Supplementary Information

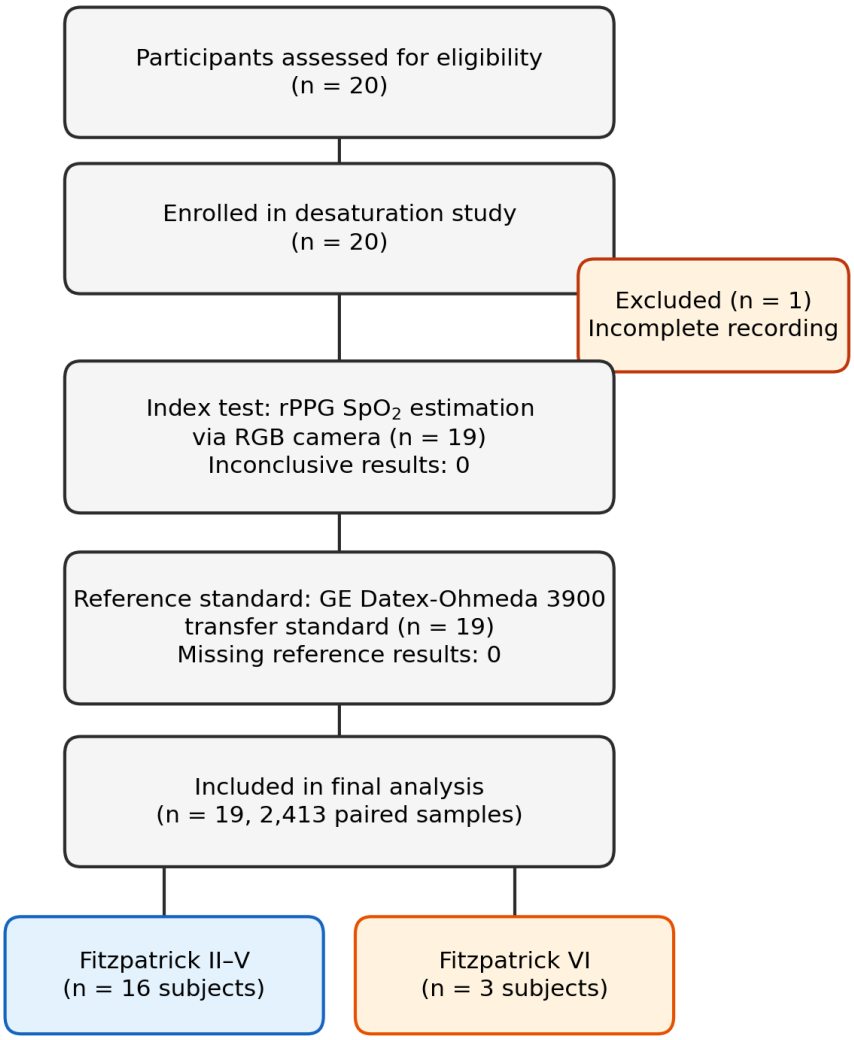
Contactless Human Oxygen Saturation Detection Using RGB Camera in an Induced Hypoxemia Study with Varied Skin Types

Li Zhu, Migyeong Gwak, Korosh Vatanparvar, Qijia Shao, Sharanya Desai, Jungmok Bae, Jilong Kuang, Alex Gao

The following supplementary materials accompany the main manuscript:

