol. Meas.* **36**, 2041 (2015).
26. Solbiati, S., Martin-Yebra, A., Vaïda, P. & Caiani, E. G. Evaluation of cardiac circadian rhythm deconditioning induced by 5-to-60 days of head-down bed rest. *Front. Physiol.* **11**, 612188 (2021).
27. Shvartz, E. & Reibold, R. Aerobic fitness norms for males and females aged 6 to 75 years: a review. *Aviat. Space Environ. Med.* **61**, 3–11 (1990).
28. Seni, G., Elder, J. Ensemble methods in data mining: improving accuracy through combining predictions, Morgan Claypool Publishers https://doi.org/10.1007/978-3-031-01899-2 (2010).
29. Abut, F., Akay, M. F. & George, J. Developing new VO2max prediction models from maximal, submaximal and questionnaire variables using support vector machines combined with feature selection. *Comput. Biol. Med.* **79**, 182–192 (2016).
30. Akay, M. F. et al. Development of novel maximal oxygen uptake prediction models for Turkish college students using machine learning and exercise data. In *2017 9th International Conference on Computational Intelligence and Communication Networks* **79**, 186–189 (2017).
31. Heil, D. P., Freedson, P. S., Ahlquist, P. S., Price, J., Ripp, J. M. Nonexercise regression models to estimate peak oxygen consumption. *Med. Sci. Sports Exerc.* **27** (1995).
32. George, J. D., Vehrs, P. R., Allsen, P. E., Fellingham, G. W. & Fisher, A. G. VO2max estimation from a submaximal 1-mile track jog for fit college individuals. *Med. Sci. Sports Exerc.* **25**, 401–406 (1993).
33. Ebbeling, C. B., Ward, A., Puleo, E. M., Widrick, J. & Rippe, J. M. Development of a single-stage submaximal treadmill walking test. *Med. Sci. Sports Exerc.* **23**, 966–973 (1991).
34. Nielson, D. E., George, J. D., Vehrs, P. R., Hager, R. L. & Webb, C. V. Predicting VO2max in college-aged participants using cycle ergometry and perceived functional ability. *Meas. Phys. Educ. Exerc. Sci.* **14**, 252–264 (2010).
35. Molina-Garcia, P. et al. Validity of estimating the maximal oxygen consumption by consumer wearables: a systematic review with meta-analysis and expert statement of the INTERLIVE network. *Sports Med.* **52**, 1577–1597 (2022).
36. Arbel, Y. et al. T-wave amplitude is related to physical fitness status. *Ann. Noninvasive Electrocardiol.* **17**, 214–218 (2012).
37. Voulgari, C., Pagoni, S., Tesfaye, S. & Tentolouris, N. The spatial QRS-T angle: implications in clinical practice. *Curr. Cardiol. Rev.* **9**, 197–210 (2013).
38. Caiani, E. G. et al. Impaired T-wave amplitude adaptation to heart-rate induced by cardiac deconditioning after 5-days of head-down bed-rest. *Acta Astronautica* **91**, 166–172 (2013).
39. Varga, C. et al. ECG alterations suggestive of hyperkalemia in normokalemic versus hyperkalemic patients. *BMC Emerg. Med.* **19**, 1–9 (2019).
40. Wang, D. H., Xiqiang, G. Li, Electrocardiographic manifestations in severe hypokalemia. *J. Int. Med. Res.* **48** https://doi.org/10.1177/0300060518811058 (2020).
41. Gumz, M. L. & Rabinowitz, L. Role of circadian rhythms in potassium homeostasis. *Semin. Nephrol.* **33**, 229–236 (2013).
42. McKenna, H. T., Reiss, I. K. & Martin, D. S. The significance of circadian rhythms and dysrhythmias in critical illness. *J. Intensive Care Soc.* **18**, 121–129 (2017).
43. Fuller, C. A., Hoban-Higgins, T. M., Griffin, D. W. & Murakami, D. M. Influence of gravity on the circadian timing system. *Adv. Space Res. Off. J. Comm. Space Res.* **14**, 399–408 (1994).
44. Tang, H., Rising, H. H., Majji, M. & Brown, R. D. Long-term space nutrition: a scoping review. *Nutrients* **14**, 194 (2021).
45. Rautaharju, P. M., Surawicz, B. & Gettes, L. S. AHA/ACCF/HRS recommendations for the standardization and interpretation of the electrocardiogram: part IV: the ST segment, T and U waves, and the QT interval: a scientific statement from the American Heart Association

