

Article

Not peer-reviewed version

Machine Learning Models for Identifying Muscle Oxygenation Abnormalities Across Activity States in People with Post-COVID Syndrome

[Dimitrios Megaritis](#)*, Emily Hume, Enya Daynes, Rachael Evans, [Yifeng Zeng](#), Sally J. Singh, Carlos Echevarria, Peter D. Wagner, [Ioannis Vogiatzis](#)

Posted Date: 23 March 2026

doi: 10.20944/preprints202603.1803.v1

Keywords: COVID-19; spectroscopy; near-infrared; muscle oxygenation; machine learning; artificial intelligence; time-series analysis



Preprints.org is a free multidisciplinary platform providing preprint service that is dedicated to making early versions of research outputs permanently available and citable. Preprints posted at Preprints.org appear in Web of Science, Crossref, Google Scholar, Scilit, Europe PMC.

Copyright: This open access article is published under a [Creative Commons CC BY 4.0 license](#), which permit the free download, distribution, and reuse, provided that the author and preprint are cited in any reuse.

Article

Machine Learning Models for Identifying Muscle Oxygenation Abnormalities Across Activity States in People with Post-COVID Syndrome

Dimitrios Megaritis ^{1,*}, Emily Hume ¹, Enya Daynes ^{2,3}, Rachael Evans ^{2,3}, Yifeng Zeng ⁴, Sally J. Singh ³, Carlos Echevarria ^{6,7}, Peter D. Wagner ^{1,5} and Ioannis Vogiatzis ¹

¹ Faculty of Health and Life Sciences, Northumbria University, Newcastle upon Tyne, Tyne and Wear, UK

² Department of Respiratory Medicine, University of Leicester, Leicester, UK

³ NIHR Leicester Biomedical Research Centre, University Hospitals of Leicester NHS Trust, Leicester, UK

⁴ Department of Computer and Information Sciences, Faculty of Engineering and Environment, Northumbria University, Newcastle upon Tyne, Tyne and Wear, UK

⁵ Department of Medicine, University of California, San Diego, USA

⁶ Translational and Clinical Research Institute, Faculty of Medical Sciences, Newcastle University, Newcastle-upon-Tyne, UK

⁷ The Newcastle upon-Tyne Hospitals NHS Foundation Trust, Newcastle-upon-Tyne, UK

* Correspondence: d.megaritis@northumbria.ac.uk

New findings

1. What is the central question of this study?

Can machine learning models applied to near-infrared spectroscopy (NIRS) time-series data detect muscle oxygenation abnormalities across different activity states at a personalised level?

2. What is the main finding and its importance?

Machine learning models successfully identified subtle, individualised physiological impairments in muscle oxygenation patterns during activity in people with post-COVID syndrome that were not detectable through traditional group-level statistical analysis. This approach provides a reproducible, open-access AI-based pipeline for personalised physiological assessment. Importantly, the models can be retrained with new and larger data from different populations or conditions, supporting adaptable applications in physiology monitoring, individualised rehabilitation protocols, and precision medicine approaches.

Abstract

Post-COVID syndrome has been associated with potentially impaired exercise capacity. Here, we postulate that machine learning models trained on near-infrared spectroscopic (NIRS) signals collected during different physical activity states from four optodes over the quadriceps, using features selected by principal component analysis (PCA), can detect distinct patterns of oxygenation. These patterns differentiate post-COVID syndrome from healthy controls and are not evident from traditional analysis of NIRS signals. 228 time-series NIRS datasets from four optodes over different quadriceps regions were collected across multiple activity states in post-COVID syndrome and healthy participants. PCA was performed to reduce dimensionality and identify data patterns. K-Nearest Neighbour with Dynamic Time Warping, Canonical Interval Forests (CIF) and Convolutional Neural Network (CNN) models were trained on NIRS-derived features to classify

post-COVID syndrome -related muscle abnormalities versus healthy responses. PCA revealed that tissue oxygenation index (StiO₂) was the most effective parameter separating the populations, whereas the normalised total haemoglobin index (nTHI) was most sensitive to activity states. Learning models incorporating StiO₂ and nTHI, exhibited excellent performance in distinguishing between populations, with CIF and CNN exhibiting best performance (Kappa>0.69, F1-score>0.85, Sensitivity>0.85, Precision>0.88, Accuracy>0.85, AUC>0.95). However, local muscle StiO₂ heterogeneity and StiO₂ on-transient mean response time did not show significant differences between populations. Our findings demonstrate the efficacy of learning models trained on time-series muscle oxygenation data for detecting distinct muscle oxygenation patterns in post-COVID syndrome participants. This provides a novel, non-invasive approach applicable at the individual level for identifying distinctive muscle oxygenation patterns, where traditional analytical methods lack sensitivity.

Keywords: COVID-19; spectroscopy; near-infrared; muscle oxygenation; machine learning; artificial intelligence; time-series analysis

Introduction

Recent advances in analytical techniques for biomedical signals (HR, PPG, ECG, EEG), combined with machine learning, have enhanced our ability to monitor human physiology (Abbaspourazad et al., 2023; Pillai et al., 2024). These models capture high-dimensional features, detecting subtle physiological and pathological patterns and supporting data-driven clinical decisions (Esteva et al., 2019).

Long-COVID is a respiratory-related condition where machine learning may help identify distinct pathological patterns and develop novel diagnostic tools (Malhotra et al., 2023). Patients often exhibit post-exertional fatigue associated with altered skeletal muscle metabolism, including shifts toward fast-fatigable fibres and reduced mitochondrial enzyme activity (Appelman et al., 2024). Muscle biopsies and near-infrared spectroscopy (NIRS) reveal compromised vasculature and slower O₂ saturation kinetics in the vastus lateralis during exercise, indicating impaired muscle oxygen delivery and utilisation compared to healthy individuals (Appelman et al., 2024; Souza et al.).

AI has been employed to analyse cough and breathing sounds linked to COVID-19 (Brown et al., 2020; Kapoor et al., 2023; Modi et al., 2024). However, learning algorithms have not yet been applied to NIRS-derived muscle oxygenation signals in any population. NIRS is a responsive tool that reflects tissue oxygen supply-demand balance and is increasingly recognised in health research (Tuesta et al., 2022). Machine learning applied to physiological time-series may uncover latent patterns undetectable by conventional methods, supporting personalised, individual-level assessment.

In the present study we investigated whether machine learning and deep learning models could classify biosignals based on the individual level on NIRS derived vastus lateralis muscle oxygenation data. The models were trained on data collected from previously hospitalised people with post-COVID syndrome and age-matched healthy participants with no history of SARS-CoV-2 infection at rest, during graded levels of exercise and in recovery from exercise. The aim of this study was three-fold: Firstly, to apply principal component analysis (PCA) on raw muscle oxygenation time-series as a data-driven approach to explore underlying patterns, reduce dimensionality, and inform feature selection for the learning models development. Secondly, to assess the efficacy of machine and deep learning models trained with NIRS muscle oxygenation time-series data for distinguishing between participants with post-COVID syndrome muscle abnormalities and normal controls during different activity states. Thirdly, to investigate potential differences in vastus lateralis muscle oxygenation patterns between people with post-COVID syndrome and healthy participants as measured by validated statistical methods that assesses the distribution of regional muscle blood flow in relation to regional muscle metabolic rate as well as the kinetic responses (Vogiatzis et al., 2015).

