nature medicine

Supplementary information

https://doi.org/10.1038/s41591-024-03243-9

Artificial intelligence guided screening for cardiomyopathies in an obstetric population: a pragmatic randomized clinical trial

In the format provided by the authors and unedited

SUPPLEMENTAL APPENDIX

Table of Contents

INVESTIGATOR LIST 2

SUPPLEMENTAL FIGURES 3

Figure S1. Receiver operating characteristic for detection of cardiomyopathy (LVEF <50% A and LVEF <40% B.) using a digital stethoscope (maximum prediction) in an obstetric population in Nigeria. 4

Figure S2. Receiver operating characteristic for detection of cardiomyopathy (LVEF <50% A and LVEF <40% B.) using the FDA cleared 12-lead AI-ECG model in an obstetric population in Nigeria. . 5

Figure S3. Receiver operating characteristic for detection of cardiomyopathy (LVEF <50% A and LVEF <40% B.) using the original Mayo Clinic 12-lead AI-ECG model in an obstetric population in Nigeria. 6

Figure S4. Enrollment sites and participants 7

Figure S5. Digital stethoscope and chest recording location. 8

Figure S6. Progression through the study visits stratified by clinical site and pregnancy status 9

SUPPLEMENTAL TABLES 10

Table S1. Sensitivity Analysis for Secondary Outcomes – AI Enabled Digital Stethoscope in the Intervention Arm at Study Entry (n=587) 10

Table S2. Echocardiographic parameters at baseline visit for participants in the intervention arm 11

Table S3. Demographic and clinical characteristics at baseline visit for participants in the intervention arm. 12

INVESTIGATOR LIST

Mayo Clinic (Jacksonville, FL & Rochester, MN): Demilade A. Adedinsewo, MD MPH; Andrea Carolina Morales-Lara, MD; KL Venkatachalam, MD; Jennifer Dugan, BA1; Xiaoxi Yao, PhD; Hanna J. Sledge, BS; Patrick W. Johnson, BS; Mikolaj A. Wieczorek, BS; Zachi I. Attia, PhD; Sabrina D. Phillips, MD; Mohamad H. Yamani, MD; Yvonne Butler Tobah, MD; Carl H. Rose, M; Emily E. Sharpe, MD; Francisco Lopez-Jimenez, MD, MBA; Paul A. Friedman, MD; Peter A. Noseworthy, MD; and Rickey E. Carter, PhD

United Kingdom (Consultant): Cynthia Yohanna, MD

Lagos University Teaching Hospital (LUTH): Bosede B. Afolabi, MBChB, Oyewole A. Kushimo, MBBS, Amam C. Mbakwem, MBBS, Ms. Damilola Onietan (lead study coordinator), Ms. Rachel Quao

University College Ibadan (UCH): Okechukwu S. Ogah, MBBS, Gbolahan Obajimi, MBBS, Mr. Solomon Aborisade, Ms. Oluwatosin Makinde, Ms. Victoria Ojo, Ms. Khadijat Adenike Olatunde, Dr. Olanike Allison Orimolade, Mr. Salawu Taofeek Oladotun, Mr. Oluwaseun Alabi, Dr. Akinsola Teslim Sanusi, Dr. Temitope Azeez Olukunle, Ms Ifeoluwa Alabede

University of Illorin Teaching Hospital (UITH): Kehinde F. Ibiyemi, MBBS, Ayodele Ogunmodede, MBBS, Hadijat Olaide Raji, MBBS, Ms. Adeola Adedokun, Ms. Omowunmi Ishola, Prof. Munir'deen A. Ijaiya, Dr. Abiodun Adeniran

Rasheed Shekoni Specialist Hospital (RSSH): Sadiq H. Ringim, MBBS; Abdullahi A. Habib MBBS; Sabiu M. Hamza, MBBS, Mrs. Asma'u Abdullahi, Mrs. Rabi Musa

Olabisi Onabanjo University Teaching Hospital (OOUTH): Olusoji E. Jagun, MBBS, Francisca O. Inofomoh, MBChB, Adeolu Temitope, MBBS, Ms. Ajifola Temilola, Adedeji Adekolade, MBChB, Adeniyi Akiseku, MBChB

Aminu Kano Teaching Hospital (AKTH): Kamilu M. Karaye, BMBCH, Sule A. Gaya, MBBS, Isiaka Alfa, MBBS, Dr. Naser A Ishaq, Mr. Yahaya Sa'ad

SUPPLEMENTAL FIGURES

Figure S1 Receiver operating characteristic curves for detection of left ventricular systolic dysfunction (A- LVEF <50%, B- LVEF <40%) using a digital stethoscope (maximal prediction) in an obstetric population in Nigeria.

Panel A shows the receiver operating characteristic (ROC) curve for the digital stethoscope maximum model output (over the three recording positions acquired) for detecting a left ventricular ejection fraction (LVEF) < 50% while Panel B shows the ROC curve for detection of LVEF <40%. Diagnostic performance metrics are embedded within each image at the model operating point, which is denoted as the red circle on the ROC curve. Stethoscope recordings deemed to be of sufficient quality were used for assessing model performance. All recordings were obtained on the same day as the echocardiogram. AUC – Area under the curve, LRT – Likelihood ratio.

Figure S2. Receiver operating characteristic curve for detection of cardiomyopathy (LVEF <50% A and LVEF <40% B.) using the FDA cleared 12-lead AI-ECG in an obstetric population in Nigeria.

Panel A shows the receiver operating characteristic (ROC) curve for the FDA cleared 12-lead AI-ECG model output for detecting a left ventricular ejection fraction (LVEF) < 50% while Panel B shows the ROC curve for detection of LVEF <40%. Diagnostic performance metrics are embedded within each image at the model operating point, which is denoted as the red circle on the ROC curve. Only ECGs deemed to be of sufficient quality by the FDA cleared device were analyzed. All ECGs were obtained on the same day as the echocardiogram. AUC – Area under the curve, LRT – Likelihood ratio.

Figure S3. Receiver operating characteristic curve for detection of cardiomyopathy (LVEF <50% A and LVEF <40% B.) using the original Mayo Clinic 12-lead AI-ECG in an obstetric population in Nigeria.

Panel A shows the receiver operating characteristic (ROC) curve for the original Mayo Clinic 12-lead AI-ECG model output for detecting a left ventricular ejection fraction (LVEF) < 50% while Panel B shows the ROC curve for detection of LVEF <40%. Diagnostic performance metrics are embedded within each image at the model operating point, which is denoted as the red circle on the ROC curve. All ECGs were obtained on the same day as the echocardiogram. AUC – Area under the curve, LRT – Likelihood ratio.

