

ORIGINAL RESEARCH

Machine Learning–Based Critical Congenital Heart Disease Screening Using Dual-Site Pulse Oximetry Measurements

Heather Siefkes , MD, MSCI; Luca Cerny Oliveira , MS; Robert Koppel , MD; Whitnee Hogan, MD; Meena Garg, MD; Erlinda Manalo , MD; Nicole Cresalia , MD; Zhengfeng Lai , PhD; Daniel Tancredi , PhD; Satyan Lakshminrusimha , MD; Chen-Nee Chuah , PhD

BACKGROUND: Oxygen saturation (SpO₂) screening has not led to earlier detection of critical congenital heart disease (CCHD). Adding pulse oximetry features (ie, perfusion data and radiofemoral pulse delay) may improve CCHD detection, especially coarctation of the aorta (CoA). We developed and tested a machine learning (ML) pulse oximetry algorithm to enhance CCHD detection.

METHODS AND RESULTS: Six sites prospectively enrolled newborns with and without CCHD and recorded simultaneous pre- and postductal pulse oximetry. We focused on models at 1 versus 2 time points and with/without pulse delay for our ML algorithms. The sensitivity, specificity, and area under the receiver operating characteristic curve were compared between the SpO₂-alone and ML algorithms. A total of 523 newborns were enrolled (no CHD, 317; CHD, 74; CCHD, 132, of whom 21 had isolated CoA). When applying the SpO₂-alone algorithm to all patients, 26.2% of CCHD would be missed. We narrowed the sample to patients with both 2 time point measurements and pulse-delay data (no CHD, 65; CCHD, 14) to compare ML performance. Among these patients, sensitivity for CCHD detection increased with both the addition of pulse delay and a second time point. All ML models had 100% specificity. With a 2-time-points+pulse-delay model, CCHD sensitivity increased to 92.86% ($P=0.25$) compared with SpO₂ alone (71.43%), and CoA increased to 66.67% ($P=0.5$) from 0. The area under the receiver operating characteristic curve for CCHD and CoA detection significantly improved (0.96 versus 0.83 for CCHD, 0.83 versus 0.48 for CoA; both $P=0.03$) using the 2-time-points+pulse-delay model compared with SpO₂ alone.

CONCLUSIONS: ML pulse oximetry that combines oxygenation, perfusion data, and pulse delay at 2 time points may improve detection of CCHD and CoA within 48 hours after birth.

REGISTRATION: URL: <https://www.clinicaltrials.gov/study/NCT04056104?term=NCT04056104&rank=1>; Unique identifier: NCT04056104.

Key Words: critical congenital heart disease ■ machine learning ■ pulse oximetry

See Editorial by XXX.

Oxygen saturation (SpO₂)-based critical congenital heart disease (CCHD) screening is more sensitive than physical examination alone among asymptomatic infants.¹ SpO₂-based CCHD screening is now widely mandated in the United States.¹ While ~900 newborns with CCHD are detected by SpO₂-based

screening in the United States annually, it is also estimated to miss nearly as many infants with CCHD.² The vast majority of the infants not detected by SpO₂-based CCHD screening have systemic obstruction lesions such as coarctation of the aorta (CoA). It is estimated that 560 CoA cases annually are missed due

Correspondence to: Heather Siefkes, MD, MSCI, University of California, Davis Ticon II, 2516 Stockton Blvd, Sacramento, CA 95817. Email: hsiefkes@ucdavis.edu

This manuscript was sent to John L. Jefferies, MD, MPH, Guest Editor, for review by expert referees, editorial decision, and final disposition.

For Sources of Funding and Disclosures, see page 12.

© 2024 The Authors. Published on behalf of the American Heart Association, Inc., by Wiley. This is an open access article under the terms of the [Creative Commons Attribution-NonCommercial-NoDerivs](#) License, which permits use and distribution in any medium, provided the original work is properly cited, the use is non-commercial and no modifications or adaptations are made.

JAHA is available at: www.ahajournals.org/journal/jaha

CLINICAL PERSPECTIVE

What Is New?

- Coarctation of the aorta (CoA) remains the most common critical congenital heart defect detected late (>48 hours), and nearly half of postnatally diagnosed CoAs were detected late.
- Machine learning oximetry combining oxygen and perfusion data, including radiofemoral pulse delay, with measurements at 2 time points within 48 hours after birth has the potential to improve critical congenital heart defect and CoA detection.
- Among newborns with prenatally suspected CoA, machine learning oximetry correctly classified the newborns who were ultimately determined to be healthy.

What Are the Clinical Implications?

- Use of readily available pulse oximetry data, including perfusion and pulse-delay measurements, could further enhance the rate of early critical congenital heart defect and CoA detection.
- Machine learning pulse oximetry may help rule out critical CoA sooner and thus lower neonatal intensive care unit length of stay and normalize care such as feeding for newborns with prenatally suspected CoA.

Nonstandard Abbreviations and Acronyms

AWAD	automated waveform artifact detection
CCHD	critical congenital heart disease
CoA	coarctation of the aorta
ML	machine learning
PAI	pulse amplitude index
RFE	recursive feature elimination
SpO₂	oxygen saturation

to false-negative SpO₂ screens.² These estimates are based on the population expected to be screened, which are presumably healthy newborns without prenatal concerns for CCHD. In fact, prenatal detection of CoA is also challenging. Thus, improvement in postnatal detection is necessary due to the high rates of morbidity and death associated with late detection.³

Measurements other than SpO₂ from photoplethysmography have been suggested as possible screening tools to improve CCHD detection and specifically CoA. The most commonly studied measurement is the perfusion index, a measurement of pulsatile blood flow.

The perfusion index has been shown to be abnormal in infants with CoA and may improve CCHD detection but has unacceptable false-positive rates.⁴⁻⁷ When 2 probes are simultaneously in place, 1 on the right hand and 1 on any foot, the time difference in pulse arriving at the 2 extremities, or radiofemoral pulse delay, can also be measured and has also been previously found to be longer in infants with CoA.⁸ Diastolic and systolic components of the photoplethysmography waveform such as slopes have been studied and also noted to have differences in newborns with CoA.⁹ These photoplethysmography features have mostly been studied individually and generally at ages that are after normal newborn discharge (ie, >48 hours). We hypothesized that a combination of photoplethysmography features, in addition to SpO₂, would improve detection of CCHD compared with SpO₂-alone screening. We attempted to determine the optimal timing and type of measurements that enhance CCHD screening. Thus, we conducted a prospective cohort study of newborns with and without congenital heart disease (CHD) using machine learning (ML) techniques to identify the best combination of photoplethysmography features for CCHD detection.

METHODS

Deidentified data from participants who consented to data sharing will be made available upon request only. The ML code, however, will be publicly available at the link provided later in the methods. The University of California, Davis Institutional Review Board approved this study for all participating sites. The study was registered with [Clinicaltrials.gov](https://clinicaltrials.gov) (NCT04056104). Written informed consent was obtained for all enrolled newborns. Due to the rapidly changing physiology of newborns with CCHD, pulse oximetry recordings were permitted in any newborn suspected of CCHD with pre- and postductal pulse oximetry probes in place before obtaining consent. Data were analyzed only if written consent was obtained within 2 weeks.

This study was a 6-site multicenter prospective cohort of newborns with and without CHD. Participating centers included University of California, Davis, University of California Los Angeles, University of California San Francisco, Cohen's Children Medical Center (New York), Sutter Medical Center (Sacramento, CA), and University of Utah/Primary Children's Hospital. Enrollment occurred from October 2019 to January 2022. The cohort of newborns without CHD included asymptomatic newborns, predominantly from the well-newborn nursery, who were to undergo routine SpO₂-based CCHD screening. The exclusion criteria for this portion of the cohort were having an echocardiogram already completed (and thus no longer qualifying for CCHD screening) and concerns for

loss to follow-up and inability to confirm absence of CCHD (eg, anticipated foster placement). To capture the cohort of newborns with CHD, we enrolled newborns with prenatally or postnatally suspected CHD regardless of completion of a confirmatory postnatal echocardiogram before enrollment to ensure that collection of our measurements was not delayed while awaiting postnatal echocardiogram results. The exclusion criteria for newborns with CHD were (1) isolated patent ductus arteriosus or patent foramen ovale/atrial septal defect, (2) corrective surgical or catheter procedure completed before enrollment, and (3) active vasoactive infusions or other cardiac medications other than prostaglandin E1.

