2021 43rd Annual International Conference of the
IEEE Engineering in Medicine & Biology Society (EMBC)
Oct 31 - Nov 4, 2021. Virtual Conference

Enhanced Critical Congenital Cardiac Disease Screening by Combining Interpretable Machine Learning Algorithms

Zhengfeng Lai[1], Pranjali Vadlaputi[2], Daniel J. Tancredi[2], Meena Garg[3], Robert I. Koppel[4], Mera Goodman[4], Whitnee Hogan[5], Nicole Cresalia[6], Stephan Juergensen[6], Erlinda Manalo[7], Satyan Lakshminrusimha[2], Chen-Nee Chuah[1], *Fellow, IEEE*, and Heather Siefkes[2]

***Abstract*— Critical Congenital Heart Disease (CCHD) screening that only uses oxygen saturation (SpO2), measured by pulse oximetry, fails to detect an estimated 900 US newborns annually. The addition of other pulse oximetry features such as perfusion index (PIx), heart rate, pulse delay and photoplethysmography characteristics may improve detection of CCHD, especially those with systemic blood flow obstruction such as Coarctation of the Aorta (CoA). To comprehensively study the most relevant features associated with CCHD, we investigated interpretable machine learning (ML) algorithms by using Recursive Feature Elimination (RFE) to identify an optimal subset of features. We then incorporated the trained ML models into the current SpO2-alone screening algorithm. Our proposed enhanced CCHD screening system, which adds the ML model, improved sensitivity by approximately 10 percentage points compared to the current standard SpO2-alone method with minimal to no impact on specificity.**

***Clinical relevance*— This establishes proof of concept for a ML algorithm that combines pulse oximetry features to improve detection of CCHD with little impact on false positive rate.**

I. INTRODUCTION

Congenital heart disease (CHD) is the most common birth defect, affecting nearly 0.8% of all newborn infants [1]. Critical congenital heart disease (CCHD) is a subset of CHD accounting for almost 20% of these infants [1], representing the most severe forms of CHD. CCHD lesions require surgical or catheter-based intervention soon after birth, often including pre-procedural hospitalization and medical management. Late or missed detection of CCHD can lead to significant, preventable morbidity, as well as death [2]–[4]. Prior to mandatory oxygen-saturation (SpO2) based CCHD screening, 25% of newborns with CCHD were diagnosed after going home from the hospital [4], [5]. Although, SpO2 screening has helped with earlier diagnosis and reduced CCHD mortality, it still misses approximately 900 newborns with CCHD annually in the United States [6], [7]. The majority of the missed types of CCHD defects are those with obstructed systemic blood flow that do not commonly cause low SpO2, or hypoxemia [6].

To address this problem, we created an automated real-time data collection system [8] to collect additional pulse oximetry data in newborns, allowing us to analyze other pulse oximetry features that may augment the current screening process when added to the SpO2 screening component.

As such, it is necessary to design an interpretable machine learning model that can be directly incorporated into our current SpO2-alone screening system [9] with automatic feature selection to further improve the sensitivity of CCHD detection with little impact on specificity (at least 99%). To our best knowledge, our work is the first to analyze the feature relevance of CCHD screening by using machine learning (ML) algorithms as well as the first to incorporate ML into current standard SpO2 screening.

II. METHODS

A. Subjects

By using an automated collection system [8], we enrolled 335 newborns, including 236 newborns that have a final diagnosis (with or without CCHD) confirmed. Patients were excluded if they required vasoactive infusions other than Prostaglandin E1. The majority of patients were enrolled at University of California (UC), Davis. Four other hospitals, Sutter Medical Center in Sacramento, UC Los Angeles, UC San Francisco, and Cohen Children's Medical Center in New York, are also enrolling patients for this study. We recorded at least 5-minute dual limb (right hand and any foot) pulse oximetry measurements at three time periods: within 24 hours, 24-48 hours, and after 48 hours following the baby's birth. We analyzed healthy newborns (defined as those without any CHD) vs those with CCHD (newborns who require a surgical or catheter-based intervention within 30 days of age). We divided all measurements into two groups: (G1) 0-48 hours, which included 158 healthy and 27 CCHD newborns; and (G2) over 48 hours, which included 50 healthy and 36 CCHD newborns. The experimental procedures involving human subjects described in this paper were approved by the Institutional Review Board.