Figure S4. Enrollment sites and participants

This figure shows participant enrollment sites along with total sample size enrolled at each location, with the entire state shaded in blue and study site locations displayed on the map of Nigeria.

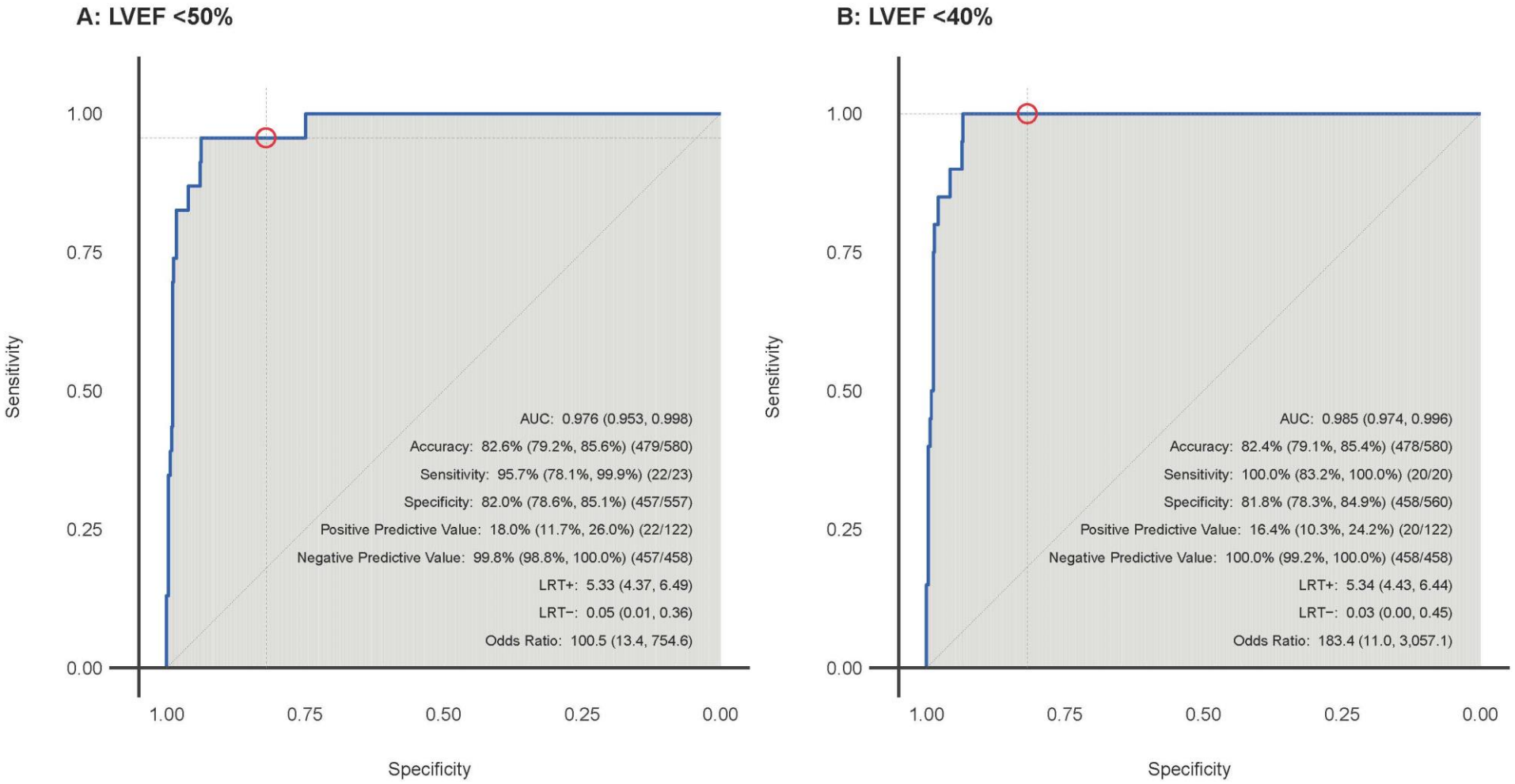
Figure S5. Digital stethoscope and chest recording location

Panel A shows the digital stethoscope with ECG and phonocardiogram recording demonstrated on a smartphone. Panel B shows V2 and angled recording positions on the chest.

Figure S6. Progression through the study visits stratified by clinical site and pregnancy status

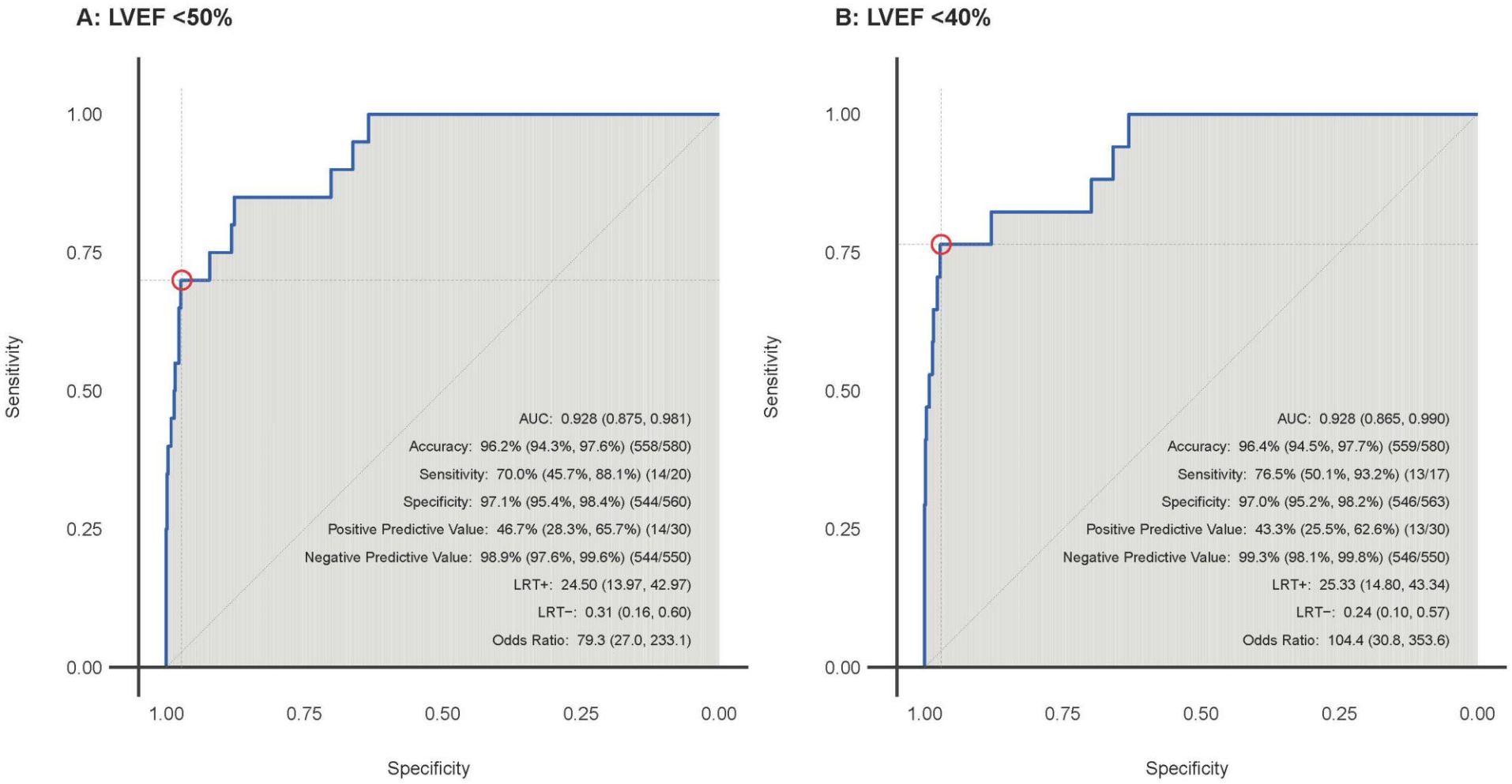
The figure shows the percentage of subjects completing the study protocol visits (up to 7) stratified by site and pregnancy status. The dashed line marks 50% the total number enrolled at each site. The median (Q1, Q3) number of visits completed was 2 (1, 3) for both the intervention and control arms. When broken down by pregnancy status, the median in the postpartum participants was 1 (1, 2) for both the intervention and control arm, and the participants pregnant at baseline had a median of 2 (1, 3) completed visits for each study arm.