Preductal (right hand) and postductal (any foot) pulse oximetry measurements were simultaneously recorded in all infants for 5 minutes up to 3 times (0–24, 24–48, and >48 hours after birth). Our pulse oximetry recording process and system has been previously described.¹⁰ Briefly, we used 2 Nonin WristOx2 3150 oximeters (Nonin Medical Inc, Plymouth, MN) that then connected via Bluetooth to a Pi-top, a laptop computer that uses Raspberry Pi microcomputers (Linux based) to record pulse oximetry data labeled with study identification numbers. The Pi-top displayed both pulse oximetry waveforms along with SpO_2 and pulse amplitude index (PAI), which is analogous to the perfusion index, values in real time.

SpO_2 -Alone CCHD Screening Classification

To classify the pass/fail status based on the SpO_2 data alone, research personnel documented single pre- and postductal SpO_2 values once the waveforms were artifact free for at least 10 seconds during each data recording, which is similar to current practice for the recommended SpO_2 -alone screening algorithm. We applied the most recent SpO_2 recommended algorithm¹¹ to these values to assign a pass/fail. The SpO_2 measurement was considered failing if (1) either the pre- or postductal SpO_2 were <90% or (2) either the pre- or postductal SpO_2 were <95% or >3% absolute difference between the pre- and postductal SpO_2 on 2 measurements. While some of our patients had SpO_2 measurements collected before 24 hours of age, the recommended algorithm recommends performing the SpO_2 measurement after 24 hours of age or sooner only if being discharged.¹¹ Thus, we applied this recommended SpO_2 algorithm to patients using the first pre- and postductal SpO_2 values after 24 hours as the start of the algorithm. When using this algorithm, if the last available research SpO_2 values resulted with a “re-test” recommendation (single SpO_2 measurement with either the pre or postductal SpO_2 were <95% or >3% absolute difference between the pre- and postductal

SpO_2), we conservatively assigned a fail to bias toward the null hypothesis for CCHD detection.

To classify patients as either having CHD or not, we defined CCHD as a defect requiring corrective or palliative surgical or catheter-based intervention within 30 days of age or requiring prostaglandin therapy until corrective or palliative intervention was completed if intervention was after 30 days of age. Patients undergoing pulmonary artery banding alone to limit excess pulmonary blood flow were not considered as CCHD and were instead labeled noncritical CHD. For the purposes of evaluating sensitivity for critical CoA, only patients with isolated CoA or with less significant defects (ie, ventricular septal defect) were classified as CoA. Thus, patients with hypoplastic left heart syndrome, single ventricle or double-outlet right ventricle with arch obstruction, or another CCHD defect with arch obstruction were not classified as isolated CoA, but instead were classified as their other CCHD defect. We confirmed that patients remained without CHD by electronic medical record review to at least 6 weeks of age. If follow-up information was not available within the electronic medical record, then direct query of parents/guardians via phone/email/text was completed. A patient was considered lost to follow-up if 5 contacts were attempted without success.

ML Methods

There are 2 components to the ML pipeline: an automated waveform artifact detection (AWAD) component followed by a CCHD detection stage (Figure 1). Both the AWAD and CCHD ML models use features extracted from pulse segments extracted from the photoplethysmography signal collected. We defined a pulse segment as the slice between 2 onsets, as shown in Figure 2. Our motion artifact ML model has been previously described in more detail.¹² Briefly, this model used multiple photoplethysmography recordings from 1 hand and 1 foot of 21 newborns (a subset from the cohort presented here) that were labeled for artifacts by 3 trained observers. A total of 6 hours and 42 minutes of recordings, which included 57 658 beats, were used to train and test the artifact detection model. This model used agreed artifact labels as ground truth, meaning a photoplethysmography pulse would only be considered normal if no annotators signaled it as an artifact. The model was trained with 44.04% of the annotated waveforms (2 hours and 57 minutes from 11 patients) and tested on the remaining 55.96% of annotated waveforms (3 hours and 45 minutes from 10 patients). A total of 12 features were taken from a 3-pulse segment consisting of each pulse being classified and its neighboring pulses (see Figure 2). These features include but are not limited to systolic phase duration, diastolic phase duration,

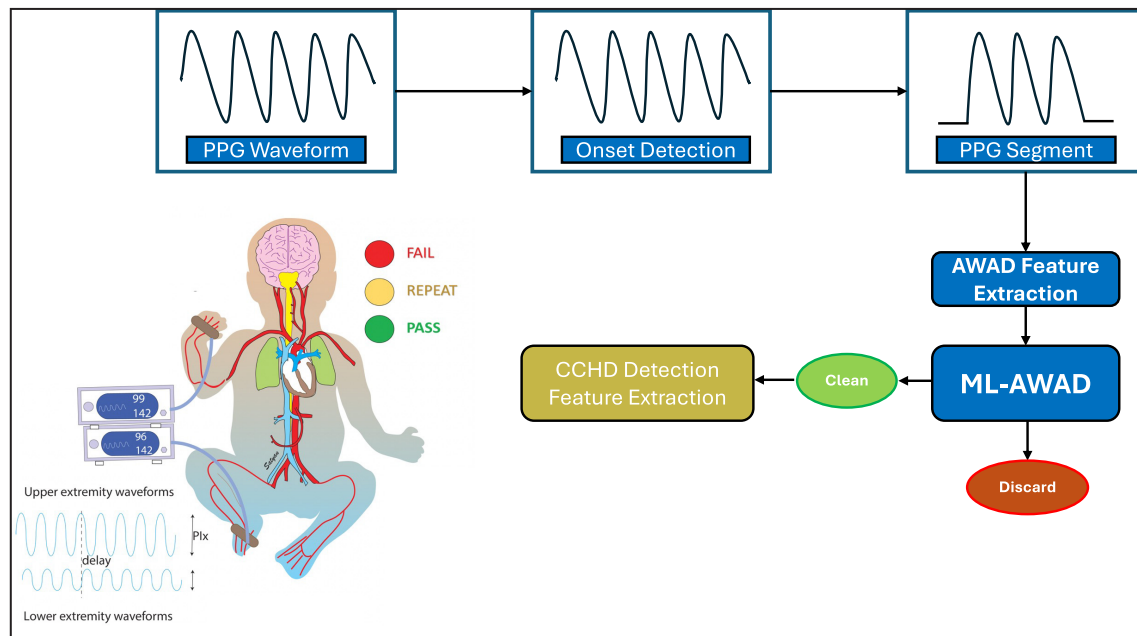


Figure 1. A schematic representation of the entire end-to-end pipeline developed during this study.

The first step is preprocessing and feature extraction of the waveform for the AWAD model. Once all artifact-positive pulses are removed, the next step is to perform feature extraction on the clean segments for CCHD prediction. All the steps above were completed for this study but not in an automated end-to-end pipeline (as this process was developed during this study). This end-to-end pipeline is now fully automated from start to finish and can be tested in situ in future studies. AWAD indicates automated waveform artifact detection; CCHD, critical congenital heart disease; and ML, machine learning. Copyright Heather Siefkes.

and dynamic time-warped Euclidean distance that describe the pulse shape and how they relate to neighboring pulses.¹² Random forest was chosen as the classifier. The best features were chosen through recursive feature elimination in 5-fold cross-validation. The AWAD model achieved 81.71% accuracy with 68.40% specificity and 89.75% sensitivity (compared with the annotators' labeling of artifact). This AWAD model was then applied to all pulse oximetry data, and only pulses considered nonartifact by AWAD were used to train the CCHD detection model, while artifact pulses detected by AWAD were discarded (as seen in Figure 1).

The features used in CCHD detection were either gathered directly from the pulse oximeters or calculated from the artifact-free photoplethysmography segments. The SpO_2 , PAI, and heart rate measurements were supplied as numerical values from the oximeters. We calculated the following features from artifact-free photoplethysmography segments: radiofemoral pulse delay, slope, average rate of rise, and average rate of fall (examples shown in Figure 3). Features from all pulses in each patient's photoplethysmography recording were leveraged to generate mean, maximum, minimum, and variance values for each feature. From each recording, we extracted a total of 70 features from foot and hand

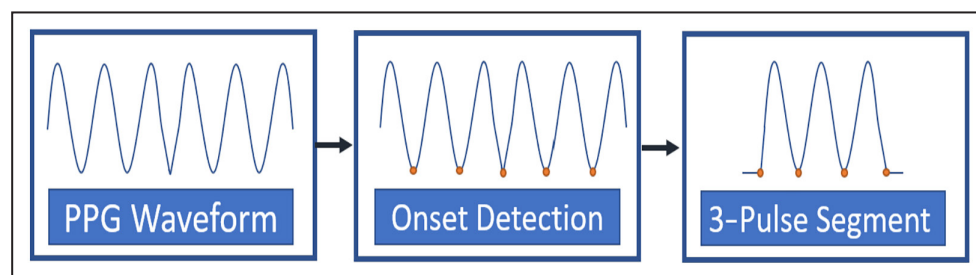


Figure 2. A schematic representation of raw photoplethysmography waveform preprocessing step before feature extraction.

