Role of artificial-intelligence-assisted automated cardiac biometrics in prenatal screening for coarctation of aorta

Taksøe-Vester, C. A.; Mikolaj, K.; Petersen, O. B.B.; Vejlstrup, N. G.; Christensen, A. N.; Feragen, A.; Nielsen, M.; Svendsen, M. B.S.; Tolsgaard, M. G.

Published in:
Ultrasound in Obstetrics and Gynecology

DOI:
10.1002/uog.27608

Publication date:
2024

Document version
Publisher's PDF, also known as Version of record

Document license:
CC BY-NC-ND

Citation for published version (APA):
Role of artificial-intelligence-assisted automated cardiac biometrics in prenatal screening for coarctation of aorta

C. A. TAKSØE-VESTER1,2,3, K. MIKOLAJ4, O. B. B. PETERSEN1,2, N. G. VEJLSTRUP5, A. N. CHRISTENSEN4, A. FERAGEN4, M. NIELSEN6, M. B. S. SVENDSEN3 and M. G. TOLSGAARD1,2,3

1 Department of Clinical Medicine, Faculty of Health and Medical Sciences, University of Copenhagen, Copenhagen, Denmark; 2 Center of Fetal Medicine, Department of Gynecology, Fertility and Obstetrics, Copenhagen University Hospital, Rigshospitalet, Copenhagen, Denmark; 3 Copenhagen Academy of Medical Education and Simulation (CAMES), Rigshospitalet, Copenhagen, Denmark; 4 Technical University of Denmark, Lyngby, Denmark; 5 Department of Cardiology, Copenhagen University Hospital, Rigshospitalet, Copenhagen, Denmark; 6 Department of Computer Science, University of Copenhagen, Copenhagen, Denmark

KEYWORDS: AI; artificial intelligence; coarctation of aorta; fetal echocardiography

CONTRIBUTION

What are the novel findings of this work?
We developed an artificial intelligence (AI) model capable of identifying standard cardiac planes and conducting automated cardiac biometric measurements. Our findings show that the use of automatic cardiac biometric measurements with AI during the 18–22-week scan has the potential to enhance the identification of fetuses that are at risk of developing coarctation of the aorta (CoA).

What are the clinical implications of this work?
AI technology provides a timesaving, objective and standardized method for conducting cardiac biometric measurements, which can eliminate interobserver variability and improve the accuracy of CoA detection compared with human measurements. Implementation of AI could improve outcomes for infants with CoA by enabling early intervention and treatment.

ABSTRACT

Objective Although remarkable strides have been made in fetal medicine and the prenatal diagnosis of congenital heart disease, around 60% of newborns with isolated coarctation of the aorta (CoA) are not identified prior to birth. The prenatal detection of CoA has been shown to have a notable impact on survival rates of affected infants. To this end, implementation of artificial intelligence (AI) in fetal ultrasound may represent a groundbreaking advance. We aimed to investigate whether the use of automated cardiac biometric measurements with AI during the 18–22-week anomaly scan would enhance the identification of fetuses that are at risk of developing CoA.

Methods We developed an AI model capable of identifying standard cardiac planes and conducting automated cardiac biometric measurements. Our data consisted of pregnancy ultrasound image and outcome data spanning from 2008 to 2018 and collected from four distinct regions in Denmark. Cases with a postnatal diagnosis of CoA were paired with healthy controls in a ratio of 1:100 and matched for gestational age within 2 days. Cardiac biometrics obtained from the four-chamber and three-vessel views were included in a logistic regression-based prediction model. To assess its predictive capabilities, we assessed sensitivity and specificity on receiver-operating-characteristics (ROC) curves.

Results At the 18–22-week scan, the right ventricle (RV) area and length, left ventricle (LV) diameter and the ratios of RV/LV areas and main pulmonary artery/ascending aorta diameters showed significant differences, with Z-scores above 0.7, when comparing subjects with a postnatal diagnosis of CoA (n = 73) and healthy controls (n = 7300). Using logistic regression and backward feature selection, our prediction model had an area under the ROC curve of 0.96 and a specificity of 88.9% at a sensitivity of 90.4%.