Methods

Participants

Participants with Post-COVID syndrome following a hospital admission for COVID-19 were recruited from the usual care group of the 'PHOSP-Rehabilitation' study (ISRCTN10980107/ISRCTN13293865)(Daynes et al., 2023) (Yorkshire & the HumberLeeds West Research Ethics Committee (20/YH/0225). Ethical approval for the data collection of the present study was issued by Northumbria University ethics committee on 07/04/2022(UNN45059). Entry criteria for the participants with post-COVID syndrome included being an adult, having been admitted to the hospital during a confirmed acute episode of COVID-19, and experiencing ongoing symptoms for >12-weeks. These symptoms include one or more of the following: reduced activity/exercise tolerance, fatigue, dyspnoea, musculoskeletal pain, and brain-fog(Sivan & Taylor, 2020). Healthy age-matched individuals with no medical conditions or COVID-19 infection history (based on hospital records and self-reports) were selected to match the post-COVID syndrome group in age and sex. Given the exploratory nature of this machine-learning study, the sample size was pragmatically determined by data availability, and model evaluation used 5-fold cross-validation, maximising data utilisation

Experimental Methods

Experiments were conducted in two visits. In visit 1, participants underwent spirometry lung function assessment that was followed by an incremental exercise test to their highest tolerable workload [peak work rate (PWR)] (Table 2). The incremental exercise test was performed on an electromagnetically braked cycle ergometer (Ergoline 800; Sensor Medics, Anaheim, CA) starting at 20 Watts and increasing by 10 Watts every minute to exhaustion, with the participants maintaining a pedalling frequency of 50-60 revolutions-per-minute. Quadriceps muscle strength was assessed as maximum voluntary contraction (QMVC), using an isokinetic dynamometer at 90° knee and hip flexion. In addition, post-COVID syndrome and healthy participants underwent the incremental shuttle walk test(ISWT) as previously described(Singh et al., 1992). Dyspnoea was assessed using the Dyspnoea-12 questionnaire(Yorke & Armstrong, 2014) and the modified Medical Research Council(mMRC) scale(Yorke et al., 2022), while fatigue was evaluated using the Functional Assessment of Chronic Illness Therapy (FACIT) fatigue scale(Webster et al., 2003). In visit 2, following resting measurements, participants completed a graded cycling protocol. Bouts of constant-load exercise were sustained for four minutes at each of three work rates, corresponding to 20%, 50%, and 80% of PWR. Constant-load exercise tests were preceded by a 3-minute rest period, followed by 3 minutes of unloaded pedalling. Graded bouts of exercise were separated by 20-minute rest periods. During the incremental and constant-load tests, heart rate and percentage O₂ saturation were determined using a pulse oximeter (Nonin 8600; Nonin Medical, North Plymouth, MN). Ratings of dyspnoea and leg discomfort were assessed at the end of each bout using the 1-10 modified borg scale(Borg, 1982).

Regional NIRS Data Collection

Four pairs of NIRS optodes were placed over the upper, middle, and lower regions of the left vastus lateralis to measure muscle oxygenation (O₂Hb, HHb, StIO₂, nTHI) at 5 Hz using a 4-channel spectrophotometer, with optodes secured and shielded from external light (online supplement, p.2).

PCA

PCA was applied to reduce the dimensionality of the NIRS data, extracting the most relevant variance-based patterns between populations and activity states. In depth details are provided in the online supplement (p.2). The first two principal components (PC1 and PC2) were plotted (Figure 1, Figures S3-9), with data points color-coded to indicate population and period, allowing visual assessment of group separability as well as temporal differences.

Feature Selection

Selected time periods were grouped into categories such as rest, unloaded exercise (warm-up), loaded exercise (sustained at 20%, 50%, and 80% of PWR), and recovery. Four different feature selection approaches were developed in this study, each using distinct input data configurations reflect steady state activity or on transient phases (Table 1).

Table 1. Summary of the four feature selection strategies used in the study. Feature sets differ in the time window of the data, input signals, and whether period labels were included, allowing evaluation of model performance under different configurations.

Feature Set	Time Window	Periods Included	Signals Used	Period Labels
A	Last 30 s	Rest, Warm-up, 20%, 50%, 80% PWR, Recovery	StiO ₂	Yes (one-hot encoded)
B	First 60 s	Rest, Warm-up, 20%, 50%, 80% PWR, Recovery	StiO ₂	Yes (one-hot encoded)
C	Last 30 s	Rest, Warm-up, 20%, 50%, 80% PWR, Recovery	StiO ₂ + nTHI	No
D	First 60 s	Rest, Warm-up, 20%, 50%, 80% PWR, Recovery	StiO ₂ + nTHI	No

Machine Learning Model’s Architectures

The study used three models for time-series classification: KNN with Dynamic Time Warping (DTW)(Tavenard et al., 2020), Canonical Interval Forests (CIF) with tree-based estimators(Middlehurst et al., 2020), and a 1D convolutional neural network for binary classification with four convolutional layers followed by a dense output layer. In-depth details for each model, including data preprocessing, input format, hyperparameter optimisation, cross-validation procedure, training, and computing environment, are presented in the online supplement (p.2-3). All model architectures and trained weights are openly available in the corresponding GitHub/Zenodo repository(Megaritis, 2025).

Performance Metrics

Cohen's Kappa, F1 Score, Sensitivity, Precision, Accuracy, and Receiver Operating Characteristic (ROC) curve with Area Under the Curve (AUC) were employed as performance metrics to evaluate the model's effectiveness in classifying the target outcomes (online supplement, p.5-6). These metrics were used to evaluate the model’s ability to classify each individual time series chunk.

Traditional Physiological Analyses

Regional heterogeneity of muscle StiO₂, quantified as the coefficient of variation across four vastus lateralis regions, and mean response time and kinetics of the quadriceps StiO₂ signal during transitions from unloaded to loaded cycling were assessed to reflect local blood flow, oxygen availability, and muscle oxygenation dynamics (full details in Supplement, p.6).

Statistical Tests

Wilcoxon rank-sum tests were used to compare demographic, clinical, and MRT data between groups. A two-way ANOVA was employed to compare StiO₂ heterogeneity between groups and across periods (rest, warm up, exercise, recovery). Data are presented as mean±SD, or median with interquartile range (IQR).

Results

Table 2 summarises participant demographics and clinical characteristics. Individuals with post-COVID syndrome commonly reported ongoing fatigue (reported by 10/12 participants) and breathlessness (9/12), and showed reduced exercise capacity.

Table 2. Demographics and clinical characteristics of the post-COVID syndrome and healthy counterparts.