The photoplethysmography waveform is first divided into pulses through onset detection (red dots on the waveform). We then grouped the pulse wave with its neighbors, forming a 3-pulse segment. The extracted features are taken from this 3-pulse segment. The prediction is assigned to the center pulse. PPG indicates photoplethysmography. Copyright Heather Siefkes.

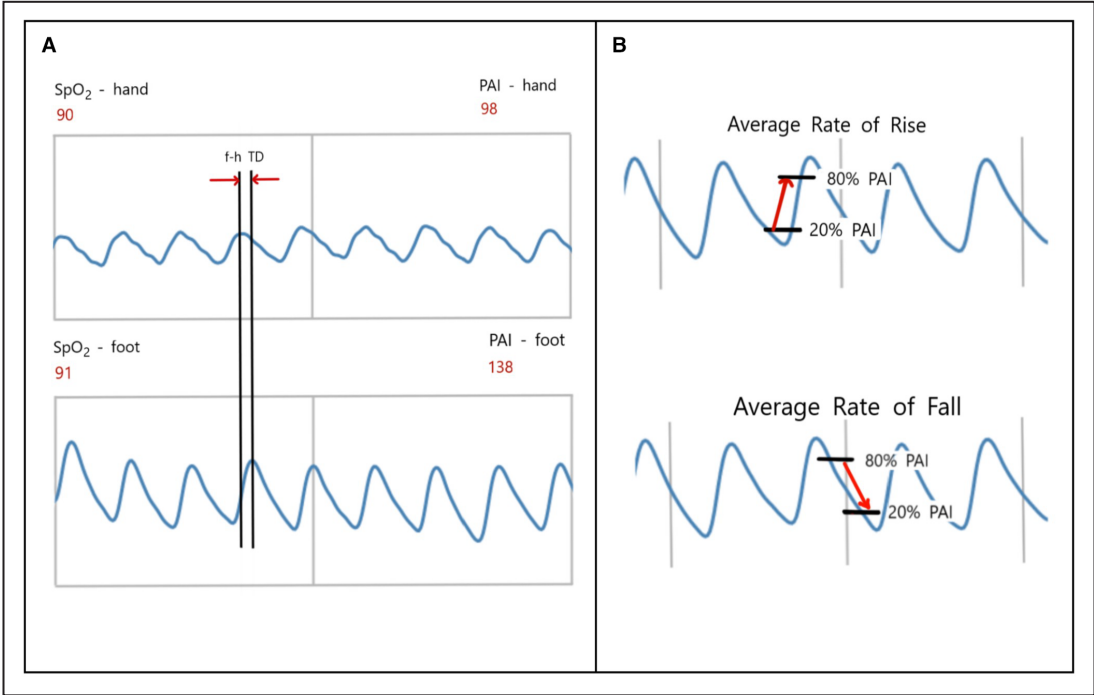


Figure 3. An example of the measurements extracted from artifact-free segments and their visual representations. The measurement depicted in (A) is radiofemoral pulse-delay, which is calculated from overlapping artifact-free segments from the foot and hand. Average rate of rise and fall as shown in (B) is measured from artifact-free photoplethysmography segments. PAI indicates pulse amplitude index; and SpO₂, oxygen saturation. Partially reproduced from Doshi et al¹⁰ under the terms and conditions of the Creative Commons Attribution-Non Commercial No Derivatives license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

combined. We used a feature selection step to prevent performance drop caused by including too many features, that is, the curse of dimensionality.¹³ We used a 3-stage feature selection mechanism followed by recursive feature elimination (RFE), similar to existing works in the literature.¹⁴ First, redundant features were observed through a Pearson correlation test (correlation coefficient >0.8, which denotes strong correlation).¹⁵ We then used χ^2 univariate analysis and removed the correlated features with the worst score. We then calculated the feature importance through the Gini index (random forest wrapped) to find the most relevant features. To select the

most relevant features and save computational time, we took the union of the 35% best-ranked features from χ^2 and Gini index analysis. After removing unimportant and redundant features, we tuned our model through k-fold cross-validation (k=5). We applied stratified data splitting for our 5 folds to ensure a similar ratio of CCHD to no CHD in the validation and training sets of each fold. In this validation step, we tuned the selected features and selected the best-performing ML model. As seen in Table 1, we evaluated logistic regression, random forest, gradient boosting, decision trees, and XGBoost models. We used RFE for

Table 1. Different Machine Learning Model Performance for k-Fold Cross-Validation (k=5) on CCHD Detection for 0–24 and 24–48 Hours With Pulse Delay With and Without Feature Selection Steps

Model used	No feature selection		Feature selection	
	Specificity for no-CHD*±SD	Sensitivity to CCHD*±SD	Specificity for no-CHD*±SD	Sensitivity to CCHD*±SD
Logistic regression	100±0.00	73.33±24.95	100±0.00	93.33±13.33
Random forest	100±0.00	86.67±16.33	100±0.00	79.99±16.33
Gradient boosting	98.46±3.08	60.00±38.87	98.46±3.08	80.00±26.67
Decision trees	98.46±3.08	73.33±32.66	98.46±3.08	80.00±26.67
XGBoost	98.46±3.08	66.67±42.16	98.46±3.08	66.67±42.16

CHD indicates congenital heart disease; and CCHD, critical congenital heart disease.
*The average specificity or sensitivity across the 5 folds for each model is shown with the SDs.

feature selection. RFE is a feature selection commonly used in ML; it identifies the features that maximize a selected performance metric on a prediction task. We chose sensitivity as the performance metric for RFE evaluation.¹⁶ RFE uses the ML model's feature importance ranking to identify the least important features. The least important feature was then eliminated, and sensitivity was evaluated. The process of removing the least important feature was repeated until sensitivity was no longer improved. The proposed feature selection methodology improves the performance of the majority of the classification algorithms. Furthermore, we observe that our best performing setting is logistic regression with feature selection steps. We found many features commonly selected through RFE; see Table 2 for a full list of features selected for each model.

We extracted features from newborns in 0 to 24 hours of age, 24 to 48 hours of age, and >48 hours of age. We trained ML models to evaluate newborns using 0 to 24 hours features and models to evaluate an ensemble of 0 to 24 hours and 24 to 48 hours features. The set of features selected by each model was

different, as they underwent RFE separately. Table 2 summarizes the selected and included features for the trained models (0–24 hours, and 0–24+24–48 hours). Once feature selection was completed, we tuned the ML models' prediction confidence threshold to maximize specificity for no CHD over the 5 folds tested. Our ML code is publicly available at the following link: <https://github.com/ucdrubinet/CCHD-ML-public>.

ML Performance Evaluation

Due to the cohort size limitations, our study was not able to build a separate holdout test set. We then reported the performance of our model on the 5 validation folds. The 5-fold cross-validation algorithm selects 80% of the set for training and the other 20% for validation, with no overlapping data between the training and validation sets in each fold. Every data point was in the validation set in 1-fold.

Using the same cohort, we also randomly selected 55 subjects for training and 24 subjects for testing (70/30 split) in 100 different configurations. We applied

Table 2. CCHD Detection Features Selected and Included by Recursive Feature Elimination

Measurement	Region	Extracted features	0- to 24-h model*	0- to 24-+24- to 48-h model*
SpO ₂	Hand	Mean	✓	✓
		Median	✓	✓
		Maximum	✓	✓
		Minimum		✓
		Variance	✓	✓
	Foot	Mean	✓	✓
		Median	✓	✓
		Maximum	✓	✓
		Variance		
Perfusion amplitude index	Hand	Mean	✓	✓
		Median	✓	✓
		Minimum	✓	✓
		Variance		✓
	Foot	Mean		✓
		Median		✓
		Maximum		✓
Heart rate	Hand	Mean		✓
		Median		✓
		Maximum	✓	
	Foot	Median		✓
		Maximum		✓
		Minimum	✓	✓
		Variance		✓
Radiofemoral delay	Hand and foot	Minimum	✓	✓
		Variance	✓	✓

CCHD indicates critical congenital heart disease; and SpO₂, oxygen saturation.