Conclusions The integration of AI technology with automated cardiac biometric measurements obtained
during the 18–22-week anomaly scan has the potential to enhance substantially the performance of screening for fetal CoA and subsequently the detection rate of CoA. Future research should clarify how AI technology can be used to aid in the screening and detection of congenital heart anomalies to improve neonatal outcomes. © 2024 The Authors. Ultrasound in Obstetrics & Gynecology published by John Wiley & Sons Ltd on behalf of International Society of Ultrasound in Obstetrics and Gynecology.

INTRODUCTION

Congenital heart disease (CHD) contributes to 30% of infant mortality resulting from congenital malformations. Coarctation of the aorta (CoA) accounts for 5–8% of CHD in children. Approximately 60% of isolated CoA cases go undetected prenatally, risking, without timely intervention, circulatory collapse and death when the arterial duct closes. Worldwide, women are encouraged to attend a scan at 18–22 weeks’ gestation performed by sonographers adhering to international guidelines. Risk stratification for CoA relies on subjective evaluation of the symmetry of the four-chamber view (4CV) and three-vessel view (3VV) without performing cardiac biometric measurements, which would be infeasible owing to time constraints in most settings. Fetal echocardiography takes a long time to master, consequently, the detection rate of CHDs largely depends on the clinician’s experience.

One of the latest advancements in ultrasound is the incorporation of artificial intelligence (AI). The current emphasis in the AI and CHD detection field is on automatic anomaly detection for certain conditions such as hypoplastic left heart syndrome, ventricular septal defect and tetralogy of Fallot. These conditions have more evident anatomical characteristics that differentiate them from healthy fetuses and, unlike CoA cases, this results in higher detection rates. Indeed, a gap exists in the literature concerning AI-assisted screening of less-apparent CHDs, such as CoA, that have lower detection rates. Several previous studies have evaluated manual methods for cardiac biometric measurements to predict CoA, however, they have mainly included third-trimester fetal echocardiography performed owing to a prior suspicion of ventricular disproportion. Furthermore, third-trimester fetal echocardiography, which is often performed at specialized centers, differs from the 18–22-week scan by including additional cardiac planes.

We developed an AI algorithm to recognize cardiac standard planes and perform automatic biometric measurements. The primary aim of this study was to use reliable quantitative fetal echocardiographic predictors of postnatal development of CoA to develop an AI-assisted screening tool for use during the 18–22-week scan. We hypothesized that performing automated biometric measurements during screening examinations would lead to more accurate identification of fetuses at risk of developing CoA.

METHODS

We developed an AI model trained to identify and evaluate the image quality of eight standard fetal cardiac planes, delineate the pertinent anatomy and automatically calculate cardiac biometric measurements. It is a deep learning model based on a convolutional neural network with a U-net architecture. All measurements for this study were performed by the cardiac AI model developed and prospectively validated by our research team. The supplementary material (Appendix S1, Figure S1, Tables S1 and S2) provides a comprehensive overview of the AI architecture, model performance scores per plane basis and evaluation of the measurements.

The study was conducted as a national retrospective observational study across multiple centers, involving pregnant women who participated in the Danish prenatal ultrasound screening program between 1st January 2008 and 31st December 2018. We collected ultrasound images and pregnancy and outcome data from women in four of five regions of Denmark: North, South, Zealand and the Capital Region. The screening program includes two ultrasound examinations, one at 12 weeks’ gestation and one at 18–22 weeks. The pregnancy and outcome data were obtained from the Danish Fetal Medicine Database and the image data were collected from regional servers. Fetal ultrasound examinations were performed using GE Logiq 7 (GE Healthcare, Milwaukee, WI, USA) or Voluson E6, E8 or E10 machines (GE Healthcare, Zipf, Austria).

The Danish Health Authorities provided permits for the extraction of ultrasound and outcome data on 602,218 pregnancies for this project. This study was approved by The Danish Data Protection Agency (protocol no. P-2019-310) and The Danish Patient Safety Authority (protocol no. 3-3031-2915/1). The study was reported according to the TRIPOD guidelines.