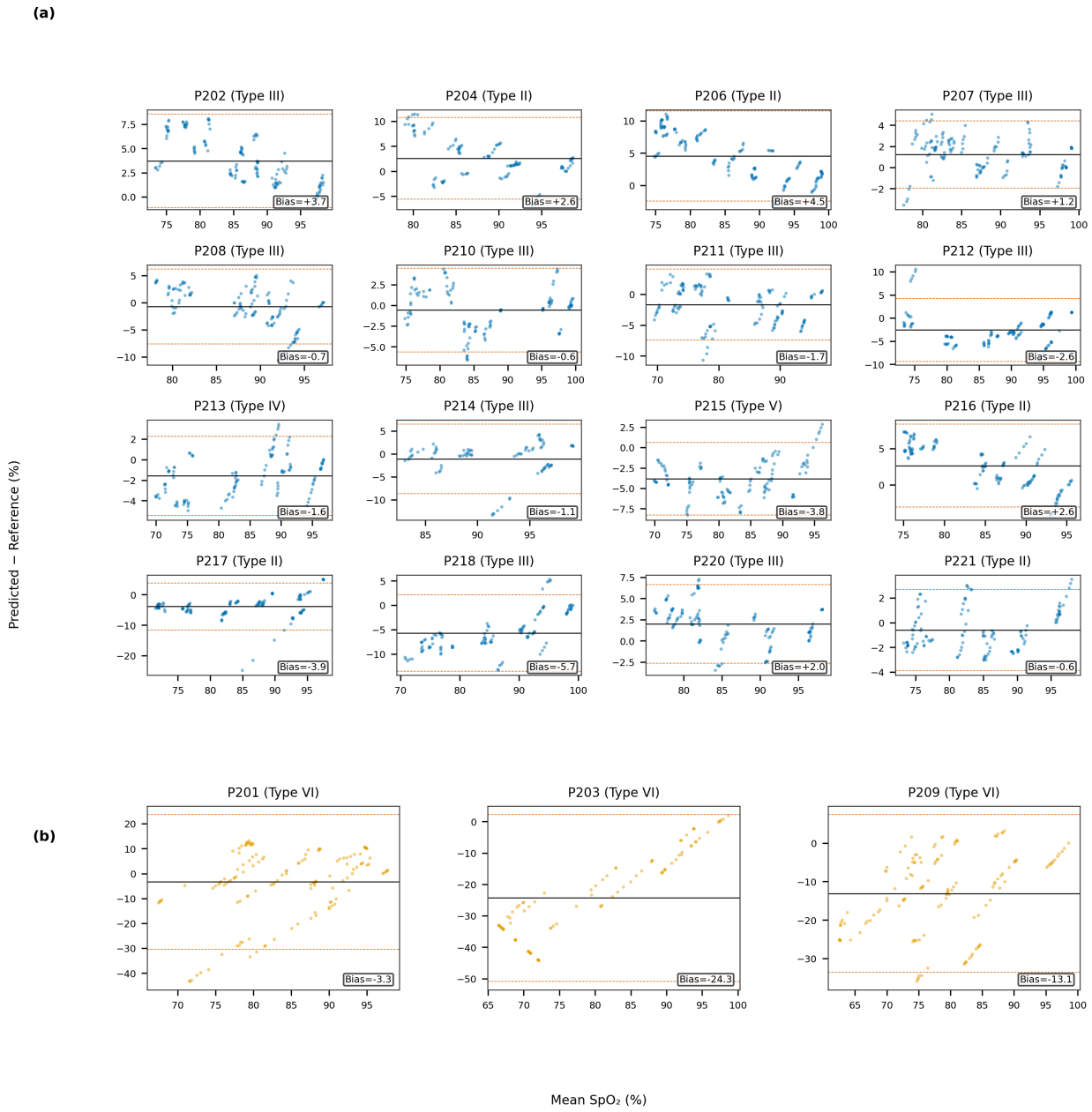
- **Supplementary Figure S1:** STARD 2015 participant flow diagram showing enrollment, exclusions, and analysis subgroups.
- **Supplementary Figure S2:** Per-subject Bland-Altman plots from the Type VI exclusion experiment.
- **Supplementary Figure S3:** Feature ablation analysis comparing $\mathrm{RoR}_{\mathrm{RB}}$-only, $\mathrm{RoR}_{\mathrm{RG}}$-only, and combined feature configurations.
- **Supplementary Figure S4:** Lighting variation effects on RoR–SpO_2 correlation across ambient lighting conditions.
- **Supplementary Figure S5:** Signal extraction detail showing per-ROI signal quality variation and multi-ROI fusion.
- **Supplementary Table S1:** Extended per-subject performance metrics.
- **Supplementary Table S2:** Bootstrap confidence intervals for aggregate performance metrics.
- **Supplementary Table S3:** TRIPOD+AI compliance checklist with item-by-item manuscript section references.

