Supplementary materials

for

Biphasic Spectral Signatures Enable Non-Linear Glucose Estimation from Smartphone Photoplethysmography

Hanna Kim[1], Wonyoung Lee[1], ByeongSeon Ahn[2], SongHee Park[2], Heeju Lee[1,3], Junyoung Jung[1,3], Juyeon Shin[1,3], Hunyeop Jeong[1], Gyuri Kim[4,7], MD, PhD, Yong-ho Lee[5,8], MD, PhD, Eui Chul Lee[2], PhD, Jee Hang Lee[2,6,*], PhD, Jinwoo Kim[1,3] PhD

[1]HAII Corporation, Seoul, [2]Sangmyung University, Seoul, [3]Yonsei University, Seoul, [4]Samsung Seoul Hospital, Seoul, [5]Yonsei Sinchon Severance Hospital, Seoul, [6]Institute for Advanced Intelligence Study, Daejeon, South Korea

Correspondence: Jee Hang Lee (jeehang@smu.ac.kr; jeehang.iais@gmail.com)

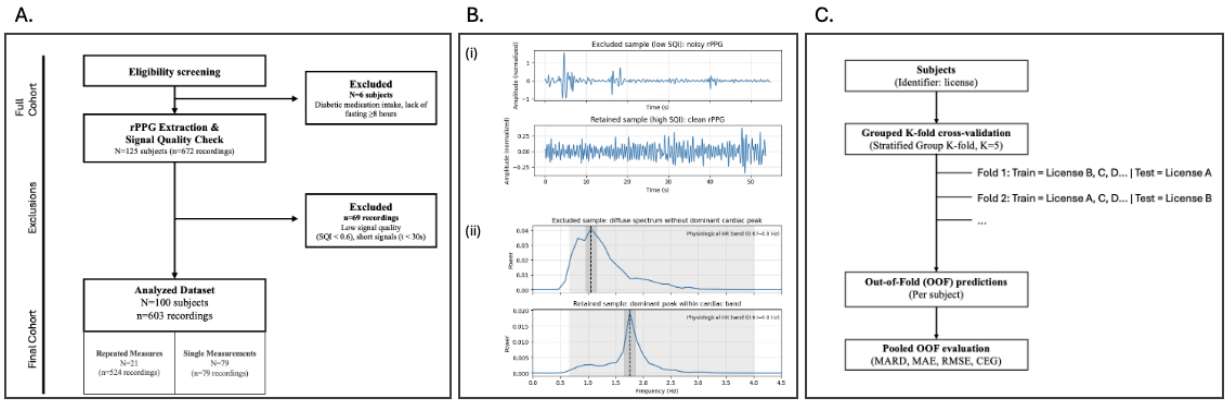

Figure S1. Study design, signal quality control, and subject-independent evaluation. (A) Cohort Flow Diagram. rPPG signals derived from facial video recordings were screened for protocol adherence and signal quality prior to inclusion. Recordings failing rPPG extraction or falling below a predefined signal quality index (SQI) threshold were excluded at the recording level. Only recordings with valid lengths (t ≥ 30 seconds) and those with matched fasting plasma glucose(FPG) were retained. The final cohort for analysis comprised 603 recordings from 100 unique subjects (licenses), and all downstream analyses were conducted using subject-independent evaluation. **(B)** rPPG extraction and signal quality control. **(i)** Representative time-domain rPPG waveforms from an excluded recording (low SQI) and a retained recording (high SQI), which illustrate noisy and clean rPPG signals for comparison. Waveforms are z-score normalized for visualization. (ii) Corresponding power spectral density (PSD) plots which demonstrate the absence of a dominant cardiac-frequency peak in noisy recordings and an identifiable dominant peak within the physiological heart-rate band (0.67–4.0 Hz) in clean recordings. Shaded regions highlight the heart-rate band and the signal band used for the SQI calculation. **(C)** Subject-independent cross-validation and evaluation. Model performance was assessed via repeated stratified group k-fold cross-validation. To ensure subject-independent evaluation, metrics were computed at the subject (license) level. All recordings from a given subject were assigned exclusively to either the training or test set within each fold, ensuring no leakage. Performance metrics were computed using pooled out-of-fold predictions ensuring that each subject was evaluated using data in the test set.