We included all cases that received a postnatal diagnosis of CoA during the study period and matched them with healthy controls at a 1:100 ratio. Our primary objective was to identify during the second-trimester anomaly scan fetuses at risk of developing CoA postnatally, so we only used cardiac images from the 18–22-week examination. If a CoA case had missing 4CV or 3VV images, video sweeps from 18–22 weeks’ gestation were reviewed and, if available, the standard planes were retrieved from the videos. We excluded CoA cases with only a prenatal diagnosis, those with a diagnosis of hypoplastic left heart syndrome or those with missing ultrasound images of the 4CV or 3VV between 18 and 22 weeks. The healthy cohort comprised singleton pregnancies with no fetal malformations, a birth weight between the 10th and 90th percentiles at term, no pre-eclampsia and spontaneous conception. We matched the healthy cases based on gestational age (GA) within ±2 days, since cardiac biometrics are closely related to GA. Additionally, we...
Taksøe-Vester et al. included only control cases with high-quality ultrasound images, specifically choosing the top 100 best images and filtering down, as the precision of measurements from the AI model is highly dependent on image quality. The AI model automatically assessed the quality of the images.

We conducted an evaluation of our AI model’s segmentations of cardiac anatomy on all CoA cases and on control cases randomly selected in a ratio of 1:5. Any incorrect segmentations were manually corrected by one annotator (C.A.T.), and cases of doubt were discussed with a fetal medicine expert (M.G.T.). Since the inclusion criteria ensured high image quality for the control cases, there was no need to make any corrections to the segmentations on control images during the inspection. This semi-automatic quality assurance approach was adopted to control for the impact of image quality on segmentation accuracy. The quality of the CoA images during the 11-year study period varied considerably and could potentially impact on the accuracy of segmentations if not semi-automatically corrected. Therefore, we needed to ensure that we could trust the measurement outputs clinically. All cardiac measurements for both CoA and controls were performed by the AI model based on the anatomical segmentations, and the AI was blinded to patient outcomes.

Of the standard planes obtained at the 18–22-week scan, the cardiac planes selected for measurements were the 4CV and the 3VV, based on prior research indicating their relevance for the detection of CoA. A significant majority of three-vessel-and-trachea (3VT) view images, which is another important plane for CoA detection, were saved with color Doppler flow in our dataset. This led to their exclusion from the analysis because the model was not trained on flow images and simply removing the flow before performing measurements would lead to an overestimation of the cardiac biometrics by the AI in these cases.

Cardiac measurements obtained from the 4CV were atrioventricular valve diameter and right and left atrial and ventricular dimensions (area, length and diameter) measured in end-diastole. Measurements from the 3VV were diameters of the descending aorta (Dao), ascending aorta (Aao) and main pulmonary artery (MPA). Additionally, we evaluated the ratio between the areas of the ventricles and the diameters of the MPA and Aao. To ensure accurate measurements, ventricular diameters were measured from the endocardium of the ventricular wall to the endocardium of the ventricular septum at the maximum transverse diameter, as suggested by previous studies. Segmentation examples of the two standard planes are provided in Figure 1 for a CoA case and a control.

Statistical analysis

The distribution of cardiac measurements used in this study is expressed in terms of mean and SD. Welch’s t-test was used to account for unequal variances for the statistical comparison between CoA and control measurements. Z-scores were calculated using pooled SD of the two groups and their respective means. Logistic regression models were fitted with postnatal CoA development as the dependent variable and echocardiographic biometrics as independent variables. The selection of logistic regression for this study was driven by its simplicity and straightforward interpretability. In this study, we considered two models. One utilized all available measurements/features, and the other model used a backward feature selection procedure. This procedure starts by including all features and iteratively removes one at a time, aiming to maximize the model’s

Figure 1 Four-chamber view (a–d) and three-vessel view (e–h) ultrasound images: examples of artificial intelligence segmentation (b,d,f,h) on which automated measurements are based, in a healthy control fetus (a,b,e,f) and a fetus with coarctation of the aorta (c,d,g,h).
performance for a given number of features. One of the primary motivations for using backward feature selection was to reduce the risk of overfitting and thereby improve the generalizability of the model.