Supplementary Figure S1: STARD Flow Diagram



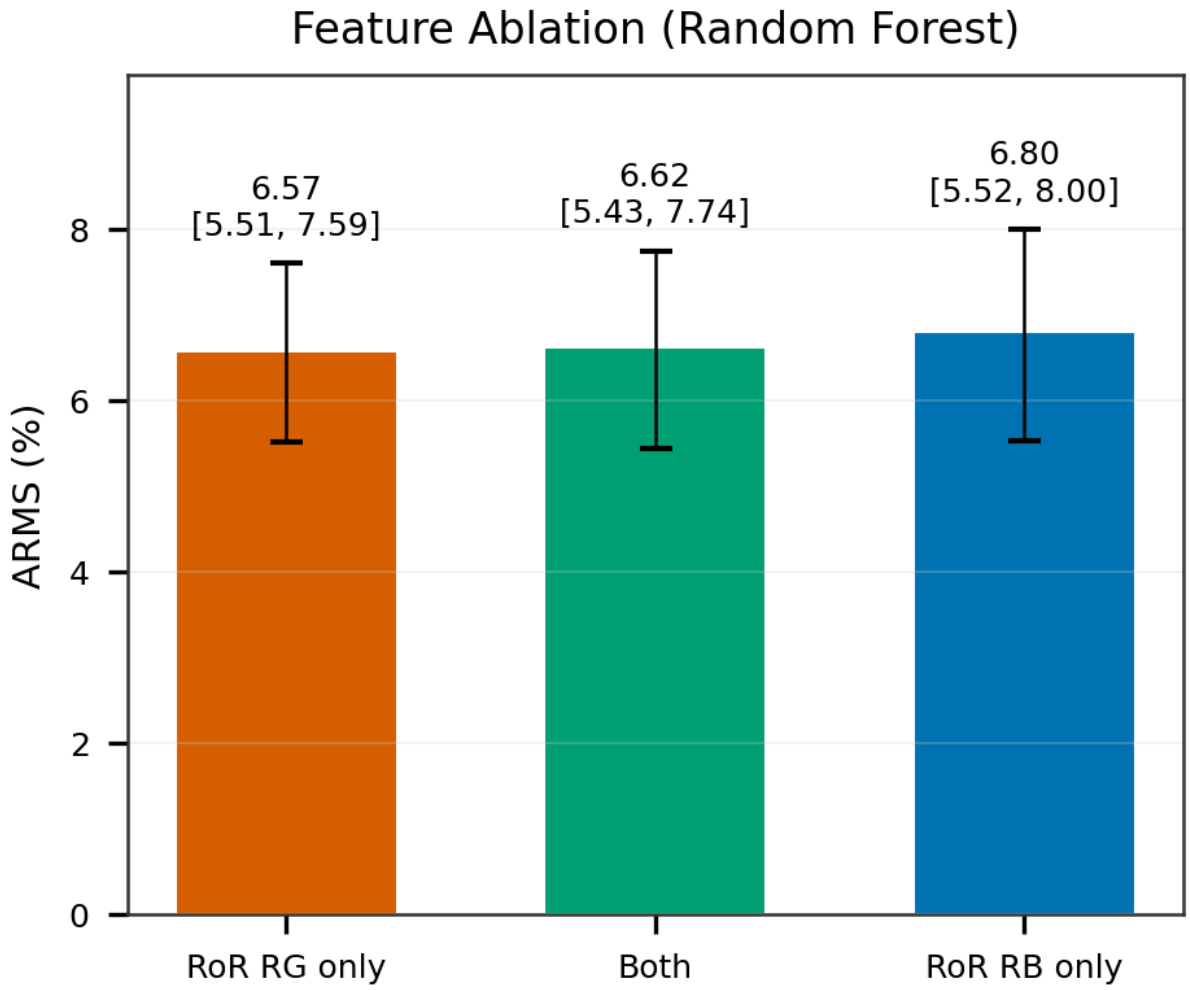
Supplementary Figure 1: **STARD 2015 participant flow diagram.** Twenty participants were assessed for eligibility, of whom 19 completed the desaturation protocol with both index test (rPPG SpO_2) and reference standard (GE Datex-Ohmeda 3900 transfer standard), yielding 2,413 paired samples. One participant was excluded due to an incomplete recording. The final cohort was stratified into Fitzpatrick Type II–V (n = 16) and Type VI (n = 3) subgroups.