[1]Z. Lai, C.-N. Chuah are with the Electrical and Computer Engineering Department, University of California Davis, Davis, CA, USA. {lzhengfeng, chuah}@ucdavis.edu

[2]P. Vadlaputi, D. Tancredi, S. Lakshminrusimha, H. Siefkes are with the Pediatrics Department, University of California Davis, Sacramento, CA , USA. {ppvadlaputi, djtancredi, slakshmi, hsiefkes}@ucdavis.edu

[3]M. Garg is with David Geffen school of Medicine UCLA, Los Angeles, CA, USA. mgarg@mednet.ucla.edu

[4]R. I. Koppel, M. Goodman are with Cohen Children's Medical Center, NY, USA. {rkoppel, Mgoodman3}@northwell.edu

[5]W. Hogan is with University of Utah Health, Salt Lake City, UT, USA. whitnee.hogan@gmail.com

[6]N. Cresalia, S. Juergensen are with University of California San Francisco, San Francisco, CA, USA. {Nicole.Cresalia, Stephan.Juergensen}@ucsf.edu

[7]E. Manalo is with Sutter Health, Sacramento, CA, USA. manaloe@sutterhealth.org

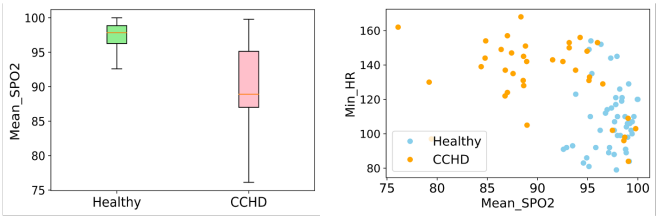
978-1-7281-1178-0/21/$31.00 ©2021 IEEE

B. Spot SpO2-alone Screening

Single pre and post-ductal SpO2 values were recorded during the measurements as well. These spot values were used to assign pass or fail to all newborns per the current standard SpO2-alone screen [9]. If the last recorded spot SpO2 measurements resulted in a ”repeat” assignment from this algorithm, we assigned them a ”fail” in a ”Conservative spot SpO2-alone” algorithm to bias towards the null for CCHD detection.

C. Features Extraction & Analysis

Pulse oximetry features evaluated for discrimination of healthy vs CCHD include: heart rate (HR), perfusion amplitude index (PAI), also known as perfusion index (PIx), and oxygen saturation (SpO2). We removed values likely associated with artifact: HR larger than 250 and SpO2 larger than 100 (the pulse oximeter assigns a value of 127 for SpO2 when the measurement quality is poor). We then extracted variance, min, max, median and mean for HR, PAI and SpO2. To study the differentiation of these features, we visualized each individually and then their correlation with each other. Fig. 1 illustrates the distribution of the mean SpO2 and its correlation with min HR: the mean SpO2 for the healthy newborns is typically higher than that for the CCHD newborns and the min HR for healthy newborns is typically less than that for CCHD newborns.