After conducting feature selection, a subsequent assessment for multicollinearity was performed to identify any potential redundant measurements. The models’ performance is presented as receiver-operating-characteristics (ROC) curves. Moreover, owing to the limited number of CoA cases, we utilized 5-fold cross-validation during training and testing to ensure that all cases were included in estimating the model performance.

Our consideration was to arrive ultimately at the most effective model; however, recognizing the gradual integration of AI into all equipment, we also considered the merits of simplicity. Hence, we opted to present the three most informative measures that can currently be ascertained by manual measurement, along with the composite model after feature selection. The results are presented using thresholds that achieve 90% sensitivity, which represents the point at which the test operates efficiently, considering its role as a screening test.

Statistical analysis was performed using Python 3.9.12.

RESULTS

Ninety-nine fetuses with a postnatal diagnosis of CoA born between 2008 and 2018 were identified. Of these, 26 were excluded from further analysis: one case because of a diagnosis of hypoplastic left heart syndrome, three cases owing to inadequate fetal imaging, four cases had only a 3VV image available and 18 images had only a 4CV image available in our dataset. Therefore, 73 CoA cases with both 3VV and 4CV images available were included into the prediction model and matched to 7300 healthy controls. Nine images were retrieved from video sweeps. The mean GA at the time of fetal echocardiography was 140.2 ± 4.7 days for CoA cases and 140.5 ± 4.7 days for healthy matched controls, with mean estimated fetal weights of 312.8 ± 50.5 g and 326.9 ± 38.2 g, respectively. The mean year of scan for the CoA fetuses was 2014 ± 2 and for controls it was 2016 ± 2. Table 1 provides background characteristics of the cases and controls.

During the semi-automatic evaluation, 24.7% of 4CV images and 52.1% of 3VV images from the CoA cases had segmentations corrected, whereas none of the healthy control images were corrected during the inspection.

Fetuses that had been diagnosed with CoA postnatally displayed significant deviations from healthy controls in terms of their cardiac structure. Specifically, these fetuses displayed significant deviations from healthy controls in terms of their cardiac structure. Specifically, these fetuses displayed significant deviations from healthy controls in terms of their cardiac structure. Specifically, these fetuses displayed significant deviations from healthy controls in terms of their cardiac structure. Specifically, these fetuses displayed significant deviations from healthy controls in terms of their cardiac structure. Specifically, these fetuses displayed significant deviations from healthy controls in terms of their cardiac structure.

As shown in Table 1, there was a significant difference in estimated fetal weight between CoA cases and controls, therefore, to adequately account for this potential confounding factor it was incorporated as a covariate in the logistic regression model. Nevertheless, this variable was omitted during the prediction phase to maintain a more clinically applicable model that required fewer measures. Missing data were imputed with mean values from the dataset for the logistic regression model.

<table>
<thead>
<tr>
<th>Characteristic</th>
<th>CoA cases (n = 73)</th>
<th>Controls (n = 7300)</th>
<th>p*</th>
</tr>
</thead>
<tbody>
<tr>
<td>Maternal age (years)</td>
<td>30.2 ± 5.1</td>
<td>30.5 ± 4.7</td>
<td>0.62</td>
</tr>
<tr>
<td>Body mass index (kg/m²)</td>
<td>24.6 ± 5.3</td>
<td>21.7 ± 3.4</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>Parity</td>
<td>0.8 ± 0.9</td>
<td>0.8 ± 0.8</td>
<td>1</td>
</tr>
<tr>
<td>GA at scan (days)</td>
<td>140.2 ± 4.7</td>
<td>140.5 ± 4.7</td>
<td>0.59</td>
</tr>
<tr>
<td>EFW (g/1)</td>
<td>312.8 ± 50.5</td>
<td>326.9 ± 38.2</td>
<td>0.0027</td>
</tr>
<tr>
<td>Year of scan</td>
<td>2014 ± 2</td>
<td>2016 ± 2</td>
<td>—</td>
</tr>
<tr>
<td>GA at birth (days)</td>
<td>271.9 ± 18.7</td>
<td>279.8 ± 10.8</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>Birth weight (g)</td>
<td>3217.7 ± 871.4</td>
<td>3527.0 ± 413.3</td>
<td>0.003</td>
</tr>
<tr>
<td>Male sex</td>
<td>41 (56.2)</td>
<td>3715 (50.9)</td>
<td>—</td>
</tr>
<tr>
<td>Spontaneous conception</td>
<td>64 (87.7)</td>
<td>7300 (100)</td>
<td>—</td>
</tr>
</tbody>
</table>