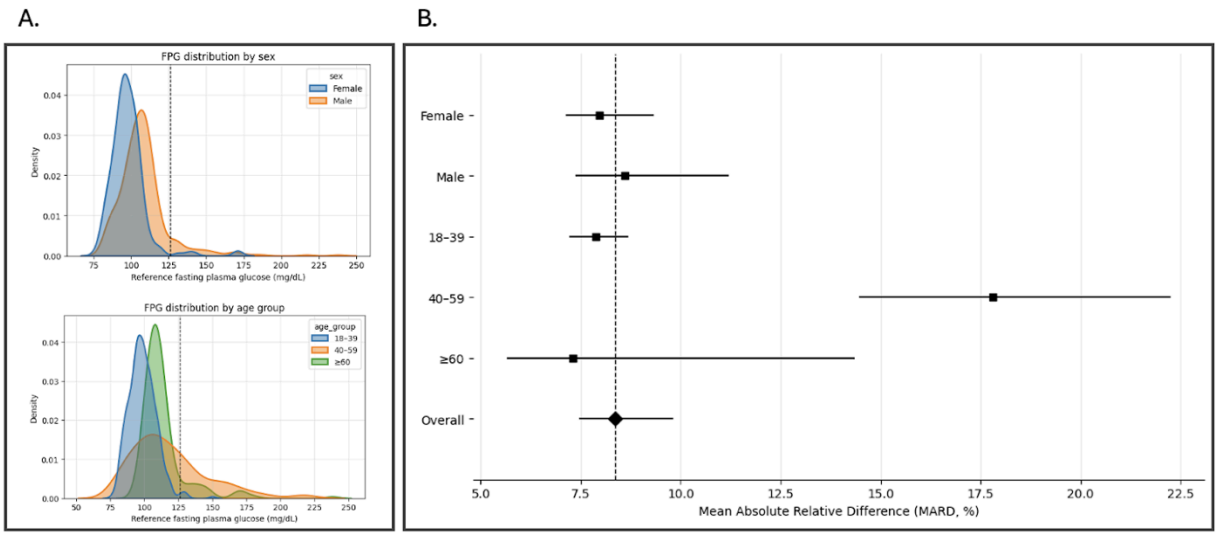

Figure S2. Distribution of fasting plasma glucose and subgroup model performance. (A) Kernel density estimates of reference fasting plasma glucose (FPG) values stratified by sex (top) and age group (bottom). The vertical dashed line indicates the diagnostic threshold for diabetes (126 mg/dL). **(B)** Mean absolute relative difference (MARD) stratified by sex and age group. Points indicate subgroup mean MARD values, with horizontal error bars representing 95% confidence intervals estimated via subject-level bootstrap resampling. The dashed vertical line denotes the overall cohort mean.