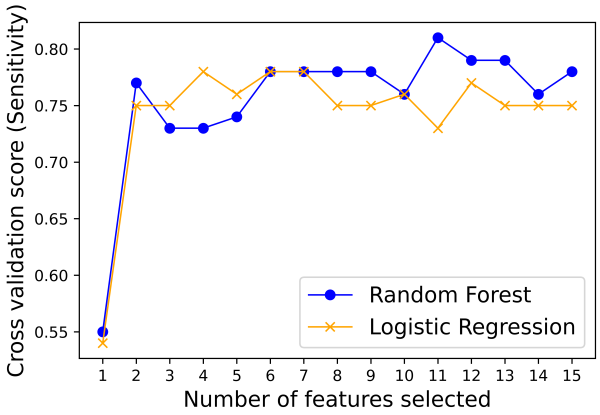


(a) Mean SpO2 (b) Mean SpO2 and Min HR

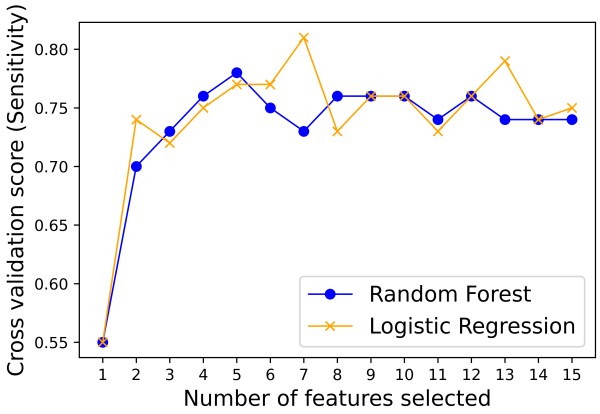
Fig. 1. Feature Analysis between Healthy vs CCHD Over 48 Hour of Age.

D. Classification Algorithms

Several ML classifiers were tested for CCHD detection during this study, including Random Forest, Logistic Regression, and Multilayer Perceptron. Although a recent study [10] claimed random forest classifier has the best performance, we comprehensively investigated the above classification algorithms by Recursive Feature Elimination (RFE) [11] with 5-fold cross-validation on each algorithm separately. RFE can help determine the best performance of each model and the corresponding optimal feature set. RFE achieves this by searching for the most relevant subset of features to optimize our performance metric. We used cross-validation to optimize sensitivity by setting it as the score of RFE. To achieve the most optimal sensitivity, we started with all features from the training dataset as the input and fit the ML models, which ranked features by importance, discarded the least important features and refit the model. This process was repeated until the desired number of features resulted in the highest sensitivity (as shown in Fig. 2).



(a) Machine Learning Models for 0-48 Hour



(b) Machine Learning Models for over 48 Hour

Fig. 2. Recursive Feature Elimination (RFE) by using Machine Learning Models for Healthy vs CCHD at Different Ages.

1) 0-48 Hours of age to evaluate for no-CHD vs CCHD: In this setting, we found that Random Forest Classifier resulted in the highest sensitivity in the 5-fold cross-validation for the 0-48 hour age. Fig. 2(a) shows the optimal sensitivity was achieved by using 11 features.

2) Over 48 Hours of age to evaluate for no-CHD vs CCHD: We found that Logistic Regression performed best in this group, as Fig. 2(b) demonstrates, we found a subset of 7 optimal features using Logistic Regression.

The fewer number of CCHD cases compared to healthy cases resulted in a class imbalance problem, making it challenging to train a relatively unbiased ML model. To deal with this issue, we applied Random Oversampling (ROS), Random Undersampling (RUS), and Synthetic Minority Oversampling Technique (SMOTE) [12], separately. SMOTE is one type of oversampling method, where the minority class is oversampled by generating ”composite” examples [12]. We tried different sampling ratios from 0.2 to 0.9, but the results had trivial improvements. Thus, we used the balanced loss function inside of the ML models, which led to the optimal results we could achieve in this study.

E. Performance Evaluation

CCHD screening is a binary classification problem between healthy and CCHD, thus we used the following metrics

to comprehensively evaluate the performance of our model: Sensitivity (Sens) (1) and Specificity (Spec) (2).

$$\text{Sens} = \text{TPR} = \frac{TP}{TP + FN} \tag{1}$$

$$\text{Spec} = \frac{TN}{TN + FP} \tag{2}$$

$$\text{FPR} = \frac{FP}{FP + TN} \tag{3}$$

where:

- TP: the number of CCHD predicted as CCHD
- FP: the number of healthy predicted as CCHD
- TN: the number of healthy predicted as healthy
- FN: the number of CCHD predicted as healthy

We also calculated the Area Under the Receiver Operating Characteristics curve (AUROC) by plotting true positive rate (TPR) (1) against false positive rate (FPR) (3) with the discrimination threshold increasing from 0 to 1.