	Post-COVID	Healthy	P
n	12	7	
Age, years, mean±SD	54.4±8.8	58.8±11.4	0.27
Sex, female, n(%)	1 (9)	2 (28)	-
BMI, kg/m², mean±SD	31.3±4.9	26.3±3.2	0.03
FEV ₁ % predicted, mean±SD	103±11	110±3	0.85
Dyspnoea 12, mean±SD	5.8±5.9	0±0	0.01
mMRC, median±SEM	1±0.3	0±0	0.03
FACIT, mean±SD	19.0±12.1	8.1±3.6	0.04
Days since hospital discharge, mean±SD	651 ± 114	-	-
Hospital length of stay, days, median (IQR)	7 (4.7, 11.7)	-	-
Intensive Care Unit utilisation, n (%)	4 (33.3)	-	-
CPAP during hospital stay, n (%)	4(33.3)	-	-
O ₂ supplementation during hospital stay, n (%)	11(92)	-	-
ISWT, m, mean±SD	638 ± 219	1025 ± 389	0.004
Dyspnoea at ISWT (Borg), median (IQR)	3.5 (2.75, 4)	1 (0, 1)	0.005
Leg discomfort at ISWT (Borg), median (IQR)	1 (0, 3.25)	1 (0, 1)	0.32
Peak Work Rate, Watts, mean±SD	136±35	170±77	0.19
Peak Work Rate, % predicted, mean±SD (van de Poppe et al., 2018)	51.2±10.9	76.7±16.8	0.001
Dyspnoea at Peak Work Rate (Borg), median (IQR)	5 (4, 5)	2 (0, 4)	0.17
Leg discomfort at Peak Work Rate (Borg), median (IQR)	5 (4.5, 6.5)	4.5 (4, 5)	0.34
SpO ₂ at Peak Work Rate, mean±SD	97.3±1.2	97.9±0.8	0.51
HR at Peak Work Rate, mean±SD	141±21	148±19	0.49
Quadriceps muscle force, kg, mean±SD	50.1±16.6	51.7±21.5	0.91
Quadriceps muscle force, % predicted, mean±SD	94.7±27.7	93.8±27.5	0.95

Data are presented as mean±SD or median with IQR; BMI: body mass index; FEV₁ % predicted: forced expiratory volume in the first second expressed as a percentage of predicted normal; Dyspnoea 12: a unidimensional measure reflecting the physical and affective aspects of dyspnoea, administered during rest and assessing breathing over the past few days; mMRC: modified Medical Research Council dyspnoea scale; FACIT: self-reported fatigue scale; Dyspnoea/Leg discomfort at PWR: 10-point Borg scale during peak work rate achieved during a cardiopulmonary exercise test; CPAP: continuous positive airway pressure applied during hospital stay. SpO₂: peripheral oxygen saturation measured via pulse oximetry.

PCA

Figure 1 illustrates the distinct clustering patterns of the principal components between populations derived from the St_iO₂ combined data recorded over the first 60 seconds and the last 30 seconds of each period. Furthermore, Figure S7 shows that there are no group discrepancies regarding the periods for St_iO₂ recordings. Distinct groupings for nTHI are evident only by period (Figure S8), with no separation by population (Figure S4). All other variables show no separation by either population or period (presented in the online supplement colour coded by population (Figures S2-4) or by period (Figures S5-8)).

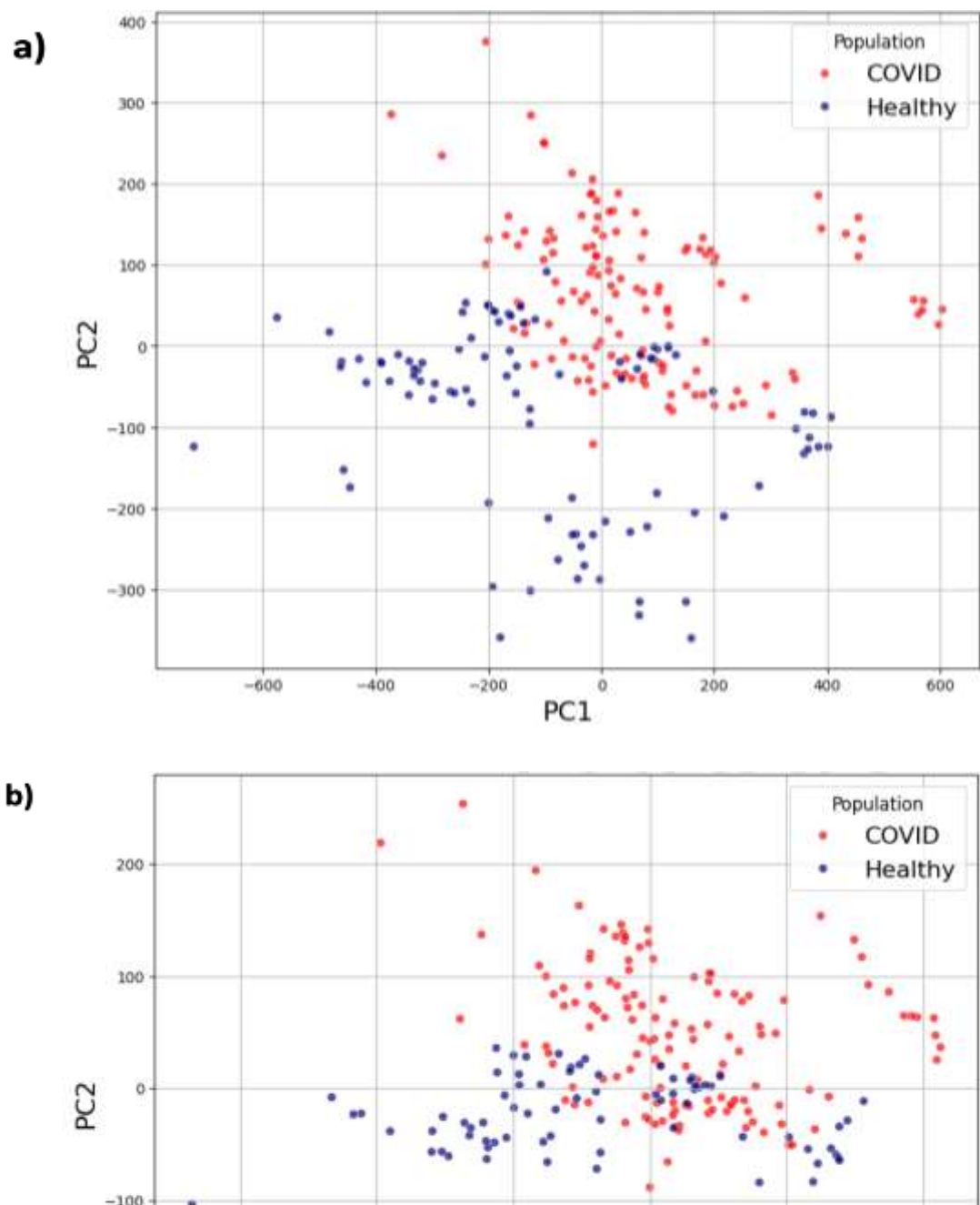


Figure 1. a) PCA of StiO₂ data from the first 60 seconds of each period; b) PCA of StiO₂ data from the last 30 seconds of each period. PC1: The first principal component represents the direction in the data that has the greatest variance. It captures the most significant pattern in the dataset; PCA2: The second principal component is orthogonal to PC1 and captures the next greatest variance in the data. It represents the second significant pattern, but without overlapping or correlating with the information captured by PC1.