*The model(s) for which the within-the-row feature was selected by recursive feature elimination is listed. The "0- to 24-h" is in reference to 1 time point within 0 to 24 h of age. The "0- to 24-+24- to 48-h" is in reference to 2 time points, one collected within 0 to 24 h and the other within 24 to 48 h of age.

stratified splitting, to ensure a similar ratio of CCHD to no CHD in training and testing sets. We applied the same features and prediction confidence threshold established in cross-validation and did not perform any tuning. We reported the average performance over the 100 different random configurations.

Comparing ML With SpO₂ Alone

We then compared the pass/fail results of the SpO₂-alone algorithms to the ML results using McNemar tests to compare sensitivity and specificity. We compared sensitivity for CCHD and for CoA. We calculated area under the receiver operating characteristic curve (AUROC) for binary outcomes and compared with SpO₂ using the rccomp command in STATA (StataCorp, College Station, TX). When comparing sensitivity, specificity, and AUROC for CCHD, we used all patients with CCHD and those without CHD.

Sample Size

Our targeted sample size was based on CoA detection with a minimum of 200 healthy newborns and 20 with critical CoA. We confirmed that this sample size would ensure sufficient power (81%) when the ML model provides clinically significant improvement in detection for CoA. SpO₂-alone screening has a near-perfect specificity

but sensitivity of $\approx 36\%$ for critical CoA,² a value that corresponds to AUROC of 68% (AUROC=0.5×[Sensitivity+Specificity]).¹⁷ We considered a clinically significant improvement in discriminative capacity to obtain an AUROC of 85%, which corresponds to improving sensitivity to 70% for a cutoff that achieves 99.9% specificity.

RESULTS

A total of 553 patients were enrolled, of whom 20 were withdrawn (ineligible, 2; consenting error, 1; unable to obtain measurements after consent, 1; consenting ability of parents changed after initial consent and unable to determine follow-up contact, 3; unable to obtain written consent after the first measurement for CHD patients, 12; parent request, 1) and 10 (1.8%) were lost to follow-up, resulting in 523 included in the analysis (no CHD, 317; CHD, 74; CCHD, 132). There were 21 infants with critical CoA included in the analysis. Site enrollment varied, with site A enrolling the most patients (N=327), followed by sites B (N=78), C (N=56), D (N=26), E (N=26), and F (N=10). Demographic data for enrolled patients are shown in Table 3.

For those with CHD, 95.1% (196/206) were diagnosed between birth (including prenatal) and 48 hours of age (Figure 4). Two-thirds of patients with CCHD (88/132) were suspected prenatally. Most infants

Table 3. Demographic Details for Enrolled Patients With and Without CHD

	No CHD, n=317	All CHD, n=206	CCHD, n=132	Critical CoA, n=21
Site, n (%)				
A	247 (77.9)	80 (38.8)	40 (30.3)	6 (28.6)
B	33 (10.4)	45 (21.8)	25 (18.9)	6 (28.6)
C	20 (6.3)	36 (17.5)	26 (19.7)	0
D	0	26 (12.6)	24 (18.2)	8 (38.1)
E	15 (4.7)	11 (5.3)	9 (6.8)	0
F	2 (0.6)	8 (3.9)	8 (6.1)	1 (4.8)
Female, n (%)	144 (45.6)	94 (45.6)	58 (43.9)	10 (47.6)
Race, n (%)				
White	148 (46.8)	96 (46.6)	63 (47.7)	15 (71.4)
Asian	42 (13.3)	20 (9.7)	12 (9.1)	3 (14.3)
Black	41 (13)	21 (10.2)	13 (9.8)	0
Native Hawaiian/Other Pacific Islander	6 (1.9)	2 (1)	2 (1.5)	0
American Indian/Alaska Native	1 (0.3)	1 (0.5)	0	0
>1	9 (2.8)	20 (9.7)	11 (8.3)	1 (4.8)
Unknown	69 (21.8)	46 (22.3)	31 (23.5)	2 (9.5)
Ethnicity, n (%)				
Hispanic	79 (25)	67 (32.5)	46 (34.8)	5 (23.8)
Gestational age, wk, mean±SD	39±1.5	38±1.9	38.2±1.8	38.5±1.2
Birth weight, g, mean±SD	3335±695	3034±647	3100±790	3131±566
Cesarean section delivery, n (%)	134 (42.4)	91 (44.9)	53 (41.1)	9 (42.9)
Family history of CHD, n (%)	16 (5.0)	18 (8.8)	16 (12.1)	1 (4.8)

Frequencies shown are within column frequencies. CCHD indicates critical congenital heart disease; CHD, congenital heart disease; and CoA, coarctation of the aorta.

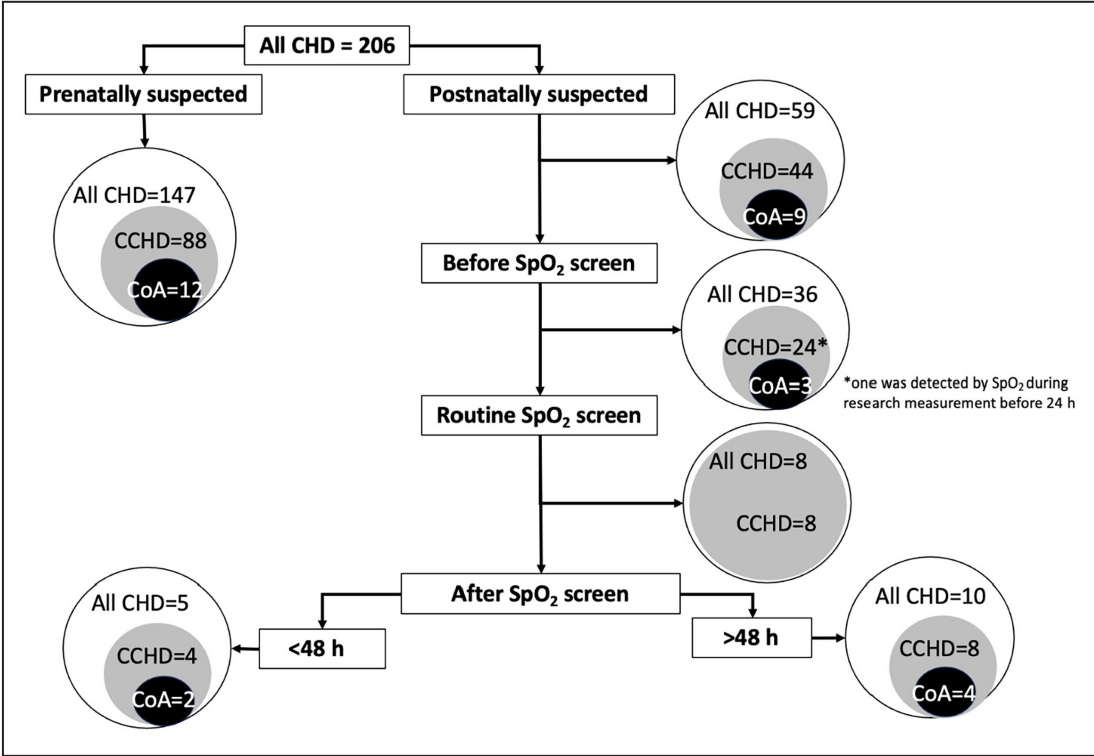


Figure 4. Flowchart of neonates with congenital heart disease and timing of diagnosis. CCHD indicates critical congenital heart disease; CHD, congenital heart disease; CoA, coarctation of the aorta; and SpO₂, oxygen saturation. Copyright Heather Siefkes.

(54.5%) with postnatally diagnosed CCHD were diagnosed before 24 hours of age due to clinical symptoms. Six percent of all CCHDs were diagnosed due to a failed SpO₂ screen, which accounted for 18% of newborns diagnosed postnatally. One patient was diagnosed due to abnormal SpO₂ values during the first research measurement collection, which was done before 24 hours of age and thus before the routine SpO₂ screening. Of the

21 newborns with critical CoA, 4 (19%) were detected after 48 hours of age (3 diagnosed after discharge), which made up half of the 8 newborns with CCHD who were diagnosed after 48 hours (and half of the 6 CCHDs diagnosed after discharge). Among the newborns with CoA diagnosed postnatally, 44% were detected after 48 hours. For all CCHDs, when excluding those prenatally suspected or who developed symptoms within 24 hours, less than half (42.8%, 9/21) would have been detected by the routine SpO₂ alone.