Figure S1. Receiver operating characteristic for detection of cardiomyopathy (LVEF <50% A and LVEF <40% B.) using a digital stethoscope (maximum prediction) in an obstetric population in Nigeria.



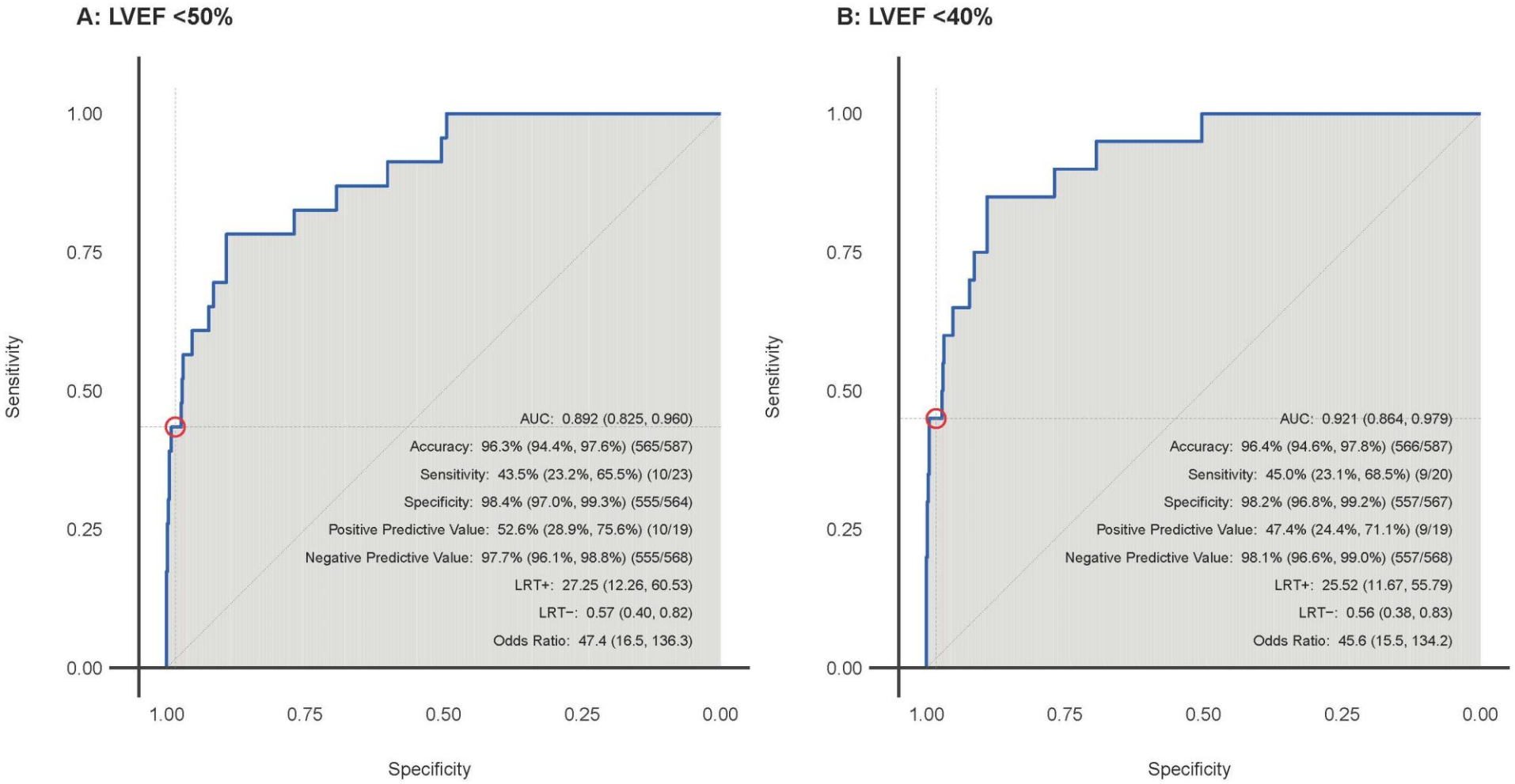
Panel A shows the receiver operating characteristic (ROC) curve for the digital stethoscope maximum model output (over the three recording positions acquired) for detecting a left ventricular ejection fraction (LVEF) < 50% while Panel B shows the ROC curve for detection of LVEF <40%. Diagnostic performance metrics are embedded within each image at the model operating point, which is denoted as the red circle on the ROC curve. Stethoscope recordings deemed to be of sufficient quality were used for assessing model performance. All recordings were obtained on the same day as the echocardiogram. AUC – Area under the curve, LRT – Likelihood ratio.