Machine Learning Models Overall Performance

KNN

Models demonstrated poor to strong performance (Table 3). The ROC-AUC for all models are shown in Figure 2.

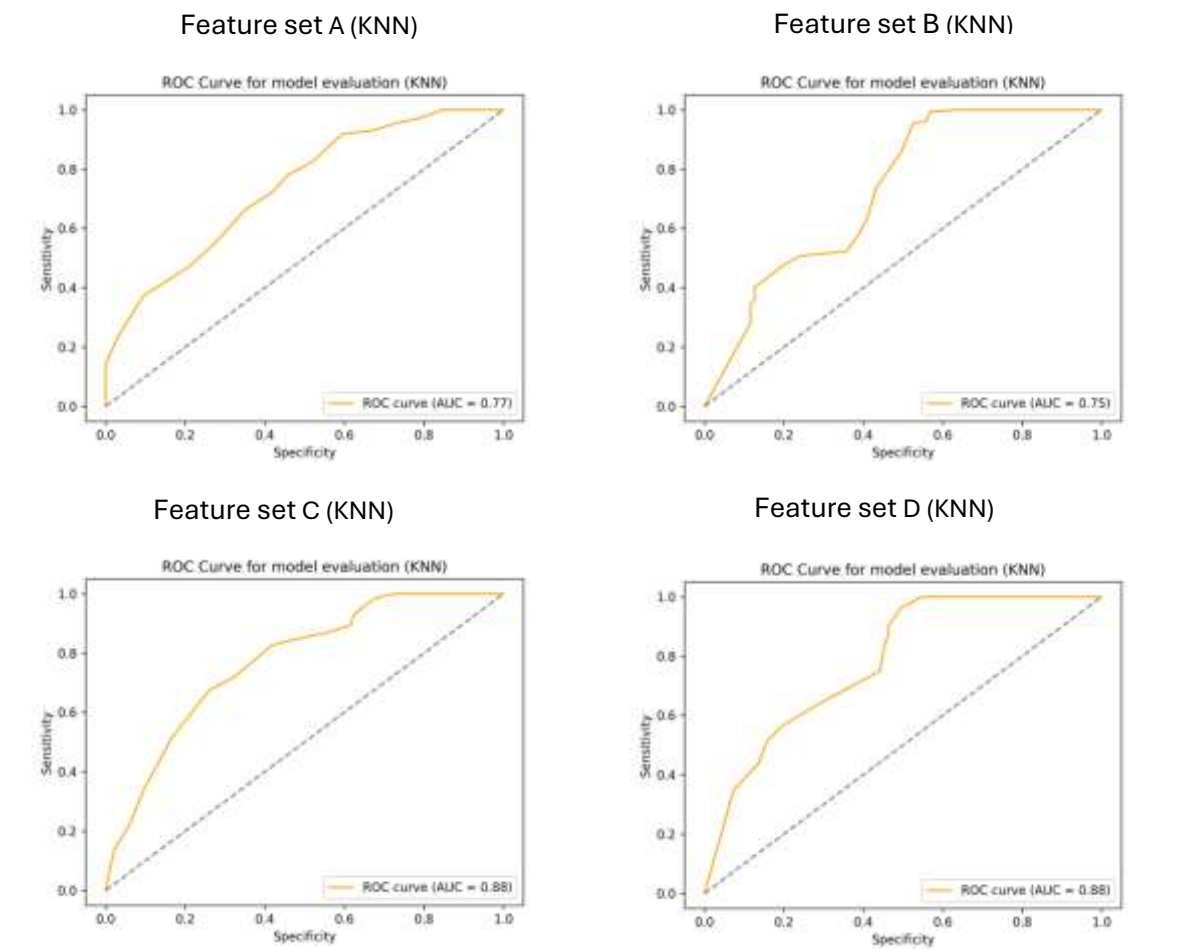


Figure 2. ROC-AUC for KNN machine learning models

CIF

Models demonstrated good to strong performance metrics (Table 3). The ROC-AUC for all models are shown in Figure 3.