Table 4. Results for SpO₂ Screen Algorithm

	No CHD, n=257 (%)	All CHD, n=191* (%)	CCHD, n=122* (%)	Critical CoA, n=17* (%)
SpO ₂ screen results starting with first SpO ₂ after 24 h of age				
Pass	244 (94.94)	79 (41.36)	32 (26.23)	9 (52.94)
Fail	0	75 (39.27)	65 (53.28)	3 (17.65)
Repeat	13 (5.06)	37 (19.37)	25 (20.49)	5 (29.41)
Conservative SpO ₂ screen results starting with first SpO ₂ after 24 h of age (assigning last "repeats" as fails)				
Fail	13 (5.06)	112 (58.64)	90 (73.77)	8 (47.06)

Description of the SpO₂ algorithm assignments: We used single pre- and postductal SpO₂ values documented as point-of-care values during the collected data points to assign pass/fail SpO₂-based screening results. We used the first value after 24 h of age as the start of the algorithm. If the last value available was a "repeat," we conservatively assigned the infant a "fail." CCHD indicates critical congenital heart disease; CHD, congenital heart disease; CoA, coarctation of the aorta; and SpO₂, oxygen saturation. *The total number for groups do not equal all patients enrolled, as not all patients had research SpO₂ values collected after 24 h of age.

SpO₂ Screening Results

Table 4 shows the SpO₂ screen results for our cohort when starting with the first available SpO₂ after 24 hours of age. It also shows the results for the "conservative" assignment, which assigned a fail to any "repeats" that did not have a repeated SpO₂ value available. When using the recommended SpO₂ screening approach, 26.23% infants with CCHD would not have been detected. Fifty-three percent of the patients with CoA would not have been identified by the SpO₂ screening event when conservatively assigning a fail to any screen with a final repeat. When excluding the final repeats, all infants without CHD passed the recommended SpO₂ screen. A total of 949 pulse oximetry recordings were collected among the included 523 patients. Due to a software change midstudy, only 335 patients (64%) had

recordings with millisecond time stamps that allowed for calculation of radiofemoral pulse delay. Not all patients had pulse oximetry recordings collected for all time periods (0–24 hours, 24–48 hours, and >48 hours). The period for the measurements varied between patients with and without CHD (Table 5). More patients without CHD had measurements collected between 0 and 24 hours of age (71.3% of no-CHD patients) and 24 to 48 hours of age (76.7%). More patients with CHD (82.5%), including CCHD (81.8%) and critical CoA (71.4%) had measurements collected after 48 hours of age.

ML CCHD Detection Results (5-Fold Analysis)

Due to lack of pulse-delay data on some of our cohort and the need for simultaneous artifact-free photoplethysmography hand and foot waveforms to calculate the radiofemoral pulse delay, we sought to evaluate the potential benefit from including it as a feature. We focused on earlier detection within 48 hours of age, using 1 time point versus 2 time points, thus ultimately testing 4 models (Table 6). The analysis was restricted to include only those patients (N=79) who had data allowing them to be in all 4 models. The 2-time-point model performed better than the 1-time-point model both with and without pulse delay included. Additionally, including pulse delay improved sensitivity for CCHD for both the 1- and 2-time-point models. All 4 models had 100% specificity. The best-performing model included 2 time points and pulse delay and improved CCHD detection to 92.86% from the reference SpO₂-alone 71.43%, although this difference was not statistically significant. Detection of CoA also improved from 0% to 66.67% but was not statistically significant on this small sample (CoA, 3). The AUROCs for both CCHD and CoA detection versus infants without CHD significantly improved compared with SpO₂-alone using the 2-time-point with pulse-delay model (Table 3).

There were 10 patients with prenatally suspected CoA who were ultimately determined to be normal. All 4 of our ML models correctly classified 100% (8 included in the ML cohort) as infants without CHD.

70/30 Split Performance

As seen in Table 7, the performance of the ML models on the 70/30 test splits followed a similar trend as the 5-folds results. We again observed an increase in sensitivity and AUROC when adding pulse delay and an increase in sensitivity and AUROC when we added a second time point (24–48 hours). In every test iteration, the 2-time-point model with pulse delay maintained 100% specificity. Additionally, the 2-time-point model with pulse delay had 100% accuracy in 48 of the 100 iterations.

DISCUSSION

CoA remains the most common defect detected after 48 hours of age and nearly half of the patients postnatally diagnosed with CoA were missed during the first 48 hours after birth. With ML techniques, our automated algorithm filtered out artifact, combined pre- and postductal pulse oximetry measurements, and used all photoplethysmography features for CCHD screening and improved CCHD and CoA AUROCs significantly compared with SpO₂ alone. Our best-performing algorithm combined measurements from 2 time points (0–24 hours and 24–48 hours) and included radiofemoral pulse delay as a feature.

Early diagnosis of CCHD is crucial to improve outcomes, as late detection is associated with higher mortality rates, with up to 27% with late diagnosis dying.²⁰ A common definition of timely versus late detection is whether or not the infant is diagnosed before leaving the hospital after birth. Before universal SpO₂ screening, up to 25% of infants with CCHD left the hospital undiagnosed, and another 5% were diagnosed only at autopsy.²¹ SpO₂ screening has been shown to have higher sensitivity than physical examination alone for CCHD detection¹ and thus has been widely adopted throughout the United States and is becoming more commonly used worldwide.¹¹ However, the impact of SpO₂ screening on death and early detection appears to be lower than prior predictions or estimates. Abouk et al previously estimated that mandated SpO₂

Table 5. Time Period of Pulse Oximetry Measurements for Patients With and Without CHD

	No CHD, n=317 (%)	All CHD, n=206 (%)	CCHD, n=132 (%)	Critical CoA, n=21 (%)
Pulse oximetry measurement time period collected*				
0–24 h of age	225 (71.3)	102 (49.5)	60 (45.5)	8 (36.1)
24–48 h of age	244 (76.7)	115 (55.8)	71 (53.8)	6 (28.6)
>48 h of age	93 (29.3)	170 (82.5)	108 (81.8)	15 (71.4)
2 time periods collected (0–24 and 24–48 h of age)	160 (50.5)	81 (39.3)	46 (34.8)	4 (19)
All 3 time periods collected	33 (10.4)	64 (31.1)	37 (28)	3 (14.3)

CCHD indicates critical congenital heart disease; CHD, congenital heart disease; and CoA, coarctation of the aorta.

*Column frequencies for pulse oximetry measurements obtained during the 3 time periods totals >100% because infants could have measurements during >1.

Table 6. SpO₂ Results Compared With Machine Learning Results for CCHD Detection

	SpO ₂ alone* No-CHD=65, CCHD=14 (including CoA=3)	Pulse oximetry machine learning algorithms† No-CHD=65, CCHD=14 (including CoA=3)			
		No pulse delay		With pulse delay	
		1-time 0- to 24-h	2-time 0- to 24-+24- to 48-h	1-time 0- to 24-h	2-time 0- to 24-+24- to 48-h
Sensitivity for CCHD (including CoA), %	71.43 <i>reference</i>	71.43; <i>P</i> >0.9 (95% CI, -0.3 to 0.3)	85.71; <i>P</i> =0.5 (95% CI, -0.15 to 0.43)	78.57; <i>P</i> >0.99 (95% CI, -0.19 to 0.34)	92.86; <i>P</i> =0.25 (95% CI, -0.11 to 0.51)
Specificity for no-CHD, %	95.38 <i>reference</i>	100; <i>P</i> =0.25 (95% CI, -0.02 to 0.13)	100; <i>P</i> =0.25 (95% CI, -0.02 to 0.13)	100; <i>P</i> =0.25 (95% CI, -0.02 to 0.13)	100; <i>P</i> =0.25 (95% CI, -0.02 to 0.13)
AUROC (CCHD vs no-CHD)	0.83 (95% CI, 0.71 to 0.96) <i>reference</i>	0.86 (0.73 to 0.98); <i>P</i> =0.67	0.93 (0.83 to 1); <i>P</i> =0.06	0.89 (0.78 to 1); <i>P</i> =0.36	0.96 (0.89 to 1); <i>P</i> =0.03
Sensitivity for CoA (vs no-CHD), %	0 <i>reference</i>	0	33.3; <i>P</i> >0.99 (95% CI, -0.63 to 0.91)	33.3; <i>P</i> >0.99 (95% CI, -0.63 to 0.91)	66.67; <i>P</i> =0.5 (95% CI, -0.54 to 0.99)
AUROC (CoA vs no-CHD)	0.48 (95% CI, 0.45 to 0.5); <i>reference</i>	0.5 (0.47 to 0.53)‡; <i>P</i> =0.08	0.67 (0.34 to 1); <i>P</i> =0.26	0.67 (0.34 to 1); <i>P</i> =0.26	0.83 (0.51 to 1); <i>P</i> =0.03

P values in each cell concern within-row between-column comparisons for the row parameter, comparing the parameter estimate in the given cell with the parameter value in the leftmost column for that row (the reference cell). For these paired comparisons, we used McNemar's test for sensitivity and specificity, and the STATA roccomp command for AUROCs (which implements the nonparametric method proposed by De Long et al).¹⁸ An exception is noted below. The CIs reported for paired comparisons of sensitivity/specificity parameters were obtained using the McNemarExactDP command from the exact2x2 package in RStudio (R Foundation for Statistical Computing, Vienna, Austria), which implements the method proposed by Fay and Lumbard.¹⁹ AUROC indicates area under the receiver operating characteristic curve; CCHD, critical congenital heart disease; CoA, coarctation of the aorta; and SpO₂, oxygen saturation.