Figure S2. Receiver operating characteristic for detection of cardiomyopathy (LVEF <50% A and LVEF <40% B.) using the FDA cleared 12-lead AI-ECG model in an obstetric population in Nigeria.



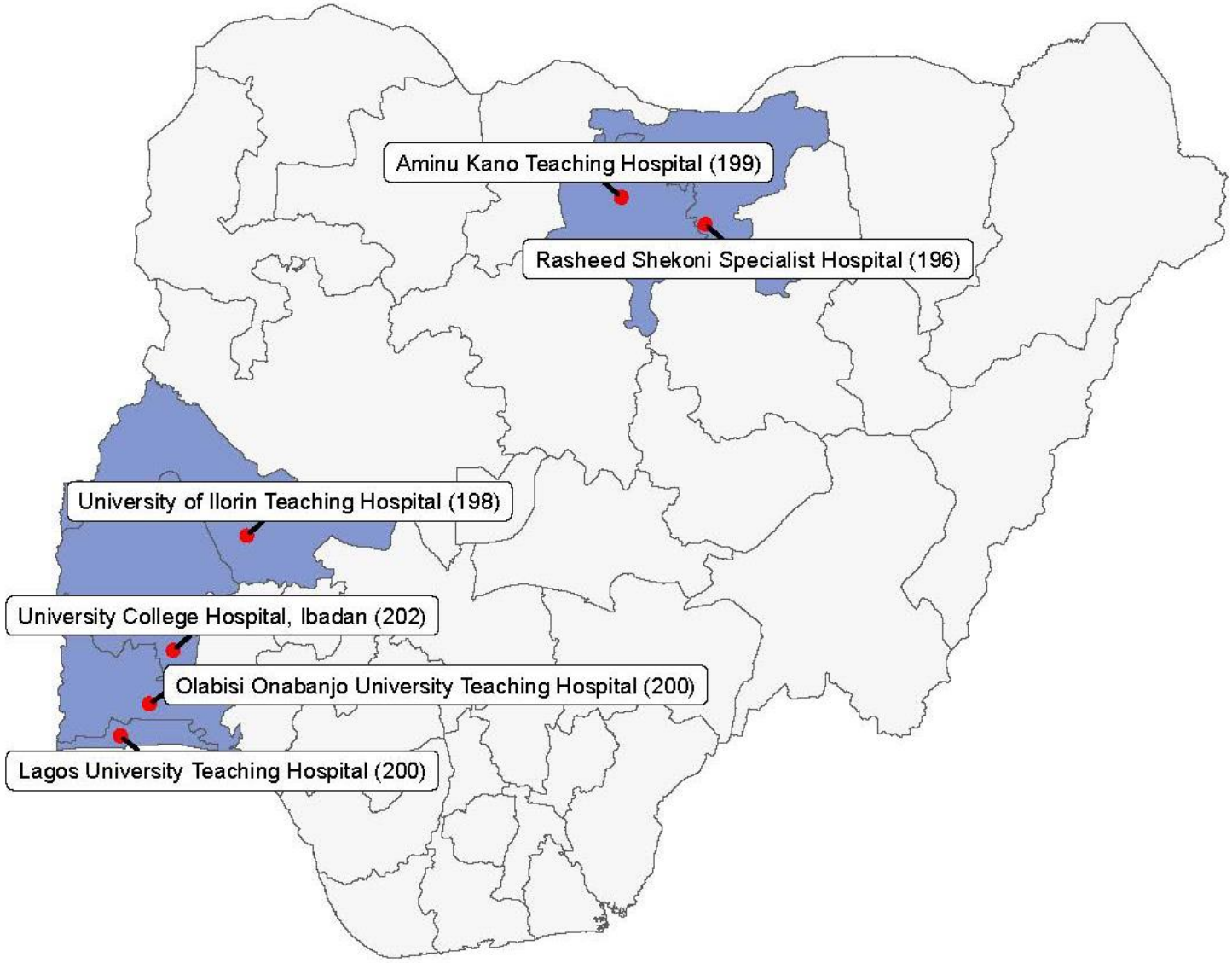
Panel A shows the receiver operating characteristic (ROC) curve for the FDA cleared 12-lead AI-ECG model output for detecting a left ventricular ejection fraction (LVEF) < 50% while Panel B shows the ROC curve for detection of LVEF <40%. Diagnostic performance metrics are embedded within each image at the model operating point, which is denoted as the red circle on the ROC curve. Only ECGs deemed to be of sufficient quality by the FDA cleared device were analyzed. All ECGs were obtained on the same day as the echocardiogram. AUC – Area under the curve, LRT – Likelihood ratio.

Figure S3. Receiver operating characteristic for detection of cardiomyopathy (LVEF <50% A and LVEF <40% B.) using the original Mayo Clinic 12-lead AI-ECG model in an obstetric population in Nigeria.



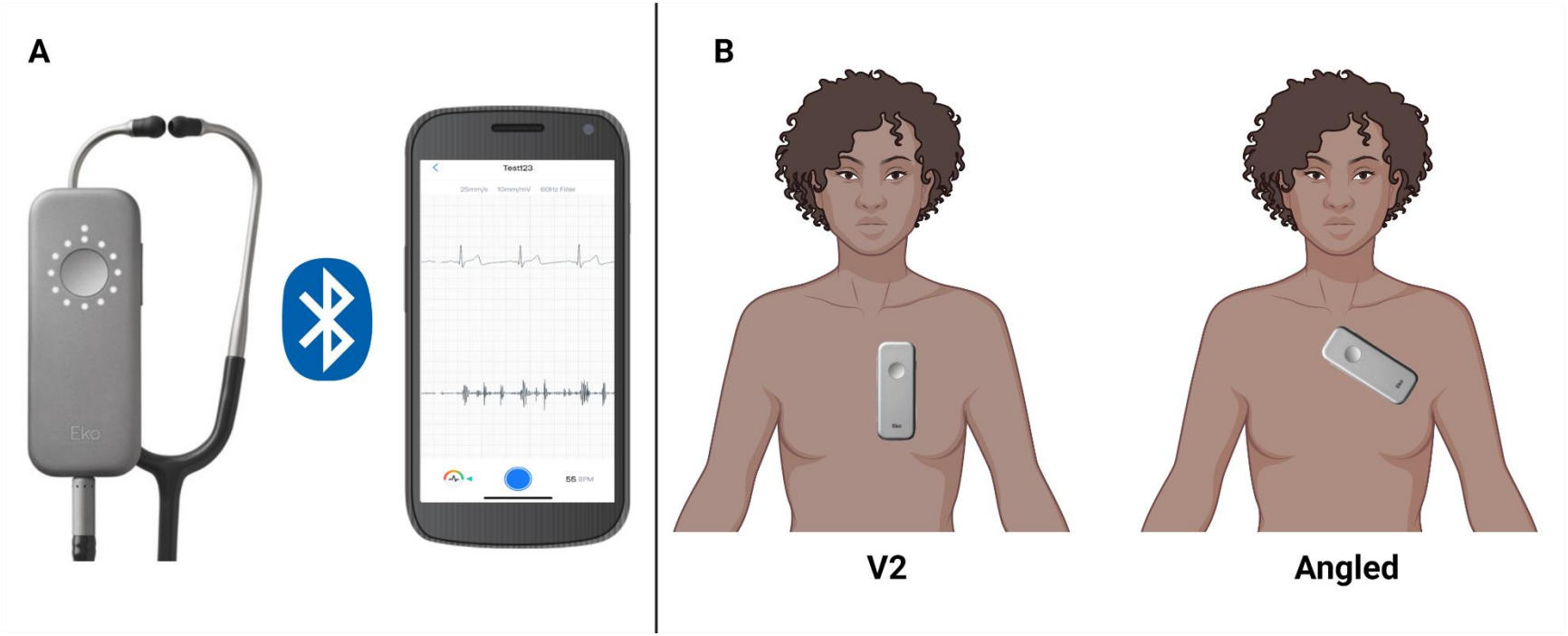
Panel A shows the receiver operating characteristic (ROC) curve for the original Mayo Clinic 12-lead AI-ECG model output for detecting a left ventricular ejection fraction (LVEF) < 50% while Panel B shows the ROC curve for detection of LVEF <40%. Diagnostic performance metrics are embedded within each image at the model operating point, which is denoted as the red circle on the ROC curve. All ECGs were obtained on the same day as the echocardiogram. AUC – Area under the curve, LRT – Likelihood ratio.