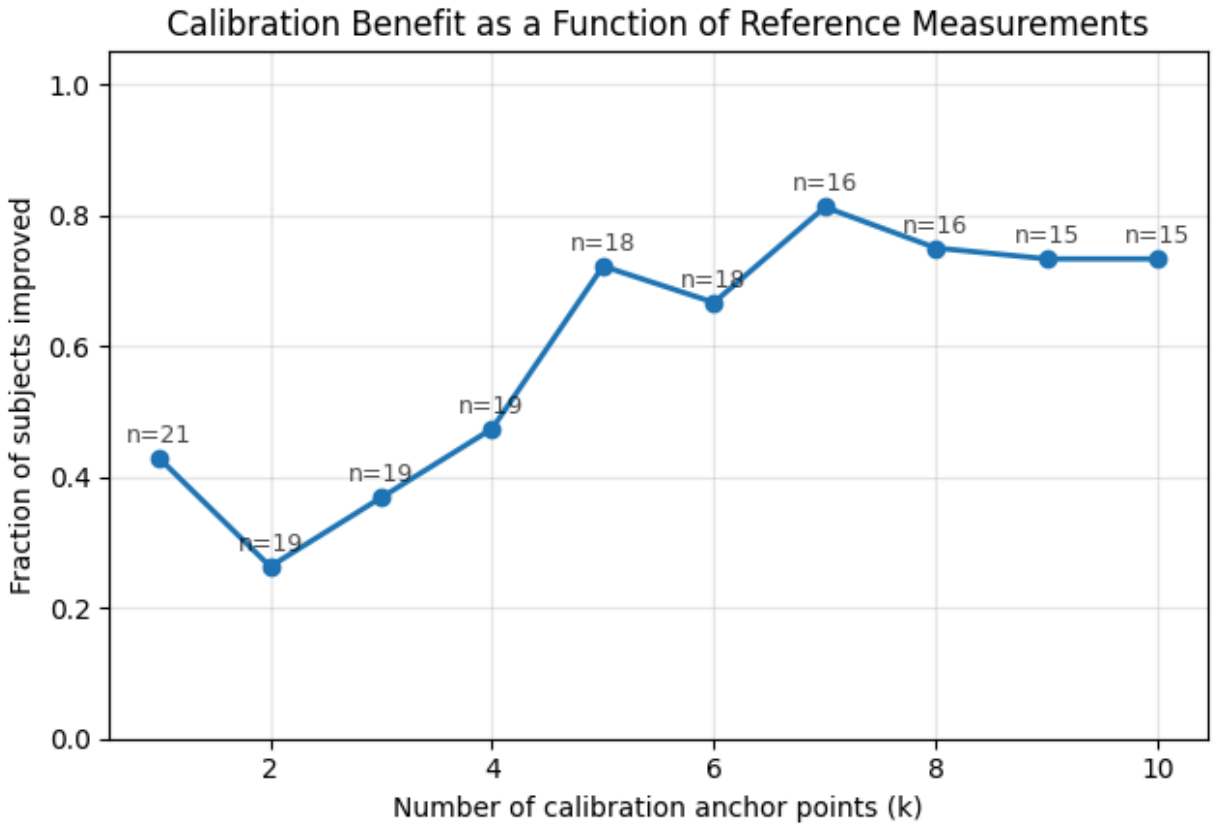

Figure S3. **Dependence of subject-level calibration benefit on the number of reference fasting glucose measurements.**