*Conservatively assigns a fail to any patients whose last measurement available would have prompted a "repeat." Uses the first measurement after 24 h of age as the first measurement to follow the algorithm.

†Four machine learning algorithms are displayed. The differences are presence/absence of radiofemoral pulse delay, and 1 time point (between 0 and 24 h of age) vs 2 time points (between 0–24 h and 24–48 h age).

‡For this cell, we could not compute the AUROC or compare it with the reference AUROC using the STATA roccomp command because the machine learning algorithm within this column did not vary (it labeled all patients as "healthy"). Therefore, to report an AUROC and a paired AUROC comparison, we took advantage of the equation for a binary test result where its AUROC=0.5×(Sensitivity+Specificity)¹⁷ and created a derived variable whose mean would thus be equal to the AUROC. For each of the 4 possible combinations of test results and disease status, the values of the derived variable were as follows: If test+ and disease+, the value was 0.5/proportion disease+. If test- and disease-, the value was 0.5/proportion disease-. For the 2 other combinations, the value was 0. We were able to compute the mean and the 95% CI for this derived variable to report the AUROC for the test. In addition, we were able to create the corresponding derived variable for the reference test and then compute the mean and 95% CI for the within-individual differences for these 2 derived variables to get valid estimates for the change in the AUROC.

screening was associated with a 33.4% reduction in infant deaths from CCHD.²² Presumably, this reduction was secondary to early detection; however, due to use of administrative registry data, Abouk et al were not able to determine timing of CCHD detection. Subsequent studies using clinical registries, including timing and mechanisms leading to CCHD detection, have shown that SpO₂ screening has not led to earlier detection for infants who are postnatally diagnosed with CCHD or been associated with a reduction in death.^{23,24} Instead, it has been noted that there have been increases in prenatal detection during this time and thus brought the question of the utility of SpO₂

screening in areas with high prenatal detection.²³ While we did not assess CCHD detection rates over time, we observed that among postnatally suspected cases, 8 of 44 (18%) of CCHD, including 4 of 9 (44%) CoA, were detected >48 hours of age.

Similar to published reports, we have shown CoA to be the most common defect diagnosed late, often after discharge, despite SpO₂ screening.^{23–25} Ailes et al previously estimated that the majority of infants with a late detection of CCHD would be those with CoA: an estimated 560 missed annually in the United States despite universal SpO₂ screening.² This is in part due to the low sensitivity of SpO₂-based screening for CoA,

Table 7. Machine Learning Model Performance on 100 Different 70/30 Test Splits

	Pulse oximetry machine learning algorithms			
	No pulse delay		With pulse delay	
	1 time, 0- to 24-h	2 times, 0- to 24-+24- to 48-h	1 time, 0- to 24-h	2 times, 0- to 24-+24- to 48-h
Sensitivity for CCHD, %	74.5±19.5	79.3±19.2	75.8±19.6	82.0±19.8
Specificity for no-CHD, %	99.6±1.3	99.5±1.7	99.85±0.9	100±0
AUROC (CCHD vs no-CHD)	0.87±0.01	0.89±0.09	0.88±0.01	0.91±0.09

AUROC indicates area under the receiver operating characteristic curve; and CCHD, critical congenital heart disease.

≈36%.²⁶ Unfortunately, CoA is also difficult to detect by prenatal ultrasound.^{27,28} Thus, even when Liberman et al found prenatal detection of CCHD overall to be increasing, CoA remained the most common CCHD with delayed detection. Up to 31% of CoA cases are detected late, accounting for 64% of the infants with a delayed diagnosis.²³ Thus, our study and prior studies demonstrate a continued need to improve postnatal CoA detection. We anticipate an increase in the need for echocardiograms for CCHD screen failure with the ML-based dual pulse oximetry screen compared with traditional SpO₂-based CCHD screen.

Our ML pulse oximetry algorithm that uses SpO₂ and other photoplethysmography data such as PAI and pulse delay during the first 48 hours after birth demonstrated the potential for increased CoA detection. We did not achieve statistical significance in improved sensitivity for CoA, as our analysis sample yielded only 3 CoA cases. The sensitivity for these 3 cases improved from 0 of 3 (0%) to 2 of 3 (67%). Although we enrolled more than our intended 20 patients with CoA, the need to develop and test the ML algorithm on consistent features, including timing of the measurement, significantly lowered our available sample size. For example, there was variation in time points collected between our healthy controls and patients with CCHD. Due to earlier discharge of our healthy controls, the majority of their measurements were within the first 48 hours of age. Patients with CCHD enrolled in our study were more likely to have measurements after 48 hours of age, commonly due to timing when the research team would know of these patients or due to patients transferring from other hospitals. A similar pattern has been seen in prior studies that have attempted to evaluate nonoxygenation pulse oximetry features as screening tools for CoA, with most measurements collected after 48 hours of age.^{7,8}

The use of ML pulse oximetry for CCHD screening is appealing as it uses the nonoxygenation data that are readily available and currently unused during the routine SpO₂-alone screen. Instead of a spot sequential check of preductal followed by postductal oximetry, our ML algorithm eliminates artifact, analyses SpO₂ and other photoplethysmography data over a period of time, and improves detection of CCHD, especially CoA. Our best-performing ML algorithm, however, would require 2 time points, which is different than the current recommended screen. The current recommended screen includes the potential for a second measurement but only dependent on the results of the first.¹¹ In fact, most patients do not require the potential second screen, as the majority either pass or fail the test with a single measurement.²⁹ In the ML-based algorithm, the second measurement would occur before 24 hours of age, theoretically enhancing the potential for early diagnosis.

The addition of a required measurement before 24 hours of age in our algorithm could result in earlier

detection, and interestingly, 1 patient in our study who was enrolled as a control was diagnosed with CCHD before 24 hours because the SpO₂ values were noted to be low during our study measurement. The timing of SpO₂ screening has been shown to be more sensitive if done within 24 hours after birth but with a higher false-positive rate.³⁰ While the false-positive rate is higher, it is important to note that most of those patients are found to have a condition other than CCHD, such as noncritical CHD, pneumonia, or sepsis, that required treatment.^{11,31,32} All of our algorithms, including the single time point within 24 hours of age, had 100% specificity suggesting that an early ML-based screen may not be associated with a high false-positive rate. We will need to validate this ML-based algorithm in larger studies.

Another potential benefit of our algorithm warranting further evaluation is the ability to correctly classify patients with prenatally suspected CoA that are postnatally determined to be without CCHD. Prenatal ultrasound has a high false-positive rate for isolated CoA, as high as 94%.³³ Patients with prenatally suspected critical CoA are often admitted to the neonatal intensive care unit for monitoring while the ductus arteriosus closes, leading to delayed maternal bonding, delayed feeding, and longer length of stay (25% staying ≥9 days) compared with healthy controls and a growing interest to adjust protocols for monitoring of newborns with prenatally suspected CoA.^{33,34} Of the 8 infants with a prenatally suspected CoA who were then determined to be healthy and with data to be included in our ML analysis algorithm, our ML algorithms correctly identified all as healthy infants using measurements before 48 hours of age. However, before these algorithms are used to help rule out CoA following prenatal suspicion, the sensitivity for CoA needs to be further validated in larger studies.

Finally, the features included in ML pulse oximetry need to be further evaluated. Our models included SpO₂, the current standard for CCHD screening. However, our model included both maximum, minimum, mean, median, and variance of SpO₂, which may provide a better clinical representation of the physiology than just the maximum. Heart rate was also selected, including heart rate variability, which has been found to be abnormal in infants with CCHD.³⁵ PAI of both hand and foot were selected. PAI is synonymous with perfusion index, which has been previously noted to be abnormal in infants with CoA and possibly other CCHD.^{4–7} We noted improved detection with the addition of pulse delay, which has been previously found to be abnormal in patients with CoA.⁸ Our model did not select photoplethysmography slopes, which have been previously shown to be abnormal in CoA.⁹ We suspect this is due to elimination of correlated features and the correlation between slope and PAI.