Supplementary Figure S2: Per-Subject Bland-Altman Plots



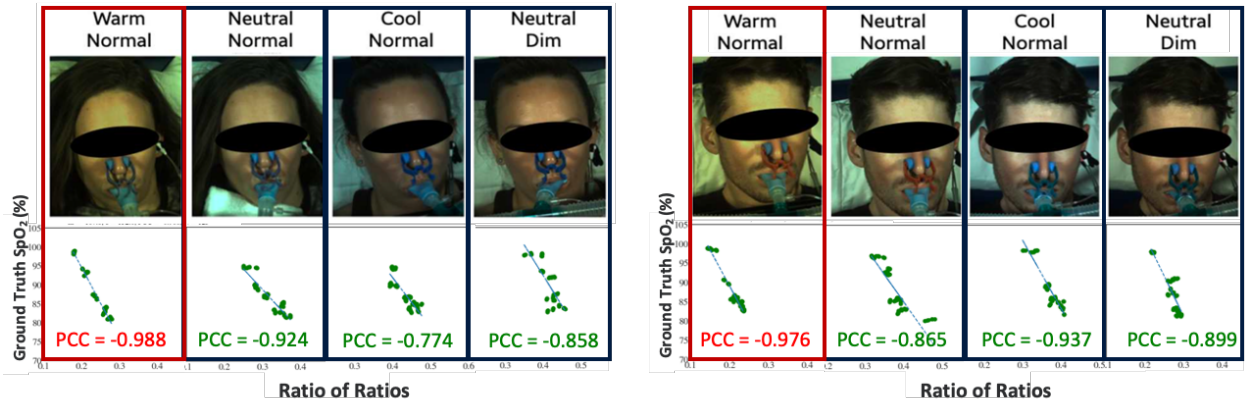
Supplementary Figure 2: **Per-subject Bland-Altman agreement plots (Type VI exclusion experiment).** (a) Individual Bland-Altman plots for 16 Type II–V participants using LOSOCV predictions from a model trained exclusively on Type II–V data (Type VI excluded from all training folds). (b) Three Type VI participants tested against the Type II–V-trained model. Solid line: mean bias; dashed lines: 95% limits of agreement. Blue: Type II–V; orange: Type VI. One-point calibrated predictions used throughout. Diagonal trends in some panels reflect proportional bias at lower SpO_2 levels.