Data are given as mean ± SD or n (%). *Differences calculated using Welch’s t-test. Some cases lacked necessary measurements for inclusion in the four-parameter Hadlock formula (missing CoA cases, n = 6; missing controls, n = 362). EFW, estimated fetal weight; GA, gestational age.
Table 2 Cardiac parameters obtained in fetuses with coarctation of the aorta (CoA) and matched healthy controls

<table>
<thead>
<tr>
<th>Parameter</th>
<th>CoA (n = 73)</th>
<th>Controls (n = 7300)</th>
<th>P*</th>
<th>Z-score†</th>
<th>All</th>
<th>Feature selection</th>
</tr>
</thead>
<tbody>
<tr>
<td>Four-chamber view</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>RV diameter (cm)</td>
<td>0.61 ± 0.10</td>
<td>0.56 ± 0.08</td>
<td>&lt; 0.0001</td>
<td>0.63</td>
<td>−0.31</td>
<td></td>
</tr>
<tr>
<td>RV length (cm)</td>
<td>1.15 ± 0.18</td>
<td>1.03 ± 0.16</td>
<td>&lt; 0.0001</td>
<td>0.75</td>
<td>0.825</td>
<td></td>
</tr>
<tr>
<td>RV area (cm²)</td>
<td>0.45 ± 0.13</td>
<td>0.38 ± 0.09</td>
<td>&lt; 0.0001</td>
<td>0.78</td>
<td>1.781</td>
<td></td>
</tr>
<tr>
<td>LV diameter (cm)</td>
<td>0.54 ± 0.11</td>
<td>0.61 ± 0.10</td>
<td>&lt; 0.0001</td>
<td>−0.7</td>
<td>−0.014</td>
<td></td>
</tr>
<tr>
<td>LV length (cm)</td>
<td>1.06 ± 0.16</td>
<td>1.10 ± 0.18</td>
<td>0.0249</td>
<td>−0.22</td>
<td>0.046</td>
<td></td>
</tr>
<tr>
<td>LV area (cm²)</td>
<td>0.38 ± 0.12</td>
<td>0.46 ± 0.13</td>
<td>&lt; 0.0001</td>
<td>−0.62</td>
<td>−2.749</td>
<td>−0.970</td>
</tr>
<tr>
<td>RV/LV area ratio</td>
<td>1.22 ± 0.26</td>
<td>0.87 ± 0.44</td>
<td>&lt; 0.0001</td>
<td>0.80</td>
<td>−1.017</td>
<td></td>
</tr>
<tr>
<td>RA diameter (cm)</td>
<td>0.58 ± 0.12</td>
<td>0.59 ± 0.10</td>
<td>0.3232</td>
<td>−0.1</td>
<td>0.701</td>
<td></td>
</tr>
<tr>
<td>RA length (cm)</td>
<td>0.76 ± 0.13</td>
<td>0.79 ± 0.13</td>
<td>0.1043</td>
<td>−0.23</td>
<td>0.136</td>
<td></td>
</tr>
<tr>
<td>RA area (cm²)</td>
<td>0.32 ± 0.11</td>
<td>0.34 ± 0.10</td>
<td>0.1366</td>
<td>−0.2</td>
<td>−0.726</td>
<td></td>
</tr>
<tr>
<td>LA diameter (cm)</td>
<td>0.48 ± 0.11</td>
<td>0.54 ± 0.11</td>
<td>&lt; 0.0001</td>
<td>−0.55</td>
<td>0.413</td>
<td></td>
</tr>
<tr>
<td>LA length (cm)</td>
<td>0.66 ± 0.12</td>
<td>0.72 ± 0.14</td>
<td>&lt; 0.0001</td>
<td>−0.43</td>
<td>0.045</td>
<td></td>
</tr>
<tr>
<td>LA area (cm²)</td>
<td>0.22 ± 0.08</td>
<td>0.27 ± 0.09</td>
<td>&lt; 0.0001</td>
<td>−0.56</td>
<td>−0.877</td>
<td>−0.388</td>
</tr>
<tr>
<td>Mitral valve diameter (cm)</td>
<td>0.46 ± 0.13</td>
<td>0.51 ± 0.14</td>
<td>0.0007</td>
<td>−0.36</td>
<td>0.073</td>
<td></td>
</tr>
<tr>
<td>Tricuspid valve diameter (cm)</td>
<td>0.59 ± 0.13</td>
<td>0.60 ± 0.13</td>
<td>0.4552</td>
<td>−0.08</td>
<td>−0.240</td>
<td></td>
</tr>
<tr>
<td>Three-vessel view</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>MPA diameter (cm)</td>
<td>0.39 ± 0.07</td>
<td>0.36 ± 0.07</td>
<td>0.0068</td>
<td>0.43</td>
<td>0.922</td>
<td>0.627</td>
</tr>
<tr>
<td>Aao diameter (cm)</td>
<td>0.23 ± 0.06</td>
<td>0.33 ± 0.06</td>
<td>&lt; 0.0001</td>
<td>−1.67</td>
<td>−1.841</td>
<td>−1.411</td>
</tr>
<tr>
<td>Dao diameter (cm)</td>
<td>0.19 ± 0.06</td>
<td>0.22 ± 0.05</td>
<td>&lt; 0.0001</td>
<td>−0.75</td>
<td>−0.317</td>
<td>−0.347</td>
</tr>
<tr>
<td>MPA/Aao diameter ratio</td>
<td>1.67 ± 0.48</td>
<td>1.09 ± 0.20</td>
<td>&lt; 0.0001</td>
<td>2.9</td>
<td>0.043</td>
<td></td>
</tr>
<tr>
<td>Intercept</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>−8.031</td>
<td>−6.657</td>
</tr>
</tbody>
</table>