Limitations

There are several limitations to our study. While we were able to enroll 132 patients with CCHD and 21 with isolated CoA, ultimately the sample included in the ML development and testing was small and underpowered. However, we did note improved AUROC in both CCHD and CoA detection. In ML development, this small sample size may have overfit our model and may overestimate CCHD sensitivity. Additionally, we were not able to test our algorithm on a true holdout test. Instead, we performed 70/30 splits on 100 iterations to test the generalizability of our algorithm, which performed well with 100% accuracy in nearly half of the iterations in our 2-time-point with pulse-delay algorithm. The model, however, will need to be tested in a true holdout sample, which we are currently conducting. We did not collect data on skin pigmentation, and our model was developed on a population that was predominantly White or unknown race. This is an important limitation to note due to concerns regarding pulse oximetry SpO_2 inaccuracies in darker-pigmented patients.^{36–38} It is important that the effect of skin pigmentation is evaluated further and that any algorithm or device created is developed using data from a diverse cohort. Finally, the majority of our patients with CCHD were receiving prostaglandin therapy to maintain ductus arteriosus patency. While it is crucial that CCHD be detected when the ductus remains patent, it is important to note that a lower perfusion index or PAI may be a marker of PDA and increases after PDA treatment.³⁹ However, the perfusion index values noted in the PDA group by Sangsari et al³⁹ were still higher than most suggested thresholds for CCHD detection.^{4–7,40} Nonetheless, this will need to be evaluated further as well as how the other features included in our ML algorithm correlate with PDA.

CONCLUSIONS

In conclusion, our ML algorithm combining 2 time points within 48 hours of age, using features of oxygenation, radiofemoral pulse delay, and perfusion increased AUROCs for CCHD and CoA. The model needs to be tested in larger cohorts and a true holdout cohort to further validate our results. Feasibility and implementation of our approach will be evaluated in future studies as well.

ARTICLE INFORMATION

Received December 14, 2023; accepted March 20, 2024.

Affiliations

Department of Pediatrics (H.S., D.T., S.L.), and Department of Electrical & Computer Engineering (L.C.O., Z.L., C-N.C.), University of California, Davis, CA, Department of Pediatrics, Cohen Children's Medical Center, Zucker School of Medicine at Hofstra/Northwell, New Hyde Park, NY (R.K.);

University of Utah, Primary Children's Hospital, Salt Lake City, UT (W.H.); Department of Pediatrics, University of California, Los Angeles, CA (M.G.); Department of Pediatrics, Sutter Sacramento Medical Center, Sacramento, CA (E.M.); and Department of Pediatrics, University of California, San Francisco, CA (N.C.).

Acknowledgments

The authors thank their patients for their involvement in this study, as well as the clinical research coordinators involved in the study.

Sources of Funding

The project described was supported by the National Center for Advancing Translational Sciences, National Institutes of Health, through grant number UL1 TR001860 and linked award KL2 TR001859; the Eunice Kennedy Shriver National Institute of Child Health & Human Development, National Institutes of Health, through grant number 1R21HD099239-01; and the University of California, Davis Artificial Intelligence Seed Grant, and Venture Catalyst DIAL Grant. Additional support was provided by the Doris Duke Charitable Foundation COVID-19 Fund to Retain Clinical Scientists awarded to University of California, Davis School of Medicine by the Burroughs Wellcome Fund. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health or the University of California, Davis.

Disclosures

Drs Siefkes and Lakshminrusimha and C.-N. Chuah, and Z. Lai are listed as inventors on a pending patent owned by University of California, Davis regarding technology discussed in this study. Dr Siefkes is the founder of NeoPOSE, a university associated start-up company to develop technology aimed at improving CCHD detection. The remaining authors have no disclosures to report.

REFERENCES

- de-Wahl Granelli A, Wennergren M, Sandberg K, Mellander M, Bejlum C, Inganas L, Eriksson M, Segerdahl N, Agren A, Ekman-Joelsson B, et al. Impact of pulse oximetry screening on the detection of duct dependent congenital heart disease: a Swedish prospective screening study in 39 821 newborns. *BMJ*. 2008;337:a3037. doi: [10.1136/bmj.a3037](https://doi.org/10.1136/bmj.a3037)
- Ailes EC, Gilboa SM, Honein MA, Oster ME. Estimated number of infants detected and missed by critical congenital heart defect screening. *Pediatrics*. 2015;135:1000–1008. doi: [10.1542/peds.2014-3662](https://doi.org/10.1542/peds.2014-3662)
- Chang R, Gurvitz M, Rodriguez S. Missed diagnosis of critical congenital heart disease. *Arch Pediatr Adolesc Med*. 2008;162:969. doi: [10.1001/archpedi.162.10.969](https://doi.org/10.1001/archpedi.162.10.969)
- Siefkes H, Kair L, Tancredi DJ, Vasquez B, Garcia L, Bedford-Mu C, Lakshminrusimha S. Oxygen saturation and perfusion index-based enhanced critical congenital heart disease screening. *Am J Perinatol*. 2020;37:158–165. doi: [10.1055/s-0039-1685445](https://doi.org/10.1055/s-0039-1685445)
- Uygur O, Koroglu OA, Levent E, Toslali M, Akisu M, Yalaz M, Kultursay N. The value of peripheral perfusion index measurements for early detection of critical cardiac defects. *Pediatr Neonatol*. 2018;60:1–6. doi: [10.1016/j.pedneo.2018.04.003](https://doi.org/10.1016/j.pedneo.2018.04.003)
- Schena F, Piccioli I, Agosti M, Zuppa AA, Zuccotti G, Parola L, Pomero G, Stival G, Markart M, Graziani S, et al. Perfusion index and pulse oximetry screening for congenital heart defects. *J Pediatr*. 2017;183:74–79. doi: [10.1016/j.jpeds.2016.12.076](https://doi.org/10.1016/j.jpeds.2016.12.076)
- de-Wahl Granelli A, Ostman-Smith I. Noninvasive peripheral perfusion index as a possible tool for screening for critical left heart obstruction. *Acta Paediatr*. 2007;96:1455–1459. doi: [10.1111/j.1651-2227.2007.00439.x](https://doi.org/10.1111/j.1651-2227.2007.00439.x)
- Palmeri L, Gradwohl G, Nitzan M, Hoffman E, Adar Y, Shapir Y, Koppel R. Photoplethysmographic waveform characteristics of newborns with coarctation of the aorta. *J Perinatol*. 2017;37:77–80. doi: [10.1038/jp.2016.162](https://doi.org/10.1038/jp.2016.162)
- Sorenson M, Sadiq I, Clifford G, Maher K, Oster M. Using pulse oximetry waveforms to detect coarctation of the aorta. *Biomed Eng Online*. 2020;19:1–12. doi: [10.1186/s12938-020-00775-2](https://doi.org/10.1186/s12938-020-00775-2)
- Doshi K, Rehm G, Vadlaputi P, Lai Z, Lakshminrusimha S, Chuah C, Siefkes H. A novel system to collect dual pulse oximetry data for critical congenital heart disease screening research. *J Clin Transl Sci*. 2020;5:e56. doi: [10.1017/cts.2020.550](https://doi.org/10.1017/cts.2020.550)
- Martin GR, Ewer AK, Gaviglio A, Hom LA, Saarinen A, Sontag M, Burns KM, Kemper AR, Oster ME. Updated strategies for pulse