Figure S4. Enrollment sites and participants



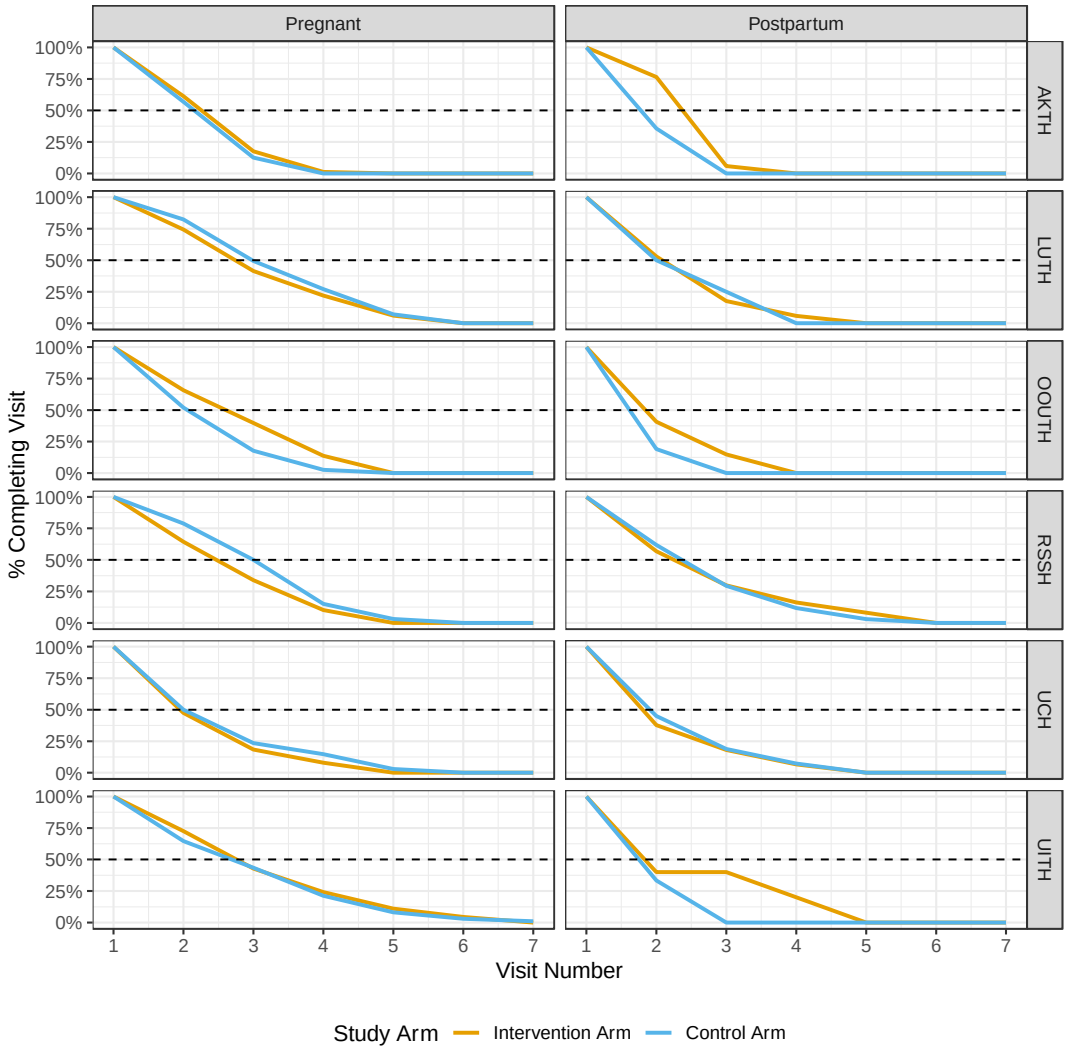
This figure shows participant enrollment sites along with total sample size enrolled at each location, with the entire state shaded in blue and study site locations displayed on the map of Nigeria.

Figure S5. Digital stethoscope and chest recording location.



Panel A shows the digital stethoscope with ECG and phonocardiogram recording demonstrated on a smartphone. Panel B shows V2 and angled recording positions on the chest. Figure illustration created using biorender.com.

Figure S6. Progression through the study visits stratified by clinical site and pregnancy status



The figure shows the percentage of subjects completing the study protocol visits (up to 7) stratified by site and pregnancy status. The dashed line marks 50% the total number enrolled at each site. The median (Q1, Q3) number of visits completed was 2 (1, 3) for both the intervention and control arms. When broken down by pregnancy status, the median in the postpartum participants was 1 (1, 2) for both the intervention and control arm, and the participants pregnant at baseline had a median of 2 (1, 3) completed visits for each study arm. Abbreviations: AKTH, Aminu Kano Teaching Hospital; LUTH, Lagos University Teaching Hospital; OOUTH, Olabisi Onabanjo University Teaching Hospital; RSSH, Rasheed Shekoni Specialist Hospital; UCH, University College Hospital, Ibadan; UITH, University of Ilorin Teaching Hospital; Q1, 25th percentile or first quartile; Q3, 75th percentile or third quartile.