Data given as mean ± SD, unless stated otherwise. *Welch’s t-test between mean of CoA and control groups. †Z-score is number of SDs by which value of CoA mean is above or below mean of controls. ‡Coefficients from logistic regression analyses. Aao, ascending aorta; Dao, descending aorta; LA, left atrium; LV, left ventricle; MPA, main pulmonary artery; RA, right atrium; RV, right ventricle.

The MPA/Aao ratio emerged as the most crucial individual feature, based on the AUC (0.90). Setting the threshold at 1.15 in MPA/Aao ratio resulted in a sensitivity of 90.3% (95% CI, 83.4–97.1%) and a specificity of 61.9% (95% CI, 60.8–63.1%) for identifying fetuses at risk of developing CoA postnatally. Figure 4 illustrates the relationship between the sensitivity and specificity as a function of the threshold.

**DISCUSSION**

We have developed a predictive AI-assisted screening model aimed at identifying fetuses at risk of developing CoA postnatally using automatic biometric measurements from the 4CV and 3VV during the 18–22-week scan. Our study showed that CoA fetuses display notable deviations in several parameters during this scan, which aligns with previous research focused on later gestations. In particular, CoA fetuses exhibited significantly larger RV dimensions, smaller Aao and Dao diameters and significantly larger ratios of RV/LV and MPA/Aao when compared with healthy controls.