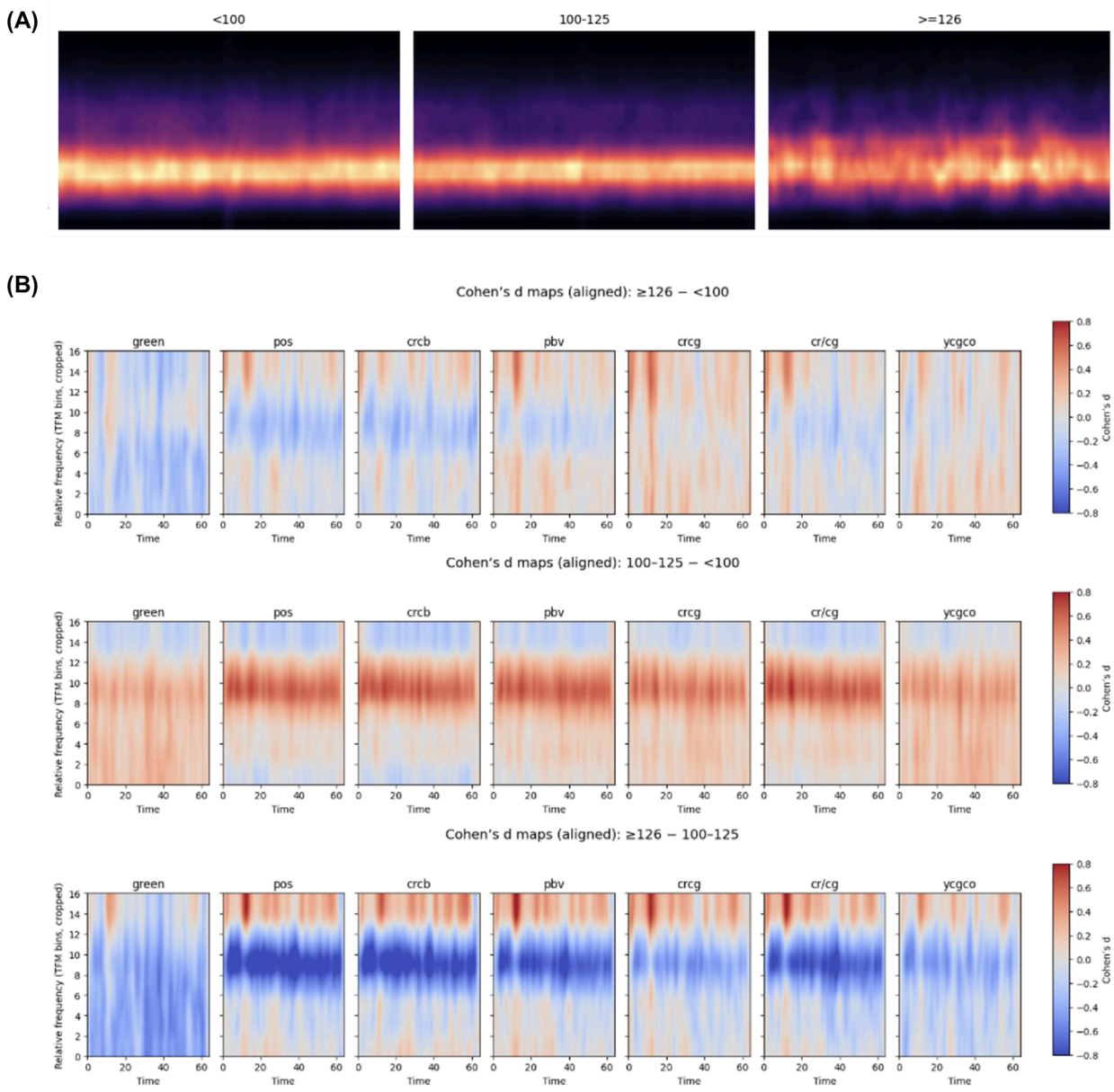

Figure S4. Time–frequency patterns associated with glucose strata. (A) Average time–frequency maps (TFMs) computed from rPPG signals for recordings stratified by reference fasting plasma glucose (FPG) range (<100 mg/dL, 100–125 mg/dL, and ≥126 mg/dL). TFMs are aligned in time and frequency to facilitate visual comparison across glucose strata. **(B)** Cohen's *d* maps highlighting differences in TFM intensity between glucose ranges across multiple rPPG-derived channels. Rows correspond to pairwise contrasts between glucose strata (≥126 vs <100, 100–125 vs <100, and ≥126 vs 100–125), and columns correspond to different color-space and rPPG extraction methods. Positive values (red) indicate higher relative power in

the higher-glucose group, whereas negative values (blue) indicate lower relative power. Effect size patterns are shown for descriptive purposes to localize spectrotemporal regions contributing to model discrimination.

Table S1. Data Demographic Summary

Cohort	Total count	Sessions	Mean Age (SD)	Sex (M/F %)	Mean FPG [95% CI] (mg/dL)	FPG <100 mg/dL (%)	FPG 100–125 mg/dL (%)	FPG≥126 mg/dL (%)
All data	72	688	43.66 ± 13.92	71.51 / 28.49	104.10 [102.84–105.36]	38.08	57.99	3.92
Train (General model)	65	378	41.50 ± 13.45	67.72 / 32.28	102.10 [100.53–103.66]	46.56	49.47	3.97
Calibration cohort	7	310	46.29 ± 14.05	76.13 / 23.87	106.55 [104.54–108.56]	27.74	68.39	3.87