III. RESULTS

A. Sensitivity and Specificity

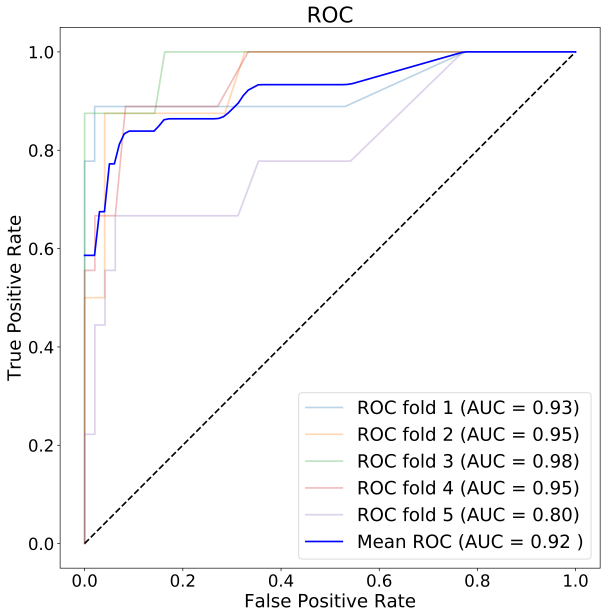
TABLE I
COMPARED WITH CURRENT SPO2 SCREENING

Methods	Specificity	Sensitivity
True Spot SpO2-alone [9]	96.8	62.8
Conservative Spot SpO2-alone [9]	96.8	76.5
ML (0-48 hrs)	**97.5**	**81.5**
ML (>48 hrs)	**100**	**83.3**
Conservative Spot SpO2-alone + ML (0-48 hrs)	96.2	**85.2**
Conservative Spot SpO2-alone + ML (>48 hrs)	96	**88.9**
Conservative Spot SpO2-alone + Any ML	95.8	**86.4**

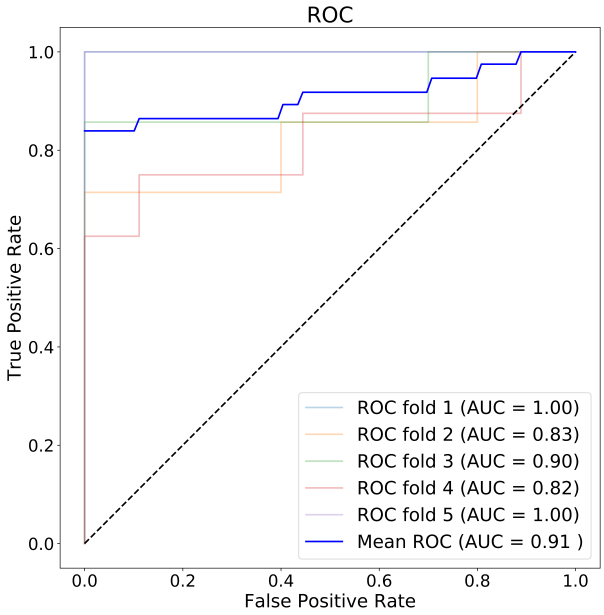
The sensitivity and specificity of our proposed method and the current SpO2 screening methods are summarized in Table. I. When comparing to the current standard CCHD screening, which includes spot right hand and any foot SpO2 measurements, the addition of the over 48 hour ML model did not lead to additional false positive results, but did detect 4 additional newborns with CCHD, hence increasing sensitivity from 76.5% to 88.9% (McNemar mid-p = 0.06) when tested on a sample of 50 healthy newborns and 36 newborns with CCHD. The addition of the 0-48 hour ML model, resulted in 3 additional false positive results and detected 3 additional newborns with CCHD, hence increasing sensitivity from 76.5% to 85.2% (McNemar mid-p = 0.13) when tested on a sample of 158 healthy newborns and 27 newborns with CCHD. Overall, the ML models improved the sensitivity from 76.5% to 86.4%, nearly 10 percentage point improvement.

B. ROC Curve

ROC curve results from 5-fold testing are shown in Fig. 3. For 0-48 hours, the average AUROC for no-CHD vs CCHD was 0.92 by using the Random Forest classifier; for over 48 hours, the average AUROC was 0.91 by using the Logistic Regression model. The estimated AUROC for our



(a) 0-48 Hours



(b) Over 48 Hours

Fig. 3. Area Under the Receiver Operating Curves (AUROC) for Models on No-CHD vs CCHD.