Prior research has shown results consistent with our findings, in which ROC curves were displayed with AUCs ranging from 0.82 to 0.98, the best of which involved technically challenging ultrasound planes, including the sagittal plane of the aortic arch, and were conducted by experts. Previous studies have centered on constructing diagnostic prediction models based on CoA cases already suspected of having ventricular imbalance and, consequently, undergoing fetal echocardiography.
Table 3 Performance of cardiac biometrics in predicting coarctation of the aorta, at a sensitivity of 90.3% or 90.4%

<table>
<thead>
<tr>
<th>Predictive parameter(s)</th>
<th>AUC</th>
<th>Sensitivity (%)</th>
<th>Specificity (%)</th>
<th>PPV (%)</th>
<th>NPV (%)</th>
<th>LR+</th>
<th>LR–</th>
<th>Threshold</th>
</tr>
</thead>
<tbody>
<tr>
<td>All measures in Table 2</td>
<td>0.9565</td>
<td>90.3</td>
<td>84.8</td>
<td>0.24</td>
<td>99.9</td>
<td>5.94</td>
<td>0.11</td>
<td>0.0051*</td>
</tr>
<tr>
<td>RVa, LVa, Daoa, Aaoa, MPAd, LAa</td>
<td>0.9625</td>
<td>90.4</td>
<td>88.9</td>
<td>0.33</td>
<td>99.9</td>
<td>8.17</td>
<td>0.11</td>
<td>0.0095*</td>
</tr>
<tr>
<td>MPAd/Aaoa ratio</td>
<td>0.8978</td>
<td>90.3</td>
<td>61.9</td>
<td>0.09</td>
<td>99.9</td>
<td>2.37</td>
<td>0.15</td>
<td>1.1500†</td>
</tr>
<tr>
<td>Aaoa</td>
<td>0.8864</td>
<td>90.4</td>
<td>59.5</td>
<td>0.09</td>
<td>99.9</td>
<td>2.23</td>
<td>0.16</td>
<td>0.3240†</td>
</tr>
<tr>
<td>RVa/LVa ratio</td>
<td>0.8784</td>
<td>90.4</td>
<td>67.1</td>
<td>0.11</td>
<td>99.9</td>
<td>2.75</td>
<td>0.14</td>
<td>0.8900†</td>
</tr>
</tbody>
</table>

Values in parentheses are 95% CI. *Threshold is applied to output of logistic regression. †Threshold is applied directly to measurement.

in the third trimester\textsuperscript{23,12,14–18,23,26}. In contrast, the primary objective of this study was to refine the CoA screening procedure to ensure that at-risk fetuses do indeed undergo the supplementary examinations.

Automating the cardiac measurement process through AI during sonographer screening enables an objective approach, making it possible to perform in busy clinical settings without fetal-cardiology expertise. The time-intensive process of taking measurements has previously posed a challenge to the incorporation of new measurements for the diagnosis of CoA\textsuperscript{23}. This study involving AI measurements on screening images has yielded results that are 20–40% better than current detection rates\textsuperscript{5,3} and comparable with existing expert-based predictions in a prescreened group\textsuperscript{16}. The comparable results can be attributed to the accurate and consistent measurements carried out by AI, the more extensive dataset compared with that of previous studies\textsuperscript{16} and the effective feature-selection process, resulting in a strong set of features derived from the extensive pool available. Furthermore, prior research suggests that models combining multiple features perform better than those focusing on single features\textsuperscript{13,27}, which supports the competitive performance of our model with six features. This promising alignment indicates that AI measurements can achieve the same level of detection on images acquired by sonographers (ultrasound technicians) during routine screening, rather than in specialized echocardiography settings. This suggests a substantial potential for AI to contribute to accurate detection in both referral and tertiary hospitals. Additionally, by flagging imbalanced ventricles for specialist examination, it could contribute to improving the detection of other cardiac lesions with similar findings.

Previous research has demonstrated that measuring the isthmus in the 3VT view is a strong predictor of CoA development\textsuperscript{18} and provided cut-off values. Additionally, other research has demonstrated a 100% specificity when examining the aortic arch in the sagittal plane and measuring the angle between the Aao and Dao\textsuperscript{5}. For a prudent

© 2024 The Authors. Ultrasound in Obstetrics & Gynecology published by John Wiley & Sons Ltd on behalf of International Society of Ultrasound in Obstetrics and Gynecology.