Supplementary Figure S3: Feature Ablation



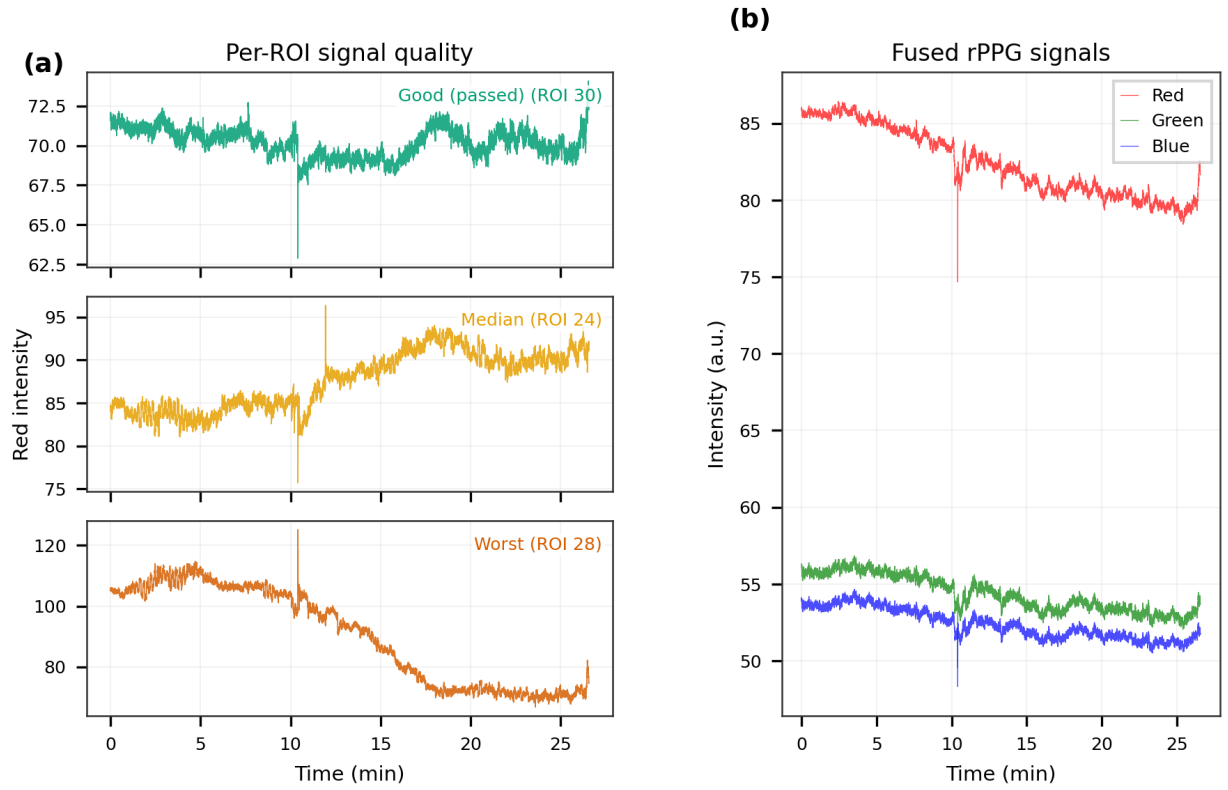
Supplementary Figure 3: **Feature ablation.** A_{RMS} for three feature configurations using the best-performing RF model: RoR_{RB} only (6.80%), RoR_{RG} only (6.57%), and both features combined (6.62%). Confidence intervals are from LOSOCV-based bootstrap within the ablation analysis (distinct from the aggregate BCa cluster bootstrap in Supplementary Table S2). Overlapping intervals indicate no statistically significant difference between single-feature and combined models; the combined model was selected as the primary result because the two-feature input is motivated by the physics-based Beer–Lambert derivation.

Supplementary Figure S4: Lighting Variation



Supplementary Figure 4: **Lighting variation effects on RoR–SpO_2 correlation.** Per-subject RoR–SpO_2 scatter plots under different ambient lighting conditions from our pilot study, demonstrating that changes in illumination affect both the slope and intercept of the RoR–SpO_2 relationship. Despite these variations, the RoR features remain correlated with ground truth SpO_2 ($|\rho| = 0.77$–0.99) across all conditions, confirming the robustness of the physics-based approach. Participants provided written informed consent for the use of facial images in publication.