SUPPLEMENTAL TABLES

Table S1. Sensitivity Analysis for Secondary Outcomes – AI Enabled Digital Stethoscope in the Intervention Arm at Study Entry (n=587)

Outcome	Model (threshold)	N	AUC	Sensitivity	Specificity	Positive Predictive Value	Negative Predictive Value
LVEF <=35%	Angled (0.43)	581/587	0.970 (0.937, 1.000)	94.1% (71.3%, 99.9%) 16/17	89.7% (86.9%, 92.1%) 506/564	21.6% (12.9%, 32.7%) 16/74	99.8% (98.9%, 100.0%) 506/507
	V2 (0.43)	582/587	0.960 (0.927, 0.992)	94.1% (71.3%, 99.9%) 16/17	86.4% (83.3%, 89.1%) 488/565	17.2% (10.2%, 26.4%) 16/93	99.8% (98.9%, 100.0%) 488/489
	Handheld (0.43)	572/587	0.968 (0.946, 0.990)	70.6% (44.0%, 89.7%) 12/17	95.0% (92.8%, 96.6%) 527/555	30.0% (16.6%, 46.5%) 12/40	99.1% (97.8%, 99.7%) 527/532
	Max Prediction (0.43)	587/587	0.982 (0.968, 0.997)	100.0% (80.5%, 100.0%) 17/17	78.8% (75.2%, 82.1%) 449/570	12.3% (7.3%, 19.0%) 17/138	100.0% (99.2%, 100.0%) 449/449
LVEF <40%	Angled (0.43)	581/587	0.975 (0.946, 1.000)	95.0% (75.1%, 99.9%) 19/20	90.2% (87.4%, 92.5%) 506/561	25.7% (16.2%, 37.2%) 19/74	99.8% (98.9%, 100.0%) 506/507
	V2 (0.43)	582/587	0.960 (0.932, 0.988)	95.0% (75.1%, 99.9%) 19/20	86.8% (83.8%, 89.5%) 488/562	20.4% (12.8%, 30.1%) 19/93	99.8% (98.9%, 100.0%) 488/489
	Handheld (0.43)	572/587	0.956 (0.928, 0.985)	65.0% (40.8%, 84.6%) 13/20	95.1% (93.0%, 96.8%) 525/552	32.5% (18.6%, 49.1%) 13/40	98.7% (97.3%, 99.5%) 525/532
	Max Prediction (0.43)	587/587	0.983 (0.970, 0.996)	100.0% (83.2%, 100.0%) 20/20	79.2% (75.6%, 82.5%) 449/567	14.5% (9.1%, 21.5%) 20/138	100.0% (99.2%, 100.0%) 449/449
LVEF <45%	Angled (0.43)	581/587	0.961 (0.928, 0.994)	87.0% (66.4%, 97.2%) 20/23	90.3% (87.6%, 92.6%) 504/558	27.0% (17.4%, 38.6%) 20/74	99.4% (98.3%, 99.9%) 504/507
	V2 (0.43)	582/587	0.958 (0.930, 0.986)	91.3% (72.0%, 98.9%) 21/23	87.1% (84.1%, 89.8%) 487/559	22.6% (14.6%, 32.4%) 21/93	99.6% (98.5%, 100.0%) 487/489
	Handheld (0.43)	572/587	0.931 (0.892, 0.970)	56.5% (34.5%, 76.8%) 13/23	95.1% (92.9%, 96.7%) 522/549	32.5% (18.6%, 49.1%) 13/40	98.1% (96.6%, 99.1%) 522/532
	Max Prediction (0.43)	587/587	0.973 (0.947, 0.998)	95.7% (78.1%, 99.9%) 22/23	79.4% (75.9%, 82.7%) 448/564	15.9% (10.3%, 23.1%) 22/138	99.8% (98.8%, 100.0%) 448/449
LVEF <50%	Angled (0.43)	581/587	0.961 (0.928, 0.994)	87.0% (66.4%, 97.2%) 20/23	90.3% (87.6%, 92.6%) 504/558	27.0% (17.4%, 38.6%) 20/74	99.4% (98.3%, 99.9%) 504/507
	V2 (0.43)	582/587	0.958 (0.930, 0.986)	91.3% (72.0%, 98.9%) 21/23	87.1% (84.1%, 89.8%) 487/559	22.6% (14.6%, 32.4%) 21/93	99.6% (98.5%, 100.0%) 487/489
	Handheld (0.43)	572/587	0.931 (0.892, 0.970)	56.5% (34.5%, 76.8%) 13/23	95.1% (92.9%, 96.7%) 522/549	32.5% (18.6%, 49.1%) 13/40	98.1% (96.6%, 99.1%) 522/532
	Max Prediction (0.43)	587/587	0.973 (0.947, 0.998)	95.7% (78.1%, 99.9%) 22/23	79.4% (75.9%, 82.7%) 448/564	15.9% (10.3%, 23.1%) 22/138	99.8% (98.8%, 100.0%) 448/449

The results provided in this table are based on baseline assessments only where all participants in the intervention arm had a confirmatory echocardiogram done for validation of AI model performance. All results include digital stethoscope recordings acquired on the same day as the echocardiogram.

The Angled, V2, Handheld models are based on the digital stethoscope recordings. The max prediction is derived as the maximum model output across the three recording positions (i.e., any positive prediction). The model and operating cut point of 0.43 were developed to detect an LVEF <40%. This table provides a sensitivity analysis that retains all data points including poor-quality digital stethoscope recordings.

LVEF, left ventricular ejection fraction, US FDA, United States Food and Drug Administration.