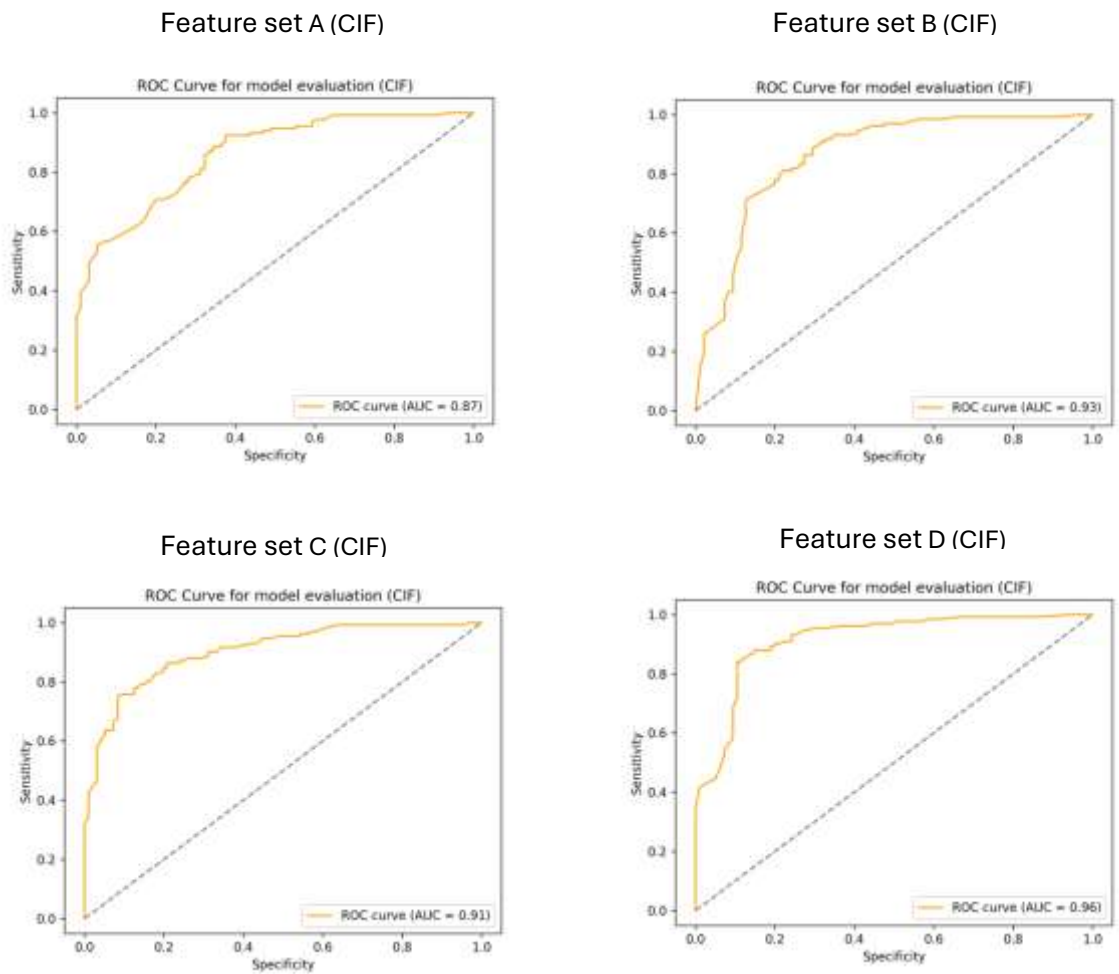


Figure 3. ROC-AUC for CIF machine learning models

CNN

Models demonstrated average to strong performance metrics (Table 3). The ROC-AUC for all models are shown in Figure 4.

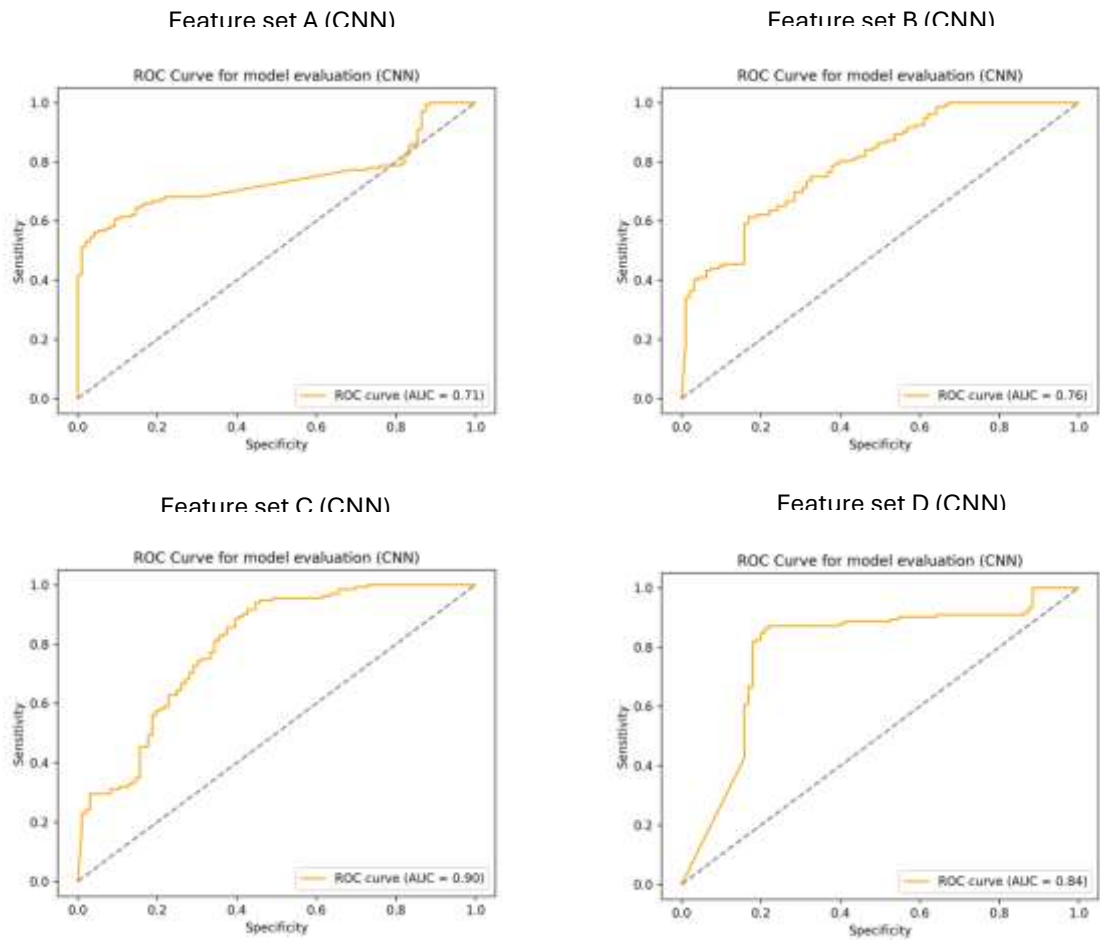


Figure 4. ROC-AUC for all models

Table 3. Performance metrics for all models and feature sets

KNN				
Metric	Feature set A (KNN)	Feature set B (KNN)	Feature set C (KNN)	Feature set D (KNN)
Kappa	0.253	0.266	0.395	0.417
F1	0.669	0.648	0.658	0.714
Sensitivity	0.647	0.65	0.718	0.723
Precision	0.648	0.648	0.778	0.701
Accuracy	0.685	0.659	0.689	0.745
AUC	0.775	0.748	0.882	0.876
CIF				
Metric	Feature set A (CIF)	Feature set B (CIF)	Feature set C (CIF)	Feature set D (CIF)
Kappa	0.406	0.594	0.562	0.697
F1	0.746	0.795	0.810	0.854
Sensitivity	0.719	0.789	0.786	0.850

Precision	0.740	0.871	0.821	0.882
Accuracy	0.778	0.810	0.817	0.859
AUC	0.87	0.933	0.910	0.957
CNN				
Metric	Feature set A (CNN)	Feature set B (CNN)	Feature set C (CNN)	Feature set D (CNN)
Kappa	0.127	0.374	0.431	0.654
F1	0.528	0.656	0.733	0.829
Sensitivity	0.558	0.683	0.763	0.829
Precision	0.502	0.700	0.729	0.831
Accuracy	0.572	0.710	0.742	0.830
AUC	0.711	0.763	0.898	0.837

Heterogeneity Analysis

There was no significant interaction between population and time on the heterogeneity of St*i*O₂ (p=0.98)(Figure S9).

On transient MRT Analysis

There was no difference (p=0.23) in the mean St*i*O₂ mean response time between the post-COVID syndrome (17.01±5.22sec) and the healthy participants (14.63±5.38sec). The effect size was moderate (Cohen’s d≈0.45, 95%CI:–0.56 to 1.47).

Discussion

The main findings are summarised as follows: i) PCA indicated that St*i*O₂ was the most sensitive variable for distinguishing between populations, while nTHI was the best indicator for differentiating across activity states. ii) The learning models demonstrated average to excellent performance in classifying muscle oxygenation time-series data as healthy or pathological. Among these, the CIF and CNN models with feature set D emerging as the most successful with excellent performance. iii) Traditional physiological statistical methods, applied at the population level, may not capture subtle individual-level variations.