Supplementary Figure S5: Signal Extraction Detail



Supplementary Figure 5: **Signal extraction detail.** (a) Per-ROI signal quality: representative high-quality (top, green), median (middle, orange), and low-quality (bottom, red) ROIs from the Red channel, illustrating the signal fidelity variation across the superpixel grid. Quality screening excludes noisy ROIs from downstream processing. (b) SNR-weighted multi-ROI fusion showing fused Red, Green, and Blue channel signals after CHROM-based pulse extraction and SNR-weighted fusion.

Supplementary Table S1: Extended Per-Subject Performance Metrics

Supplementary Table 1: Extended per-subject performance metrics

Subject	Skin Type	n	GT Range	ARMS (%)	MAE (%)	RoR PCC	Bias (%)	SD (%)
P201	VI	143	73–97	8.42	7.35	-0.130	-2.06	8.20
P202	III	143	71–98	10.79	9.40	-0.978	+8.93	6.07
P203	VI	102	83–98	10.13	8.08	-0.671	-8.01	6.23
P204	II	111	74–98	5.14	4.39	-0.831	-2.18	4.68
P206	II	152	71–98	5.63	4.75	-0.969	+2.13	5.23
P207	III	140	77–98	2.72	2.20	-0.966	+0.87	2.59
P208	III	125	76–98	5.57	4.62	-0.857	+3.30	4.50
P209	VI	137	73–99	10.59	9.42	-0.547	-0.34	10.63
P210	III	121	75–99	6.51	5.91	-0.953	-4.16	5.03
P211	III	140	70–97	5.01	4.15	-0.941	+2.89	4.10
P212	III	134	70–99	5.21	4.57	-0.859	-2.58	4.54
P213	IV	123	72–97	4.22	3.63	-0.980	+1.13	4.08
P214	III	85	83–98	4.00	3.08	-0.726	-0.28	4.02
P215	V	121	71–95	7.86	6.91	-0.962	-2.03	7.63
P216	II	134	71–95	6.40	5.64	-0.968	+5.35	3.53
P217	II	128	74–97	4.29	3.75	-0.898	+2.39	3.57
P218	III	123	76–99	6.37	5.25	-0.941	-4.24	4.78
P220	III	124	74–96	3.57	3.11	-0.946	-0.77	3.50
P221	II	127	74–96	4.70	4.36	-0.978	-1.56	4.45

A_{RMS}: accuracy root mean square per ISO 80601-2-61 ($= \sqrt{\text{bias}^2 + \text{SD}^2}$). RoR PCC: Pearson correlation between RoR_{RG} feature and reference SpO_2.

Supplementary Table S2: Bootstrap Confidence Intervals

Supplementary Table 2: Bootstrap confidence intervals for aggregate metrics

Metric	CI Level	Estimate	CI Lower	CI Upper	SE	Method
ARMS	0.95	6.62	5.60	7.93	0.59	BCa_cluster
ARMS	0.99	6.62	5.31	8.34	0.59	BCa_cluster
MAE	0.95	5.34	4.51	6.42	0.48	BCa_cluster
MAE	0.99	5.34	4.28	6.78	0.48	BCa_cluster
PCC	0.95	0.595	0.373	0.732	0.087	BCa_cluster
PCC	0.99	0.595	0.296	0.770	0.087	BCa_cluster
AUC (90%)	0.95	0.797	0.669	0.870	0.047	BCa_cluster
AUC (90%)	0.99	0.797	0.609	0.887	0.047	BCa_cluster

All intervals computed using 10,000 BCa cluster-adjusted bootstrap resamples across 19 subjects.

Supplementary Table S3: TRIPOD+AI Compliance Checklist

This checklist documents compliance with the Transparent Reporting of a multivariable prediction model for Individual Prognosis or Diagnosis—Artificial Intelligence (TRIPOD+AI) reporting guideline. All 27 items and their sub-items are listed with manuscript section references. Items not applicable to this study are marked N/A with brief justification.