Ultrasound Obstet Gynecol 2024; 64: 36–43.
approach to also reduce the false-positive rate, sonographers can acquire the 3VT plane and subsequently the aortic arch in the sagittal plane if the prediction model, based on the 4CV and 3VV, flags an increased risk for CoA. This preliminary step may eliminate the need for comprehensive echocardiography performed by a fetal medicine expert, improving efficiency and cost-effectiveness. However, this approach is feasible only in settings in which local fetal medicine experts are available to assess the additional images. In order to evaluate the system’s effectiveness, prospective testing is essential to determine the potential increase in unnecessary scans, particularly in settings in which this additional step is not feasible.

Limitations of our study include the lack of follow-up in medical records to differentiate severe and mild CoA cases, as well as the exclusion of the 3VT view. The exclusion hampers the direct applicability to current screening programs, of which the 3VT view is an integral component. Another limitation is that the images in our sample spanned over a decade, with images being up to 14 years old, which necessitated human correction of segmentations in low-quality images. The issue was specific to the CoA images, with the healthy control images not having this problem since they were more recent and partly chosen based on image quality. Moreover, a higher body mass index (BMI) was observed in mothers of the CoA cases compared with the control group, which partly explains the lower image quality in the CoA group. The control group’s low BMI could potentially constrain the external applicability of this model to different populations. Nonetheless, these potential confounding factors are diminished by the fact that the prediction model relies on measurements rather than AI-based textural analysis. A semi-automatic approach, in which the sonographer accepts or corrects the AI model’s segmentations before relying on the measurements and conclusions, would overcome the issue of image quality impairing the accuracy of the AI system and ensure autonomy for the clinician.

Trained on screening images obtained by sonographers from four distinct regions in Denmark, our model is expected to have a high level of generalizability. Additionally, previous research indicates that AI models based on the same dataset have shown effective generalization to other European populations. The utilization of an AI-assisted screening approach gives rise to several ethical challenges. These challenges involve questions about responsibility when AI systems make mistakes, leading to unnecessary anxiety and distress among affected families. Engaging clinicians in the process, as suggested in this study, can provide an essential human touch. Our proposed measurements not only offer explanations to clinicians, but also streamline workflow and enable cardiac assessment during routine screening within busy clinical settings.

Determining a threshold between sensitivity and specificity and allowing for additional scans as a means of improving detection is a nuanced blend of political and health-economic considerations. While our study contributes valuable insights and potential strategies, we acknowledge that assessment of health resources is beyond the scope of our research. Nonetheless, identifying at-risk fetuses is important for their management irrespective of organization, reimbursement strategy or health politics.

In conclusion, this study pioneers a predictive screening model for early CoA suspicion during the 18–22-week scan, targeting reduced postnatal cardiovascular risk. Our approach utilizes AI’s potential to improve CoA detection rates and addresses one of the most elusive CHD diagnoses in fetal medicine.

ACKNOWLEDGMENTS

The Project is supported by The Novo Nordisk Foundation through grant NNF170030576, the Danish Regions’ AI Signature Project, through the Centre for Basic Machine Learning Research in Life Science (NNF20OC006206) and the Pioneer Centre for AI, DNRF grant no. P1.

REFERENCES


© 2024 The Authors. Ultrasound in Obstetrics & Gynecology published by John Wiley & Sons Ltd on behalf of International Society of Ultrasound in Obstetrics and Gynecology.
AI-assisted prenatal screening for CoA


SUPPORTING INFORMATION ON THE INTERNET

The following supporting information may be found in the online version of this article:

Appendix S1 Comprehensive overview of the artificial intelligence (AI) architecture, the model performance scores per plane basis and evaluation of the measurements

Figure S1 Model architecture.

Table S1 Model performance score per plane basis

Table S2 Measurement errors of the model