ML algorithms combining pulse oximetry features (PAI and HR) appears similar to or better than the current SpO2-alone screen [13], [14].

C. Interpretable Machine Learning

Compared to the current standard SpO2-alone CCHD screening, our ML models have the potential to achieve better sensitivity by incorporating features related to HR and PAI (or PIx). The optimal subset for the 0-48 hours Random Forest classifier includes: HR (median, mean, max, variance), SpO2 (min, max, median, mean), PAI or PIx (mean, median, max). The optimal subset for the over 48 hours Logistic Regression ML model includes: HR (min, max, variance),

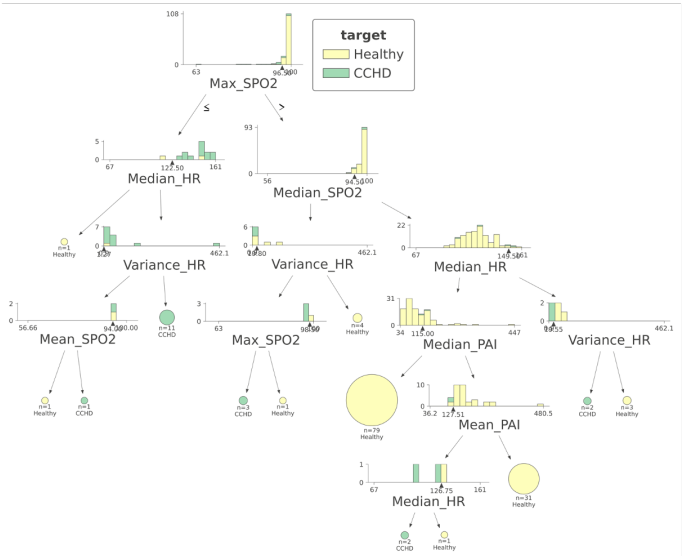


Fig. 4. Visualization of One Decision Tree.

SpO2 (median, mean), PAI or PIx (mean, min). Therefore, from our experiments, we can conclude the features extracted from HR and PAI have potentials for CCHD detection.

From Logistic Regression, we achieved importance ranking of features based on the trained weights. For the Random Forest classifier, we visualized decision trees as shown in (Fig. 4): the decision tree mainly relies on the features from SpO2, but the features related to HR and PAI are also considered in deeper decision-making layers of the tree.

IV. DISCUSSION

In this study, we investigated ML algorithms with optimized feature selection to provide an enhanced CCHD screening algorithm. We first applied the current standard CCHD screening (including both True Spot SpO2-alone and Conservative Spot SpO2-alone) to our enrolled newborns as the benchmark. Then, we tested our proposed ML models on these newborns, which gained approximately 10 percentage points increase in sensitivity of CCHD detection. Our proposed system improved detection of defects currently missed by SpO2-alone which would serve as a promising enhanced CCHD screening tool. Furthermore, by using these interpretable ML models, we found potential benefit of PAI and HR as features to differentiate between healthy newborns and newborns with CCHD.

In our future work, we will test the generalizability of our proposed system on more newborns, and we will study the potential relations between PAI and specifically systemic CCHD. To further improve the sensitivity, other potential features related to CCHD diagnosis such as radiofemoral pulse delay and photoplethysmography slopes will be incorporated and evaluated in the future.

ACKNOWLEDGMENT

The project described was supported by the National Center for Advancing Translational Sciences, National Institutes of Health (NIH), through grant number UL1 TR001860 and linked award KL2 TR001859, the Eunice Kennedy Shriver National Institute of Child Health & Human Development, NIH, through grant number 1R21HD099239-02, the University of California, Davis (UCD) Artificial Intelligence Seed Grant and UCD Venture Catalyst DIAL Grant. The content is solely the responsibility of the authors and does not necessarily represent the official views of the NIH or UCD.