Table S2. Echocardiographic parameters at baseline visit for participants in the intervention arm

Characteristic	N	Median (Q1 - Q3) or No. (%) of patients			
		Overall (N=587)	LVEF<50% (N=23)	LVEF≥50% (N=564)	P-value
Baseline LVEF %	587	61.00 (56.75-65.20)	24.50 (21.50-32.75)	61.00 (57.00-66.00)	<0.001
Cardiac Output (L/m)	564	5.36 (4.00-6.78)	3.90 (2.95-5.30)	5.40 (4.10-6.87)	<0.001
Cardiac Index (L/min/m2)	564	3.00 (2.32-3.85)	2.49 (1.95-3.62)	3.02 (2.33-3.85)	0.035
LVEDD (mm)	587	46.80 (43.00-50.00)	60.00 (57.00-66.00)	46.13 (42.71-49.63)	<0.001
LVESD (mm)	587	31.80 (29.00-34.50)	53.00 (49.00-58.65)	31.18 (29.00-34.00)	<0.001
Septal thickness in mm	587	8.30 (7.20-9.70)	7.80 (7.00-8.60)	8.30 (7.20-9.82)	0.019
Posterior wall thickness in mm	587	8.00 (7.10-10.00)	8.00 (7.00-8.75)	8.03 (7.16-10.00)	0.12
LAVI (ml/m2)	580	27.84 (21.74-34.70)	44.30 (39.28-64.53)	27.30 (21.30-34.00)	<0.001
RAVI (ml/m2)	472	22.00 (16.78-27.83)	38.00 (29.65-55.06)	21.90 (16.50-27.00)	<0.001
Left ventricular mass (g)	584	131.00 (105.77-159.06)	177.00 (150.00-237.00)	130.00 (104.00-158.00)	<0.001
LVMI (g/m2)	584	75.00 (61.58-90.84)	113.00 (96.50-146.00)	73.12 (61.00-89.40)	<0.001
Left ventricular RWT (%)	585	0.36 (0.31-0.42)	0.25 (0.20-0.30)	0.36 (0.31-0.43)	<0.001
LV longitudinal peak systolic strain (%)	164	-18.35 (-20.60--16.08)	-9.20 (-9.20--9.20)	-18.40 (-20.60--16.15)	0.087
TAPSE (mm)	561	23.90 (20.80-26.60)	18.00 (14.20-19.40)	24.00 (21.00-26.90)	<0.001
TR velocity (m/sec)	170	2.02 (1.54-2.50)	3.11 (2.80-3.28)	1.95 (1.50-2.32)	<0.001
E wave velocity (m/sec)	583	0.87 (0.71-1.00)	1.09 (0.91-1.31)	0.86 (0.71-0.99)	<0.001
A wave velocity (m/sec)	554	0.59 (0.49-0.68)	0.41 (0.34-0.63)	0.59 (0.49-0.69)	0.009
e' velocity (m/sec)	577	0.11 (0.09-0.13)	0.07 (0.06-0.09)	0.11 (0.09-0.13)	<0.001
E/e' ratio	578	7.72 (6.32-9.64)	14.70 (11.10-17.00)	7.60 (6.25-9.43)	<0.001
VHD (left sided)	587	16 (2.7%)	12 (52.2%)	4 (0.7%)	<0.001
VHD (right sided)	587	45 (7.7%)	11 (47.8%)	34 (6.0%)	<0.001

P-values result from a Wilcoxon rank sum test (continuous variables) or Chi-Square test (categorical variables) evaluated at the alpha = 0.05 level of significance (two-sided). No adjustments for multiple comparison were performed.

*GLS – global longitudinal peak systolic strain, LAVI- Left atrial volume index LVEDD - Left ventricular end diastolic diameter, LVESD - Left ventricular end systolic diameter, LVMI - Left ventricular mass index, RAVI - Left atrial volume index, RWT - relative wall thickness, TR – Tricuspid regurgitation, VHD -Valvular heart disease (moderate or severe stenosis or regurgitation, left sided VDH includes mitral valve, aortic valves and right sided VHD includes tricuspid, pulmonic valves); TAPSE -Tricuspid Annular Plane Systolic Excursion Using M-Mode.

Table S3. Demographic and clinical characteristics at baseline visit for participants in the intervention arm.

		Median (Q1 - Q3) or No. (%) of patients			
Characteristic	N	Overall (N=587)	LVEF<50% (N=23)	LVEF≥50% (N=564)	P-value
Age, years	587	31 (27-35)	30 (24-34)	31 (27-35)	0.20
Race (Black)	587	587 (100.0%)	23 (100.0%)	564 (100.0%)	
Ethnicity	587				<0.001
Hausa		163 (27.8%)	16 (69.6%)	147 (26.1%)	
Igbo		61 (10.4%)	0 (0.0%)	61 (10.8%)	
Other		35 (6.0%)	2 (8.7%)	33 (5.9%)	
Yoruba		328 (55.9%)	5 (21.7%)	323 (57.3%)	
Status at Entry	587				<0.001
Pregnant		423 (72.1%)	2 (8.7%)	421 (74.6%)	
Postpartum		164 (27.9%)	21 (91.3%)	143 (25.4%)	
Timepoint of Pregnancy/Postpartum	587				<0.001
First trimester		37 (6.3%)	0 (0.0%)	37 (6.6%)	
Second trimester		150 (25.6%)	1 (4.3%)	149 (26.4%)	
Third trimester		238 (40.5%)	0 (0.0%)	238 (42.2%)	
Time of delivery or up to 6 weeks after delivery		116 (19.8%)	9 (39.1%)	107 (19.0%)	
Greater than 6 weeks and up to 3 months after delivery		23 (3.9%)	6 (26.1%)	17 (3.0%)	
Greater than 3 months and up to 5 months after delivery		7 (1.2%)	1 (4.3%)	6 (1.1%)	
Greater than 5 months and up to 12 months after delivery		16 (2.7%)	6 (26.1%)	10 (1.8%)	
Weight, kg	587	70 (60-81)	51 (49-59)	70 (61-82)	<0.001
Height, cm	587	161 (157-165)	159 (155-163)	161 (157-165)	0.048
Systolic Blood Pressure, mmHg	587	110 (100-120)	100 (100-118)	110 (100-120)	0.13
Diastolic Blood Pressure, mmHg	587	70 (60-80)	70 (60-88)	70 (60-80)	0.41
Resting Heart Rate, bpm	587	87 (80-95)	105 (97-112)	86 (80-94)	<0.001
Hemoglobin (g/dL)	516	11 (10-12)	10 (9-11)	11 (10-12)	0.023
Hematocrit (%)	534	33 (30-35)	31 (28-33)	33 (30-35)	0.038
Blood type	553				0.74
A		98 (17.7%)	3 (14.3%)	95 (17.9%)	
B		106 (19.2%)	5 (23.8%)	101 (19.0%)	
AB		29 (5.2%)	2 (9.5%)	27 (5.1%)	
O		320 (57.9%)	11 (52.4%)	309 (58.1%)	
Hemoglobin Genotype	553				0.63
AA		402 (72.7%)	16 (84.2%)	386 (72.3%)	
AS		131 (23.7%)	2 (10.5%)	129 (24.2%)	

Characteristic	N	Median (Q1 - Q3) or No. (%) of patients			P-value
		Overall (N=587)	LVEF<50% (N=23)	LVEF≥50% (N=564)	
SS		5 (0.9%)	0 (0.0%)	5 (0.9%)	
Sc		1 (0.2%)	0 (0.0%)	1 (0.2%)	
Other		14 (2.5%)	1 (5.3%)	13 (2.4%)	
Infectious Screen					
HIV	587	7 (1.2%)	0 (0.0%)	7 (1.2%)	1.00
Hepatitis C	587	3 (0.5%)	0 (0.0%)	3 (0.5%)	1.00
Syphilis	587	4 (0.7%)	0 (0.0%)	4 (0.7%)	1.00
Hepatitis B	587	9 (1.5%)	0 (0.0%)	9 (1.6%)	1.00
Urinalysis positive for protein	565	61 (10.8%)	6 (31.6%)	55 (10.1%)	0.010

P-values result from a Wilcoxon rank sum test (continuous variables) or Chi-Square test (categorical variables) evaluated at the alpha = 0.05 level of significance (two- sided).
No adjustments for multiple comparison were performed.
HIV -Human immunodeficiency virus.