electrocardiography and arrhythmias committee, council on clinical cardiology; the American College of Cardiology Foundation; and the Heart Rhythm Society: endorsed by the International Society for Computerized Electrocardiology. *Circulation* **119**, e241–e250 (2009).
46. Mazur, A. et al. Relationship between autonomic control of heart rate and QT dispersion in patients with acute anterior wall myocardial infarction. *Ann. Noninvasive Eectrocardiol.* **4**, 152–157 (1999).
47. Nahshoni, E. et al. Complexity of the dynamic QT variability and RR variability in patients with acute anterior wall myocardial infarction: a novel technique using a non-linear method. *J. Electrocardiol.* **37**, 173–179 (2004).
48. Perhonen, M. A. et al. Cardiac atrophy after bed rest and spaceflight. *J. Appl. Physiol.* **91**, 645–653 (2001).
49. Henry, W. L., Epstein, S. E., Griffith, J. M., Goldstein, R. E., Redwood, D. R. Effect of prolonged space flight on cardiac function and dimensions. In *NASA. Johnson Space Center Proc. of the Skylab Life Sci. Symp* 2 (1974).
50. Koenig, S. C. et al. Evidence for increased cardiac compliance during exposure to simulated microgravity. *Am. J. Physiol.* **275**, R1343–R1352 (1998).
51. Tuday, E. C., Meck, J. V., Nyhan, D., Shoukas, A. A. & Berkowitz, D. E. Microgravity-induced changes in aortic stiffness and their role in orthostatic intolerance. *J. Appl. Physiol. (Bethesda, Md.1985)* **102**, 853–858 (2007).
52. Leach, C. S. et al. Regulation of body fluid compartments during short-term spaceflight. *J. Appl. Physiol. (Bethesda, Md. 1985)* **81**, 105–116 (1996).
53. White, R. J., Blomqvist, C. G. Central venous pressure and cardiac function during spaceflight. *J. Appl. Physiol. (Bethesda, Md.1985)* **85**, 738–746 (1998).
54. Wheatley, C. M., Snyder, E. M., Johnson, B. D. & Olson, T. P. Sex differences in cardiovascular function during submaximal exercise in humans. *SpringerPlus* **3**, 445 (2014).
55. Angerer, O., Larina, I., Cromwell, R. Guidelines for standardization of bed rest studies in the spaceflight context. In Sundblad, P., Orlov, O. *The International Academy of Astronautics* June 2014, https://iaaspace.org/product/guidelines-for-standardization-of-bed-rest-studies-in-the-spaceflight-context/.
56. Mortara, D. W. Automated QT measurement and application to detection of moxifloxacin-induced changes. *Ann. Noninvasive Electrocardiol.* **14**, S30–S34 (2009).
57. Bazett, H. An analysis of the time-relations of electrocardiograms. *Ann. Noninvasive Electrocardiol.* **2**, 177–194 (1997).
58. Fridericia, L. S. The duration of systole in an electrocardiogram in normal humans and in patients with heart disease. *Ann. Noninvasive Electrocardiol.* **8**, 343–351 (2003).
59. Sagie, A., Larson, D. E., Goldberg, D., Bengtson, P. M. & Levy, D. An improved method for adjusting the QT interval for heart rate (The Framingham Heart Study). *Am. J. Cardiol.* **70**, 797–801 (1992).
60. Hodges, M. Bazett's QT correction reviewed-evidence that a linear QT correction for heart is better. *J. Am. Coll. Cardiol.* **1**, 694 (1983).
61. Shaffer, F. & Ginsberg, J. P. An overview of heart rate variability metrics and norms. *Front. Public Health* **5**, 258 (2017).
62. Solbiati, S. & Caiani, E. G. Heart rate variability adaptation due to head-down tilt bed rest: insights from pooled data of six campaigns. *npj Microgravity* **11**, 48 (2025).
63. Refinetti, R., Cornelissen, G. & Halberg, F. Procedures for numerical analysis of circadian rhythms. *Biol. Rhythm Res.* **38**, 275–325 (2007).
64. Burger, H. C. A theoretical elucidation of the notion ventricular gradient. *Am. Heart J.* **53**, 240–246 (1957).
65. Emmert-Streib, F. & Dehmer, M. High-dimensional LASSO-based computational regression models: regularization, shrinkage, and selection. *Mach. Learn. Knowl. Extraction* **1**, 359–383 (2019).

Acknowledgements

We would like to thank all the involved ESA, MEDES, DLR personnel for their invaluable competence in organizing such experiments, as well as all the participating volunteers. This work was supported by the Italian Space Agency (Contracts 2018-7-U.0 (QT-BED), 2022-9-U.0 (CMR-PP), 2022-10-U.0 (CMR-SDM), PI and recipient E.G.C.).

Author contributions

E.G.C. conceived and designed the research. E.G.C. contributed to the data acquisition. All authors contributed to the development of analytical tools. S.S., M.C.F. and R.B. analysed the data. S.S., M.C.F., S.M. and E.G.C. wrote the manuscript. All the authors read and approved the manuscript.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary information The online version contains supplementary material available at https://doi.org/10.1038/s41526-025-00542-4.

Correspondence and requests for materials should be addressed to Enrico Gianluca Caiani.