TRIPOD+AI Compliance Checklist

Item	Section	Checklist Item	Location
1	Title	Identify the study as developing or evaluating the performance of a multivariable prediction model, the target population, and the outcome to be predicted	Title
2	Abstract	See TRIPOD+AI for Abstracts checklist (13 sub-items covering background, objectives, setting, participants, outcome, predictors, sample size, missing data, analysis, results, conclusions, registration, funding)	Abstract
3a	Introduction	Explain the healthcare context (including whether diagnostic or prognostic) and rationale for developing or evaluating the prediction model	Introduction, paragraphs 1–2
3b	Introduction	Describe the target population and the intended purpose of the prediction model in the context of the care pathway	Introduction, paragraphs 2–3
3c	Introduction	Describe any known health inequalities between sociodemographic groups	Introduction, paragraph 3 (skin type equity)
4	Introduction	Specify the study objectives, including whether the study describes the development or validation of a prediction model	Introduction, final paragraph (objectives)
5a	Methods	Describe the sources of data separately for the development and evaluation datasets	Methods > Study Design and Participants
5b	Methods	Specify the dates of the collected participant data, including start and end of participant accrual	Methods > Study Design and Participants
6a	Methods	Specify key elements of the study setting including the number and location of centres	Methods > Study Design and Participants
6b	Methods	Describe the eligibility criteria for study participants	Methods > Study Design and Participants

Continued on next page

Table continued from previous page

Item	Section	Checklist Item	Location
6c	Methods	Give details of any treatments received, and how they were handled during model development or evaluation	N/A — no therapeutic intervention; induced hypoxemia is the study protocol, not a treatment
7	Methods	Describe any data pre-processing and quality checking, including whether this was similar across sociodemographic groups	Methods > Signal Extraction Pipeline (quality filters, step removal, CHROM processing)
8a	Methods	Clearly define the outcome that is being predicted and the time horizon, including how and when assessed	Methods > Reference Standard (SpO_2 from pulse oximeter)
8b	Methods	If outcome assessment requires subjective interpretation, describe the qualifications and demographic characteristics of outcome assessors	N/A — SpO_2 reference is automated pulse oximeter reading, no subjective interpretation
8c	Methods	Report any actions to blind assessment of the outcome to be predicted	N/A — reference device operates independently of camera system
9a	Methods	Describe the choice of initial predictors and any pre-selection of predictors before model building	Methods > Signal Extraction Pipeline; Methods > Ratio-of-Ratios Feature Extraction
9b	Methods	Clearly define all predictors, including how and when they were measured	Methods > Ratio-of-Ratios Feature Extraction (RoR_{RB}, RoR_{RG} defined)
9c	Methods	If predictor measurement requires subjective interpretation, describe the qualifications and demographic characteristics of assessors	N/A — RoR features are algorithmically computed, no subjective interpretation
10	Methods	Explain how the study size was arrived at and justify that it was sufficient to answer the research question	Methods > Sample Size
11	Methods	Describe how missing data were handled. Provide reasons for omitting any data	Methods > Study Design and Participants (exclusion criteria); Results > Participant Flow
12a	Methods	Describe how the data were used in the analysis, including whether the data were partitioned	Methods > Machine Learning Models (LOSOCV description)
12b	Methods	Depending on the type of model, describe how predictors were handled in the analyses	Methods > Machine Learning Models (two RoR features as direct inputs)