CONSORT-AI

CONSORT-AI checklist of information to include when reporting a randomised trials of AI interventions

Section	Item	CONSORT 2010 Item[a]	CONSORT-AI Item		Addressed on Page No[b]
Title and Abstract					
Title and Abstract	1a	Identification as a randomised trial in the title	CONSORT-AI 1a,b Elaboration	(i) Indicate that the intervention involves artificial intelligence/machine learning in the title and/or abstract and specify the type of model.	1
	1b	Structured summary of trial design, methods, results, and conclusions (for specific guidance see CONSORT for abstracts)		(ii) State the intended use of the AI intervention within the trial in the title and/or abstract.	3
Introduction					
Background and objectives	2a	Scientific background and explanation of rationale	CONSORT-AI 2a (i) Extension	Explain the intended use of the AI intervention in the context of the clinical pathway, including its purpose and its intended users (e.g. healthcare professionals, patients, public).	4,5
	2b	Specific objectives or hypotheses			4,5
Methods					
Trial design	3a	Description of trial design (such as parallel, factorial) including allocation ratio			34-36
	3b	Important changes to methods after trial commencement (such as eligibility criteria), with reasons			40-41
Participants	4a	Eligibility criteria for participants	CONSORT-AI 4a (i) Elaboration	State the inclusion and exclusion criteria at the level of participants.	35
			CONSORT-AI 4a (ii) Extension	State the inclusion and exclusion criteria at the level of the input data.	37-38, 39-40
	4b	Settings and locations where the data were collected	CONSORT-AI 4b Extension	Describe how the AI intervention was integrated into the trial setting, including any onsite or offsite requirements.	35-37
Interventions	5	The interventions for each group with sufficient details to allow replication, including how and when they were actually administered	CONSORT-AI 5 (i) Extension	State which version of the AI algorithm was used.	37
			CONSORT-AI 5 (ii) Extension	Describe how the input data were acquired and selected for the AI intervention.	36-37
			CONSORT-AI 5 (iii) Extension	Describe how poor quality or unavailable input data were assessed and handled.	37-38, 39-40
			CONSORT-AI 5 (iv) Extension.	Specify whether there was human-AI interaction in the handling of the input data, and what level of expertise was required of users.	36-37

			CONSORT-AI 5 (v) Extension	Specify the output of the AI intervention	36
			CONSORT-AI 5 (vi) Extension	Explain how the AI intervention's outputs contributed to decision-making or other elements of clinical practice.	36-37
Outcomes	6a	Completely defined pre-specified primary and secondary outcome measures, including how and when they were assessed			38-40
	6b	Any changes to trial outcomes after the trial commenced, with reasons			N/A
Sample size	7a	How sample size was determined			40-41
	7b	When applicable, explanation of any interim analyses and stopping guidelines			N/A
Randomisation					
Sequence generation	8a	Method used to generate the random allocation sequence			35
	8b	Type of randomisation; details of any restriction (such as blocking and block size)			35
Allocation concealment mechanism	9	Mechanism used to implement the random allocation sequence (such as sequentially numbered containers), describing any steps taken to conceal the sequence until interventions were assigned			35
Implementation	10	Who generated the random allocation sequence, who enrolled participants, and who assigned participants to interventions			35
Blinding	11a	If done, who was blinded after assignment to interventions (for example, participants, care providers, those assessing outcomes) and how			N/A
	11b	If relevant, description of the similarity of interventions			35-38
Statistical methods	12a	Statistical methods used to compare groups for primary and secondary outcomes			40-42
	12b	Methods for additional analyses, such as subgroup analyses and adjusted analyses			40-42
Results					
Participant flow (a diagram is strongly recommended)	13a	For each group, the numbers of participants who were randomly assigned, received intended treatment, and were analysed for the primary outcome			6
	13b	For each group, losses and exclusions after randomisation, together with reasons			6

Recruitment	14a	Dates defining the periods of recruitment and follow-up			6
	14b	Why the trial ended or was stopped			6
Baseline data	15	A table showing baseline demographic and clinical characteristics for each group			22-23
Numbers analysed	16	For each group, number of participants (denominator) included in each analysis and whether the analysis was by original assigned groups			6-7
Outcomes and estimation	17a	For each primary and secondary outcome, results for each group, and the estimated effect size and its precision (such as 95% confidence interval)			7-11
	17b	For binary outcomes, presentation of both absolute and relative effect sizes is recommended			7-11
Ancillary analyses	18	Results of any other analyses performed, including subgroup analyses and adjusted analyses, distinguishing pre-specified from exploratory			7-11
Harms	19	All important harms or unintended effects in each group (for specific guidance see CONSORT for harms)	CONSORT-AI 19 Extension	Describe results of any analysis of performance errors and how errors were identified, where applicable. If no such analysis was planned or done, explain why not.	11, 34, 37-40
Discussion					
Limitations	20	Trial limitations, addressing sources of potential bias, imprecision, and, if relevant, multiplicity of analyses			16-18
Generalisability	21	Generalisability (external validity, applicability) of the trial findings			17
Interpretation	22	Interpretation consistent with results, balancing benefits and harms, and considering other relevant evidence			12-19
Other Information					
Registration	23	Registration number and name of trial registry			34
Protocol	24	Where the full trial protocol can be accessed, if available			34
Funding	25	Sources of funding and other support (such as supply of drugs), role of funders	CONSORT-AI 25 Extension.	State whether and how the AI intervention and/or its code can be accessed, including any restrictions to access or re-use.	19-20, 34

[a] We strongly recommend reading this statement in conjunction with the CONSORT 2010 Explanation and Elaboration for important clarifications on all the items.
[b] Indicates page numbers to be completed by authors during protocol development.