PCA

This dimensionality reduction approach condensed the high-frequency time-series data from four channels into a single data point in the PC1-PC2 space, summarising complex data. The PCA results for St*i*O₂ reveal clear separation between the two population groups with minimal overlap, indicating distinct underlying patterns in muscle oxygenation across all activity states. Differences between groups were more pronounced along PC2 than PC1, suggesting the secondary variance component contributes more to population separation. Although the physiological meaning of PCA components is not established, PC1 may reflect global oxygen supply-utilisation balance, while PC2 may capture local perfusion or metabolic fluctuations. nTHI, which resembles blood volume, clustered by period, with higher changes corresponding to higher PC2 values (i.e., exercise)(Figure S8). This interpretation aligns with PCA theory(Bro & Smilde, 2014) and its physiological applications(Muhammad et al., 2013), though further work is needed to define the mechanisms.

Hence, PCA effectively captures intrinsic StiO_2 patterns distinguishing populations, independent of activity state.

The absence of distinct StiO_2 clusters across activity periods indicates stable signal patterns independent of metabolic state. This suggests that population differences are intrinsic to underlying physiology rather than temporary activity effects, supporting StiO_2 as a robust marker for distinguishing groups.

The relevance of StiO_2 in this data-centric approach likely stems from its capacity to measure the proportion of oxygenated to total haemoglobin. This ratio reflects the balance between oxygen supply and utilisation within the tissue, which is influenced by the difference in oxygen consumption (VO_2) and blood flow (Q), as described by the Fick principle ($\text{VO}_2 = Q \times [\text{CaO}_2 - \text{CvO}_2]$) (Fick, 1870). StiO_2 , as a single measure of oxygenated Hb, Mb, and cytochromes, will decrease as CvO_2 falls and increase as CvO_2 rises, making it a reflection of the VO_2/Q ratio, indicating the amount of haemoglobin that remains saturated after passing through the tissue and thereby reflecting oxygen delivery and utilisation within the tissue. nTHI clustered by activity states rather than groups because it reflects changes in total haemoglobin volume during different activities rather than underlying physiological differences; thus, it serves as a complementary feature in models sensitive to activity-related variations. No other variables (HHb, O_2Hb) showed separation by population or period (Figures S3-S9). Consequently, this data-centric approach guided the development of models incorporating StiO_2 with period labels (Feature sets A/B) or StiO_2 with nTHI without labels (C/D), integrating both physiology-driven and data-centric principles to produce more robust models.

Learning Models

Nearest neighbour classifiers with DTW have long been a strong baseline for time series classification (Middlehurst et al., 2020). However, recent advances and results from this study show that deep learning (CNN) and tree-based ensemble models (CIF) can surpass DTW-based methods (Bagnall et al., 2017).

CIF calculates numerous statistical features per data interval (e.g., autocorrelation, entropy, fluctuation scaling, outlier detection, means and standard deviations) (Middlehurst et al., 2020) capturing temporal and kinetic characteristics encoding physiological responses effectively. CNNs, in contrast, learn hierarchical features directly from raw data via convolutional filters that detect local patterns like shape changes and peaks (Zhao et al., 2017). KNN classifies by finding the closest time-series examples using DTW to align sequences (Mahato et al., 2018). The key difference is CIF explicitly encodes temporal statistics before classification, whereas CNN and KNN-DTW rely on learned feature representation and direct shape similarity, respectively. This explains CIF's advantage on feature sets with activity labels, where the statistical descriptors plus label context enhance discrimination. Overall, models informed by domain knowledge and PCA perform strongly, but CNNs may offer better adaptability in real-world applications with fewer labelled data.

Feature sets containing StiO_2 and nTHI without activity labels (C/D) outperformed those including StiO_2 with labelled activity states (A/B). This likely reflects the richer dynamic information from nTHI, which captures blood volume changes more precisely than static one-hot labels. The inclusion of one-hot encoded activity labels in A and B may have introduced contextual information unrelated to physiological signal characteristics, which could limit model generalisability. This further enforces the value of models based solely on physiological input (C/D). Furthermore, data from the first 60 seconds of each period (B/D) yielded better results than the last 30 seconds (A/C), possibly due to the kinetic changes early in the activity phase versus plateaued oxygenation later (Vogiatzis et al., 2007).

Further signal manipulations or feature construction were not applied, as the signal lacks periodicity (e.g., ECG). Moreover, models incorporating all signal parameters (O_2Hb , HHb, nTHI, StiO_2) were tested, but their performance did not surpass the features selected for feature sets C/D. As a result, these models were not presented, as the additional parameters did not provide meaningful information.

Traditional Statistical Approaches

Participants with post-COVID syndrome reported considerable fatigue and dyspnoea symptoms in their daily lives and they experienced impaired exercise tolerance compared to their healthy counterparts. There was very little heterogeneity in the StiO_2 and no change from rest to exercise indicating tight matching between $\dot{Q}$ and $\dot{\text{VO}}_2$ across progressively increasing levels of exercise in both post-COVID syndrome and healthy people. Importantly, the coefficient of variation of StiO_2 was not different between post-COVID syndrome ($\text{CV}=0.06$) and healthy participants ($\text{CV}=0.07$) across time points. Moreover, the average CV of StiO_2 ($0.06\text{--}0.08$) in the present study is comparable to these previously reported in healthy individuals (Vogiatzis et al., 2015) and in patients with COPD (Louvaris et al., 2017) indicating tight regulation of oxygen supply and demand. Additionally, during constant-load exercise at 80% peak work rate the mean response time (MRT) values StiO_2 were not significantly different between groups, albeit the post-COVID syndrome population exhibited a slower StiO_2 response compared to healthy participants. MRT values in this study are in good agreement with those in the literature for both post-COVID (Luz Goulart et al., 2024) and healthy participants (Cipriano et al., 2024). However, the moderate effect size (Cohen's $d \approx 0.45$) and wide confidence interval spanning both negative and positive values, combined with the small sample size, suggest caution in interpreting this non-significant p-value. Furthermore, there was no significant interaction between population and time point in the mean StiO_2 , indicating that the effect of time point on StiO_2 does not differ by population (Figure S12). Logistic regression models with StiO_2 as the independent variable failed to distinguish between population across different time points (online supplement p.15).

Applications

These learning models show promise for clinical and laboratory use, enabling differentiation between healthy and pathological muscle responses beyond what standard approaches may capture. Models require only short NIRS segments (30-60 seconds) to classify responses. Feature sets without time labels (C and D) reduce the need for manual labelling, broadening applicability across different protocols and real-world monitoring, including wearables. Standardising features makes models software-agnostic, improving compatibility across NIRS systems (Ishwaran & O'Brien, 2022). With further refinement, these models could objectively assess recovery likelihood and trajectory.