- oximetry screening for critical congenital heart disease. *Pediatrics*. 2020;146:e20191650. doi: [10.1542/peds.2019-1650](https://doi.org/10.1542/peds.2019-1650)
12. Oliveira LC, Lai Z, Geng W, Siefkes H, Chuah CN. A machine learning driven pipeline for automated Photoplethysmogram signal artifact detection. In: *2021 IEEE/ACM Conference on Connected Health: Applications, Systems and Engineering Technologies (CHASE)*. IEEE; 2021:149–154.
 13. Pavlenko T. On feature selection, curse-of-dimensionality and error probability in discriminant analysis. *J Stat Plan Inference*. 2003;115:565–584. doi: [10.1016/S0378-3758\(02\)00166-0](https://doi.org/10.1016/S0378-3758(02)00166-0)
 14. Rehm GB, Cortés-Puch I, Kuhn BT, Nguyen J, Fazio SA, Johnson MA, Anderson NR, Chuah C-N, Adams JY. Use of machine learning to screen for acute respiratory distress syndrome using raw ventilator waveform data. *Crit Care Explor*. 2021;3:e0313. doi: [10.1097/CCE.0000000000000313](https://doi.org/10.1097/CCE.0000000000000313)
 15. Akoglu H. User's guide to correlation coefficients. *Turk J Emerg Med*. 2018;18:91–93. doi: [10.1016/j.tjem.2018.08.001](https://doi.org/10.1016/j.tjem.2018.08.001)
 16. Breiman L. Random forests. *Mach Learn*. 2001;45:5–32. doi: [10.1023/A:1010933404324](https://doi.org/10.1023/A:1010933404324)
 17. Cantor SB, Kattan MW. Determining the area under the ROC curve for a binary diagnostic test. *Med Decis Mak*. 2000;20:468–470. doi: [10.1177/0272989X0002000410](https://doi.org/10.1177/0272989X0002000410)
 18. DeLong E, Delong D, Clark-Pearson D. Comparing the areas under two or more correlated receiver operating characteristic curves: a nonparametric approach. *Biometrics*. 1988;44:837–845. doi: [10.2307/2531595](https://doi.org/10.2307/2531595)
 19. Fay MP, Lumbard K. Confidence intervals for difference in proportions for matched pairs compatible with exact McNemar's or sign tests. *Stat Med*. 2021;40:1147–1159. doi: [10.1002/sim.8829](https://doi.org/10.1002/sim.8829)
 20. Eckersley L, Sadler L, Parry E, Finucane K, Gentles TL. Timing of diagnosis affects mortality in critical congenital heart disease. *Arch Dis Child*. 2016;101:516–520. doi: [10.1136/archdischild-2014-307691](https://doi.org/10.1136/archdischild-2014-307691)
 21. Wren C, Reinhardt Z, Khawaja K. Twenty-year trends in diagnosis of life-threatening neonatal cardiovascular malformations. *Arch Dis Child Fetal Neonatal Ed*. 2008;93:F33–F35. doi: [10.1136/adc.2007.119032](https://doi.org/10.1136/adc.2007.119032)
 22. Abouk R, Grosse S, Ailes E, Oster M. Association of US state implementation of newborn screening policies for critical congenital heart disease with early infant cardiac deaths. *JAMA*. 2017;318:2111–2118. doi: [10.1001/jama.2017.17627](https://doi.org/10.1001/jama.2017.17627)
 23. Liberman RF, Heinke D, Lin AE, Nestoridi E, Jalali M, Markenson GR, Sekhvat S, Yazdy MM. Trends in delayed diagnosis of critical congenital heart defects in an era of enhanced screening, 2004–2018. *J Pediatr*. 2023;257:113366. doi: [10.1016/j.jpeds.2023.02.012](https://doi.org/10.1016/j.jpeds.2023.02.012)
 24. Campbell MJ, Quarshie WO, Faerber J, Goldberg DJ, Mascio CE, Blinder JJ. Pulse oximetry screening has not changed timing of diagnosis or mortality of critical congenital heart disease. *Pediatr Cardiol*. 2020;41:899–904. doi: [10.1007/s00246-020-02330-1](https://doi.org/10.1007/s00246-020-02330-1)
 25. Martin GR, Schwartz BN, Hom LA, Donofrio MT. Lessons learned from infants with late detection of critical congenital heart disease. *Pediatr Cardiol*. 2022;43:580–585. doi: [10.1007/s00246-021-02760-5](https://doi.org/10.1007/s00246-021-02760-5)
 26. Prudhoe S, Abu-harb M, Richmond S, Wren C. Neonatal screening for critical cardiovascular anomalies using pulse oximetry. *Arch Child Fetal Neonatal Ed*. 2013;98:346–351. doi: [10.1136/archdischild-2012-302045](https://doi.org/10.1136/archdischild-2012-302045)
 27. Lannering K, Bartos M, Mellander M. Late diagnosis of coarctation despite prenatal ultrasound and postnatal pulse oximetry. *Pediatrics*. 2015;136:e406–e412. doi: [10.1542/peds.2015-1155](https://doi.org/10.1542/peds.2015-1155)
 28. Beattie M, Peyvandi S, Ganesan S, Moon-Grady A. Toward improving the fetal diagnosis of coarctation of the aorta. *Pediatr Cardiol*. 2017;38:344–352. doi: [10.1007/s00246-016-1520-6](https://doi.org/10.1007/s00246-016-1520-6)
 29. Sneeringer MR, Vadlapudi P, Lakshminrusimha S, Siefkes H. Lower pass threshold ($\geq 93\%$) for critical congenital heart disease screening at high altitude prevents repeat screening and reduces false positives. *J Perinatol*. 2022;42:1176–1182. doi: [10.1038/s41372-022-01491-6](https://doi.org/10.1038/s41372-022-01491-6)
 30. Thangaratnam S, Brown K, Zamora J, Khan KS, Ewer AK. Pulse oximetry screening for critical congenital heart defects in asymptomatic newborn babies: a systematic review and meta-analysis. *Lancet*. 2012;379:2459–2464. doi: [10.1016/S0140-6736\(12\)60107-X](https://doi.org/10.1016/S0140-6736(12)60107-X)
 31. Ewer AK, Martin GR. Newborn pulse oximetry screening: which algorithm is best? *Pediatrics*. 2016;138:e20161206. doi: [10.1542/PEDS.2016-1206/60410](https://doi.org/10.1542/PEDS.2016-1206/60410)
 32. Singh A, Rasiyah SV, Ewer AK. The impact of routine predischarge pulse oximetry screening in a regional neonatal unit. *Arch Dis Child Fetal Neonatal Ed*. 2014;99:297–302. doi: [10.1136/archdischild-2013-305657](https://doi.org/10.1136/archdischild-2013-305657)
 33. Hede SV, DeVore G, Satou G, Sklansky M. Neonatal management of prenatally suspected coarctation of the aorta. *Prenat Diagn*. 2020;40:942–948. doi: [10.1002/pd.5696](https://doi.org/10.1002/pd.5696)
 34. Soslow JH, Kavanaugh-Mchugh A, Wang L, Saurers DL, Kaushik N, Killen SAS, Parra DA. A clinical prediction model to estimate the risk for coarctation of the aorta in the presence of a patent ductus arteriosus. *J Am Soc Echocardiogr*. 2013;26:1379–1387. doi: [10.1016/j.echo.2013.08.016](https://doi.org/10.1016/j.echo.2013.08.016)
 35. Mulkey SB, Govindan R, Metzler M, Swisher CB, Hitchings L, Wang Y, Baker R, Larry Maxwell G, Krishnan A, du Plessis AJ. Heart rate variability is depressed in the early transitional period for newborns with complex congenital heart disease. *Clin Auton Res*. 2020;30:165–172. doi: [10.1007/s10286-019-00616-w](https://doi.org/10.1007/s10286-019-00616-w)
 36. Vesoulis Z, Tims A, Lodhi H, Lalos N, Whitehead H. Racial discrepancy in pulse oximeter accuracy in preterm infants. *J Perinatol*. 2021;2021:1–7. doi: [10.1038/s41372-021-01230-3](https://doi.org/10.1038/s41372-021-01230-3)
 37. US Food and Drug Administration. FDA in Brief: FDA warns about limitations and accuracy of pulse oximetry. 2021. Accessed June 16, 2021. <https://www.fda.gov/news-events/fda-brief/fda-brief-fda-warns-about-limitations-and-accuracy-pulse-oximeters>
 38. Sjöding MW, Dickson RP, Iwashyna TJ, Gay SE, Valley TS. Racial bias in pulse oximetry measurement. *N Engl J Med*. 2020;383:2477–2478. doi: [10.1056/NEJMc2029240](https://doi.org/10.1056/NEJMc2029240)
 39. Sangsari R, Dalili H, Kadivar M, Saeedi M, Mirnia K, Fathi A, Hakimelahi J. Evaluation of the relationship between perfusion index and the improvement of patent ductus arteriosus. *Uncorrected Proof Innov J Pediatr*. 2023;33:134709. doi: [10.5812/ijp-134709](https://doi.org/10.5812/ijp-134709)
 40. Jegatheesan P, Nudelman M, Goel K, Song D, Govindaswami B. Perfusion index in healthy newborns during critical congenital heart disease screening at 24 hours: retrospective observational study from the USA. *BMJ Open*. 2017;7:e017580. doi: [10.1136/bmjopen-2017-017580](https://doi.org/10.1136/bmjopen-2017-017580)