Continued on next page

Table continued from previous page

Item	Section	Checklist Item	Location
12c	Methods	Specify the type of model, rationale, all model building steps, and method for internal validation	Methods > Machine Learning Models; Methods > Model Development Transparency
12d	Methods	Describe if and how any heterogeneity in estimates across clusters was handled and quantified	Results > Skin Type Performance (Fitzpatrick stratification); Results > P202 Case Study
12e	Methods	Specify all measures and plots used to evaluate model performance	Methods > Statistical Analysis (A_{RMS}, PCC, AUC, Bland-Altman)
12f	Methods	Describe any model updating arising from the model evaluation	N/A — no model updating performed
12g	Methods	For model evaluation, describe how the model predictions were calculated	Methods > Machine Learning Models (LOSOCV prediction procedure)
13	Methods	If class imbalance methods were used, state why and how this was done	N/A — primary task is regression (continuous SpO_2); binary classification for hypoxemia uses natural prevalence with imbalance-aware evaluation metrics (AUC-ROC, sensitivity, specificity at optimal Youden threshold) rather than accuracy alone
14	Methods	Describe any approaches that were used to address model fairness and their rationale	Methods > Statistical Analysis (stratified analysis); Results > Skin Type Performance; Discussion > Equity
15	Methods	Specify the output of the prediction model and provide details and rationale for any classification	Methods > Machine Learning Models (continuous SpO_2 output); Methods > Statistical Analysis (90% threshold for hypoxemia classification)
16	Methods	Identify any differences between the development and evaluation data in healthcare setting and eligibility criteria	N/A — single dataset with leave-one-subject-out cross-validation
17	Methods	Name the institutional research board or ethics committee that approved the study	Ethics Approval
18a	Open Science	Give the source of funding and the role of the funders for the present study	Acknowledgements

Continued on next page

Table continued from previous page

Item	Section	Checklist Item	Location
18b	Open Science	Declare any conflicts of interest and financial disclosures for all authors	Competing Interests
18c	Open Science	Indicate where the study protocol can be accessed or state that a protocol was not prepared	Methods > Study Design and Participants (Clinimark protocol reference)
18d	Open Science	Provide registration information for the study	N/A — observational diagnostic accuracy study, not a registered clinical trial
18e	Open Science	Provide details of the availability of the study data	Data Availability
18f	Open Science	Provide details of the availability of the analytical code	Code Availability
19	PPI	Provide details of any patient and public involvement during the design, conduct, reporting, interpretation, or dissemination	N/A — no patient or public involvement occurred
20a	Results	Describe the flow of participants through the study, including the number with and without the outcome	Results > Participant Characteristics; Supplementary Figure S1 (STARD flow diagram)
20b	Results	Report the characteristics overall and, where applicable, for each data source or setting	Results > Participant Characteristics (demographics table)
20c	Results	For model evaluation, show a comparison with the development data of the distribution of important predictors	N/A — single dataset, no separate development data
21	Results	Specify the number of participants and outcome events in each analysis	Results > Model Performance (participant counts reported)
22	Results	Provide details of the full prediction model to allow predictions in new individuals	Methods > Model Development Transparency (all hyperparameters, feature definitions, and model specifications reported)
23a	Results	Report model performance estimates with confidence intervals, including for any key subgroups	Results > Model Performance; Results > Skin Type Performance (A_{RMS} with 95% BCa CIs)
23b	Results	If examined, report results of any heterogeneity in model performance across clusters	Results > Skin Type Performance; Results > P202 Case Study
24	Results	Report the results from any model updating, including the updated model and subsequent performance	N/A — no model updating performed

Continued on next page

Table continued from previous page

Item	Section	Checklist Item	Location
25	Discussion	Give an overall interpretation of the main results, including issues of fairness in the context of the objectives	Discussion > Summary; Discussion > Equity
26	Discussion	Discuss any limitations of the study and their effects on any biases, statistical uncertainty, and generalisability	Discussion > Limitations
27a	Discussion	Describe how poor quality or unavailable input data should be assessed and handled when implementing the model	Discussion > Limitations (signal quality, environmental constraints)
27b	Discussion	Specify whether users will be required to interact in the handling of the input data or use of the model	Discussion > Limitations (contactless monitoring context, no user interaction required for prediction)
27c	Discussion	Discuss any next steps for future research, with a specific view to applicability and generalisability	Discussion > Limitations, final paragraph (future directions); Conclusion