Novelty

To the best of our knowledge, this study is the first to apply machine learning models to classify healthy versus symptomatic muscle responses across activity states using time-series NIRS data. Model and feature selection were guided by physiological, clinical, and PCA-informed criteria. This demonstrates the potential of machine learning to identify pathological states at the individual level. Furthermore, the present study employed an annotated medical dataset, which despite being increasingly difficult and costly to obtain due to the complexity of health data and the challenges of labelling (Abbaspourazad et al., 2023), it could provide valuable insights into various clinical and physiological variables and their interactions. Existing approaches include self-supervised learning techniques applied to other biosignals in large, non-annotated datasets from real-world settings (Abbaspourazad et al., 2023) or general foundation models (Pillai et al., 2024). Prior research utilised large datasets for general tasks relevant to overall well-being and population monitoring, presenting model-centred approaches (Abbaspourazad et al., 2023; Pillai et al., 2024). However, these approaches might have limited integration of domain-specific knowledge, which can present challenges in effectively informing highly targeted analyses. We used this annotated dataset to develop data-centric models that identify distinct physiology patterns. Success relied on precise preprocessing, segmentation, and labelling (Jakubik et al., 2024). The methodology and publicly available model weights and pipelines provide a foundation for further research, refinement, and retraining (Megaritis, 2025).

Limitations

Given the limited number of participants, all previously hospitalised patients from the UK with specific types of COVID-19, there is a risk of overfitting, particularly for complex models such as CNNs. To mitigate this, we employed grouped 5-fold cross-validation, ensuring that all time-series data from each participant were assigned exclusively to either the training or test set in each fold. This approach prevents information leakage between individuals, provides performance metrics on unseen participants, and offers a more realistic estimate of model performance. Future work would benefit from retraining the model on larger and more diverse datasets to improve generalisability. It should also be noted that NIRS-derived physiological variables reflect composite measures of tissue oxygenation and haemodynamics, influenced by arterial, capillary, and venous haemoglobin, as well as myoglobin and cytochrome oxidase, which current methodologies do not fully separate.

Conclusions

We demonstrated that AI models trained on time-series muscle oxygenation data can effectively classify responses to distinguish between populations. A data-centric PCA approach identified StIO₂ (and nTHI) as the most sensitive physiological features, guiding model development. Among the models tested, CIF and CNN performed best. By integrating domain-specific physiological interpretation with data-centric AI, this study provides a framework for developing learning models from biosignals. Applied to larger, diverse datasets, these models could enable scalable, objective muscle-response assessment and passive monitoring.

Declaration of Interest

D. Megaritis has nothing to disclose.

E. Hume has nothing to disclose.

E. Daynes reports consulting fees from the Royal College of Physicians, lecture fees from Fisher and Pykell, and community outreach payments from Chiesi.

R. Evans has nothing to disclose.

Y. Zeng has nothing to disclose.

S. J. Singh has nothing to disclose.

C. Echevarria has nothing to disclose.

P. D. Wagner was a paid consultant for SMS Biotech and may hold shares. Received travel support from the University of Utah and from the University of Texas Southwestern.

I. Vogiatzis has nothing to disclose.

Data Availability Statement: The model architectures, training pipelines, and usage instructions are openly accessible via the GitHub repository: https://github.com/DMegaritis/BioSigClass_NIRS. The raw data used to train these models can be made available by the corresponding author upon reasonable request and subject to ethical approval.

References

1. Abbaspourazad, S., Elachqar, O., Miller, A. C., Emrani, S., Nallasamy, U., & Shapiro, I. (2023). Large-scale Training of Foundation Models for Wearable Biosignals. *ArXiv, abs/2312.05409*.
2. Appelman, B., Charlton, B. T., Goulding, R. P., Kerkhoff, T. J., Breedveld, E. A., Noort, W., Offringa, C., Bloemers, F. W., van Weeghel, M., Schomakers, B. V., Coelho, P., Posthuma, J. J., Aronica, E., Joost Wiersinga, W., van Vugt, M., & Wüst, R. C. I. (2024). Muscle abnormalities worsen after post-exertional malaise in long COVID. *Nature Communications*, 15(1), 17. <https://doi.org/10.1038/s41467-023-44432-3>
3. Bagnall, A., Lines, J., Bostrom, A., Large, J., & Keogh, E. (2017). The great time series classification bake off: a review and experimental evaluation of recent algorithmic advances. *Data Mining and Knowledge Discovery*, 31(3), 606-660. <https://doi.org/10.1007/s10618-016-0483-9>
4. Borg, G. A. (1982). Psychophysical bases of perceived exertion. *Medicine and Science in Sports and Exercise*, 14(5), 377-381.
5. Bro, R., & Smilde, A. K. (2014). Principal component analysis [10.1039/C3AY41907J]. *Analytical Methods*, 6(9), 2812-2831. <https://doi.org/10.1039/C3AY41907J>
6. Brown, C., Chauhan, J., Grammenos, A., Han, J., Hasthanasombat, A., Spathis, D., Xia, T., Cicuta, P., & Mascolo, C. (2020). *Exploring Automatic Diagnosis of COVID-19 from Crowdsourced Respiratory Sound Data* Proceedings of the 26th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining, Virtual Event, CA, USA. <https://doi.org/10.1145/3394486.3412865>
7. Cipriano, G., da Luz Goulart, C., Chiappa, G. R., da Silva, M. L., Silva, N. T., do Vale Lira, A. O., Negrão, E. M., DÁvila, L. B. O., Ramalho, S. H. R., de Souza, F. S. J., Cipriano, G. F. B., Hirai, D., Hansen, D., & Cahalin, L. P. (2024). Differential impacts of body composition on oxygen kinetics and exercise tolerance of HFrEF and HFpEF patients. *Scientific Reports*, 14(1), 22505. <https://doi.org/10.1038/s41598-024-72965-0>
8. Daynes, E., Baldwin, M., Greening, N. J., Yates, T., Bishop, N. C., Mills, G., Roberts, M., Hamrouni, M., Plekhanova, T., Vogiatzis, I., Echevarria, C., Nathu, R., McAuley, H. J. C., Latimer, L., Glennie, J., Chambers, F., Penfold, R., Hume, E., Magaritis, D.,...Singh, S. J. (2023). The effect of COVID rehabilitation for ongoing symptoms Post HOSPitalisation with COVID-19 (PHOSP-R): protocol for a randomised parallel group controlled trial on behalf of the PHOSP consortium. *Trials*, 24(1), 61. <https://doi.org/10.1186/s13063-023-07093-7>
9. Esteva, A., Robicquet, A., Ramsundar, B., Kuleshov, V., DePristo, M., Chou, K., Cui, C., Corrado, G., Thrun, S., & Dean, J. (2019). A guide to deep learning in healthcare. *Nature Medicine*, 25(1), 24-29. <https://doi.org/10.1038/s41591-018-0316-z>
10. Fick, A. (1870). Ueber die Messung des Blutquantum in den Herzventrikeln. *Sb Phys Med Ges Worzburg*, 16-17.
11. Ishwaran, H., & O'Brien, R. (2022). The Standardization And Automation Of Machine Learning For Biomedical Data. *Journal of Thoracic and Cardiovascular Surgery*, 163(1), e102-e103. <https://doi.org/10.1016/j.jtcvs.2020.07.113>
12. Jakubik, J., Vössing, M., Kühl, N., Walk, J., & Satzger, G. (2024). Data-Centric Artificial Intelligence. *Business & Information Systems Engineering*, 66(4), 507-515. <https://doi.org/10.1007/s12599-024-00857-8>
13. Kapoor, T., Pandhi, T., & Gupta, B. (2023). Cough Audio Analysis for COVID-19 Diagnosis. *SN Comput Sci*, 4(2), 125. <https://doi.org/10.1007/s42979-022-01522-1>
14. Louvaris, Z., Habazettl, H., Asimakos, A., Wagner, H., Zakynthinos, S., Wagner, P. D., & Vogiatzis, I. (2017). Heterogeneity of blood flow and metabolism during exercise in patients with chronic obstructive pulmonary disease. *Respiratory Physiology & Neurobiology*, 237, 42-50. <https://doi.org/10.1016/j.resp.2016.12.013>
15. Luz Goulart, C., Robson Fernando Borges, R. F. B., Claudia Cristina Conde Holanda Sobral, C. C. C. H. S., Thomas Beltrame, T. B., Mauricio Milani, M. M., Gerson Cipriano Junior, G. C. J., Ana Cleides Pereira Dos Santos, A. C. P. S., Ricardo Stein, R. S., Fabricio Braga, F. B., & Luiz Eduardo Fonteles Ritt, L. E. F. R. (2024). Muscle oxygenation kinetics in Long COVID-19: Illness severity and clinical implications. *European Journal of Preventive Cardiology*, 31(Supplement_1), zwae175.402. <https://doi.org/10.1093/eurjpc/zwae175.402>
16. Mahato, V., O'Reilly, M., & Cunningham, P. (2018). *A Comparison of k-NN Methods for Time Series Classification and Regression*.