CONFLICTS OF INTEREST

Chen-Nee Chuah, Satyan Lakshminrusimha, Zhengfeng Lai, and Heather Siefkes are named as inventors on a patent application "Systems and Methods for Classifying Critical Heart Defects." Heather Siefkes is the founding member of NeoPOSE Inc., a company working to develop devices to detect CCHD.

REFERENCES

[1] M. D. Reller, M. J. Strickland, T. Riehle-Colarusso, W. T. Mahle, and A. Correa, "Prevalence of congenital heart defects in metropolitan atlanta, 1998-2005," *J. Pediatr.*, vol. 153, no. 6, pp. 807–813, 2008.

[2] R.-K. R. Chang, M. Gurvitz, and S. Rodriguez, "Missed diagnosis of critical congenital heart disease," *Archives of pediatrics & adolescent medicine*, vol. 162, no. 10, pp. 969–974, 2008.

[3] R. I. Koppel, C. M. Druschel, T. Carter, B. E. Goldberg, P. N. Mehta, R. Talwar, and F. Z. Bierman, "Effectiveness of pulse oximetry screening for congenital heart disease in asymptomatic newborns," *Pediatrics*, vol. 111, no. 3, pp. 451–455, 2003.

[4] C. Wren, Z. Reinhardt, and K. Khawaja, "Twenty-year trends in diagnosis of life-threatening neonatal cardiovascular malformations," *Archives of Disease in Childhood-Fetal and Neonatal Edition*, vol. 93, no. 1, pp. F33–F35, 2008.

[5] K. Lannering, M. Bartos, and M. Mellander, "Late diagnosis of coarctation despite prenatal ultrasound and postnatal pulse oximetry," *Pediatrics*, vol. 136, no. 2, pp. e406–e412, 2015.

[6] E. C. Ailes, S. M. Gilboa, M. A. Honein, and M. E. Oster, "Estimated number of infants detected and missed by critical congenital heart defect screening," *Pediatrics*, vol. 135, no. 6, pp. 1000–1008, 2015.

[7] R. Abouk, S. D. Grosse, E. C. Ailes, and M. E. Oster, "Association of us state implementation of newborn screening policies for critical congenital heart disease with early infant cardiac deaths," *Jama*, vol. 318, no. 21, pp. 2111–2118, 2017.

[8] K. Doshi, G. B. Rehm, P. Vadlaputi, Z. Lai, S. Lakshminrusimha, C.-N. Chuah, and H. M. Siefkes, "A novel system to collect dual pulse oximetry data for critical congenital heart disease screening research," *J. Clin. Transl. Sci.*, pp. 1–7.

[9] G. R. Martin, A. K. Ewer, A. Gaviglio, L. A. Hom, A. Saarinen, M. Sontag, K. M. Burns, A. R. Kemper, and M. E. Oster, "Updated strategies for pulse oximetry screening for critical congenital heart disease," *Pediatrics*, vol. 146, no. 1, 2020.

[10] L. Wei, S. Ventura, M. Lowery, M. A. Ryan, S. Mathieson, G. B. Boylan, and C. Mooney, "Random forest-based algorithm for sleep spindle detection in infant eeg," in *2020 42nd Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC)*. IEEE, 2020, pp. 58–61.

[11] B. I. Ferleger, K. S. Sonnet, T. H. Morriss, A. L. Ko, H. J. Chizeck, and J. A. Herron, "A tablet-and mobile-based application for remote diagnosis and analysis of movement disorder symptoms," in *2020 42nd Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC)*. IEEE, 2020, pp. 5588–5591.

[12] N. V. Chawla, K. W. Bowyer, L. O. Hall, and W. P. Kegelmeyer, "Smote: synthetic minority over-sampling technique," *J. Artif. Intell. Res.*, vol. 16, pp. 321–357, 2002.

[13] M. N. Plana, J. Zamora, G. Suresh, L. Fernandez-Pineda, S. Thangaratinam, and A. K. Ewer, "Pulse oximetry screening for critical congenital heart defects," *Cochrane Database Syst. Rev.*, no. 3, 2018.

[14] S. B. Cantor and M. W. Kattan, "Determining the area under the roc curve for a binary diagnostic test," *Med. Decis. Mak.*, vol. 20, no. 4, pp. 468–470, 2000.