Table S2. **Subgroup analysis results**

Subgroup		Subjects (Recordings)	Mean FPG (mg/dL)	SD FPG (mg/dL)	Median FPG (mg/dL)	Min–Max FPG (mg/dL)
Sex	Female	34 (744)	97.7	12.1	96	76–172
	Male	66 (1,668)	107.9	17.7	106	79–238
Age	18–39	37 (1,464)	98.5	9.9	98	76–150
	40–59	31 (196)	118.7	28.5	112	81–217
	≥60	32 (752)	113.4	17.5	109	93–238

Table S3. Performance against FDA accuracy criteria for glucose monitoring systems

Reference Glucose Range (mg/dL)	Accuracy Criterion	N (in this work)	Within Tolerance % [95% CI]	Lower 95% Bound (Wilson)	FDA Threshold (%)
< 70	±15 mg/dL	0	N/A	N/A	85
70 – 180	±15%	2396	85.6 [78.0–90.4]	84.2	70
> 180	±15%	16	0.0 [0.0–0.0]	0.0	80

Table S4. Clarke Error Grid (CEG) Zone Distribution

Clarke Error Grid Zone	Proportion of Samples (%)
Zone A	92.7
Zone B	7.3
Zone C	0.0
Zone D	0.0
Zone E	0.0
Zone A+B	**100.0**

Table S5. Out-of-fold (OOF) performance with additive demographic features

Model Configuration	RMSE <100 (mg/dL)	RMSE 100–125 (mg/dL)	%Diff ≥126 (%)	MARD (%)	MAE (mg/dL)
TFM only	10.01	8.17	30.27	8.99	10.18
TFM + Age	8.14	8.77	23.39	7.94	8.98
TFM + Age + Gender (FULL)	9.31	7.75	22.25	7.91	8.80

Table S6. Channel SQI results

rPPG Channel	Mean SQI ± SD	Median SQI (IQR)
CrCb	3.97 ± 4.59	2.17 (0.86–5.15)
Green	0.89 ± 0.73	0.63 (0.47–1.04)
POS	4.13 ± 4.51	2.58 (0.95–5.50)
PBV	2.85 ± 3.18	1.53 (0.70–3.85)
CrCg	1.23 ± 1.38	0.65 (0.44–1.26)
Cr/Cg	2.77 ± 3.18	1.41 (0.67–3.35)
YCGCO	0.91 ± 0.84	0.59 (0.41–1.01)

Table S7. Calibration Model. Fraction of subjects exhibiting improved accuracy as a function of the number of reference fasting glucose measurements.

k	Fraction of subjects improved
1	0.43
2	0.26
3	0.37
4	0.47
5	0.72
6	0.67
7	0.81
8	0.75
9	0.73
10	0.73

Table S8. Impact of subject-specific affine calibration over time. Results are stratified by elapsed time since the last calibration point. Values are reported as mean with 95% confidence intervals, estimated using subject-level bootstrap resampling. Calibration measurements are disjoint from evaluation data and are not used in the primary performance analyses reported in the main manuscript.

Time since last calibration	Subjects	MARD (uncalibrated) [CI] (%)	MARD (calibrated; τ=30d) [CI] (%)	RMSE (uncalibrated) [CI]	RMSE (calibrated; τ=30d) [CI]
0–3 days	16	8.93 [6.89–10.71]	8.01 [4.82–12.89]	9.50 [6.74–12.11]	8.84 [5.74–12.89]
4–7 days	17	9.34 [7.03–11.69]	7.55 [5.41–10.75]	10.08 [7.11–12.82]	7.71 [5.83–10.75]
8–14 days	14	10.19 [6.76–15.22]	7.54 [4.82–15.25]	10.60 [7.21–15.58]	7.54 [5.34–15.25]
15–30 days	12	9.66 [6.84–12.56]	9.98 [6.34–15.85]	11.79 [7.78–15.85]	11.72 [7.78–15.85]
31–60 days	12	9.28 [7.52–11.37]	8.40 [6.65–15.29]	10.06 [7.94–15.29]	8.76 [7.94–15.29]