17. Malhotra, A. G., Borkar, P., Chowdhary, R., & Singh, S. (2023). Combating COVID-19 by employing machine learning predictions and projections. In K. Pal, S. Ari, A. Bit, & S. Bhattacharyya (Eds.), *Advanced Methods in Biomedical Signal Processing and Analysis* (pp. 175-203). Academic Press. <https://doi.org/https://doi.org/10.1016/B978-0-323-85955-4.00003-X>
18. Megaritis, D. (2025). *BioSigClass_NIRS*. In Zenodo.
19. Middlehurst, M., Large, J., & Bagnall, A. (2020, 10-13 Dec. 2020). The Canonical Interval Forest (CIF) Classifier for Time Series Classification. 2020 IEEE International Conference on Big Data (Big Data),
20. Modi, B., Sharma, M., Hemani, H., Joshi, H., Kumar, P., Narayanan, S., & Shah, R. (2024). Analysis of Vocal Signatures of COVID-19 in Cough Sounds: A Newer Diagnostic Approach Using Artificial Intelligence. *Cureus*, 16(3), e56412. <https://doi.org/10.7759/cureus.56412>
21. Muhammad, Y., Krivoshei, A., & Annus, P. (2013). Separation of cardiac and respiratory components from the electrical bio-impedance signal using PCA and fast ICA.
22. Pillai, A., Spathis, D., Kawsar, F., & Malekzadeh, M. (2024). *PaPaGei: Open Foundation Models for Optical Physiological Signals*. <https://doi.org/10.48550/arXiv.2410.20542>
23. Singh, S. J., Morgan, M. D., Scott, S., Walters, D., & Hardman, A. E. (1992). Development of a shuttle walking test of disability in patients with chronic airways obstruction. *Thorax*, 47(12), 1019-1024. <https://doi.org/10.1136/thx.47.12.1019>
24. Sivan, M., & Taylor, S. (2020). NICE guideline on long covid. *BMJ*, 371, m4938. <https://doi.org/10.1136/bmj.m4938>
25. Souza, V., Lafeta, M., Saldanha, M., Penido, F., Menezes, T., Tanni, S., Albuquerque, A., Nery, L., Ota-Arakaki, J., Ferreira, E., & Sperandio, P. Dynamic matching of oxygen uptake kinetics and muscle deoxygenation in post-COVID-19. *European Respiratory Journal*, 60(suppl 66), 2212. <https://doi.org/10.1183/13993003.congress-2022.2212>
26. Tavenard, R., Faouzi, J., Vandewiele, G., Divo, F., Androz, G., Holtz, C., Payne, M., Yurchak, R., Rußwurm, M., Kolar, K., & Woods, E. (2020). Tslearn, A Machine Learning Toolkit for Time Series Data.
27. Tuesta, M., Yáñez-Sepúlveda, R., Verdugo-Marchese, H., Mateluna, C., & Alvear-Ordenes, I. (2022). Near-Infrared Spectroscopy Used to Assess Physiological Muscle Adaptations in Exercise Clinical Trials: A Systematic Review. *Biology (Basel)*, 11(7). <https://doi.org/10.3390/biology11071073>
28. van de Poppe, D.-J., Hulzebos, H. J., & Takken, T. (2018). Reference values for maximum work rate in apparently healthy Dutch/Flemish adults: data from the LowLands fitness registry. *Acta Cardiologica*, 74, 1-8. <https://doi.org/10.1080/00015385.2018.1478763>
29. Vogiatzis, I., Habazettl, H., Louvaris, Z., Andrianopoulos, V., Wagner, H., Zakynthinos, S., & Wagner, P. D. (2015). A method for assessing heterogeneity of blood flow and metabolism in exercising normal human muscle by near-infrared spectroscopy. *Journal of applied physiology (Bethesda, Md. : 1985)*, 118(6), 783-793. <https://doi.org/10.1152/japplphysiol.00458.2014>
30. Vogiatzis, I., Zakynthinos, S., Georgiadou, O., Golemati, S., Pedotti, A., Macklem, P. T., Roussos, C., & Aliverti, A. (2007). Oxygen kinetics and debt during recovery from expiratory flow-limited exercise in healthy humans. *European Journal of Applied Physiology*, 99(3), 265-274. <https://doi.org/10.1007/s00421-006-0342-2>
31. Webster, K., Cella, D., & Yost, K. (2003). The Functional Assessment of Chronic Illness Therapy (FACIT) Measurement System: properties, applications, and interpretation. *Health Qual Life Outcomes*, 1, 79. <https://doi.org/10.1186/1477-7525-1-79>
32. Yorke, J., & Armstrong, I. (2014). The assessment of breathlessness in pulmonary arterial hypertension: Reliability and validity of the Dyspnoea-12. *European Journal of Cardiovascular Nursing*, 13(6), 506-514. <https://doi.org/10.1177/1474515113514891>
33. Yorke, J., Khan, N., Garrow, A., Tyson, S., Singh, D., Vestbo, J., & Jones, P. W. (2022). Evaluation of the Individual Activity Descriptors of the mMRC Breathlessness Scale: A Mixed Method Study. *International Journal of Chronic Obstructive Pulmonary Disease*, 17, 2289-2299. <https://doi.org/10.2147/copd.S372318>
34. Zhao, B., Lu, H., Chen, S., Liu, J., & Wu, D. (2017). Convolutional neural networks for time series classification. *Journal of Systems Engineering and Electronics*, 28(1), 162-169. <https://doi.org/10.21629/JSEE.2017.01.18>

