Taksøe-Vester Caroline (Orcid ID: 0000-0002-5634-2830)
Tolsgaard Martin Grønnebæk (Orcid ID: 0000-0001-9197-5564)

Role of AI-assisted automated cardiac biometrics in screening for fetal coarctation of aorta

Authors:

C. A. Taksoee-Vester[1,2,3], K. Mikolaj[4], O. B. B. Petersen[1,2], N. G. Vejlstrup[5], A. N. Christensen[4], A. Feragen[4], M. Nielsen[6], M. B. S. Svendsen[3], M. G. Tolsgaard[1,2,3]

[1]University of Copenhagen, Dept. of Clinical Medicine, Faculty of Health and Medical Sciences, Denmark

[2]Center of Fetal Medicine, Dept. of Gynecology, Fertility and Obstetrics, Copenhagen University Hospital, Rigshospitalet, Denmark

[3]Copenhagen Academy of Medical Education and Simulation (CAMES), Rigshospitalet, Denmark

[4]Technical University of Denmark, Lyngby, Denmark

[5]Dept. of Cardiology, Copenhagen University Hospital, Rigshospitalet, Denmark

[6]University of Copenhagen, Dept. of Computer Science, Denmark

Corresponding author: Dr C. A. Taksoee-Vester, Copenhagen University Hospital, Rigshospitalet, Blegdamsvej 9, Dept. 4071, DK – 2100 Copenhagen

(e-mail: ca_tv@hotmail.com)

Keywords: Coarctation aorta, fetal echocardiography, artificial intelligence

Running head: AI-assisted Screening for Fetal Coarctation

This article has been accepted for publication and undergone full peer review but has not been through the copyediting, typesetting, pagination and proofreading process which may lead to differences between this version and the Version of Record. Please cite this article as doi: 10.1002/uog.27608

CONTRIBUTIONS

What are the novel findings of this work?

We developed an AI model capable of identifying standard cardiac planes and conducting automated cardiac biometric measurements.

Our findings show that leveraging 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 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 inter-observer variability and improve the accuracy of CoA detection compared to human measures.

Implementation of AI could improve outcomes for infants with CoA by enabling early intervention and treatment.

Taksøe-Vester Caroline (Orcid ID: 0000-0002-5634-2830)
Tolsgaard Martin Grønnebæk (Orcid ID: 0000-0001-9197-5564)

ABSTRACT

Objectives:

Although there have been remarkable strides in fetal medicine and prenatal diagnosis of congenital heart disease, a significant percentage of newborns with isolated coarctation of the aorta (CoA) - around 60 percent - are still not identified prior to birth. The prenatal detection of CoA has been shown to have a notable impact on the survival rates of affected infants. To this end, the implementation of artificial intelligence (AI) in fetal ultrasound may represent a groundbreaking advancement. Our hypothesis is that leveraging automated cardiac biometric measurements with AI during the 18-22-week anomaly scan will enhance the identification of fetuses that are at risk of developing CoA.

Methods:

We have 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. The CoA cases from the period were paired with healthy controls in a ratio of 1:100 and matched on gestational ages of ±2 days. The cardiac biometrics on the four-chamber view and three vessel view were included in a logistic regression-based prediction model. To assess the predictive capabilities, we visualized sensitivity and specificity on Receiver Operating Characteristic (ROC) curves.

Results:

At the 18-22 week scan, the right ventricle (RV)area and length, left ventricle (LV) width, 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 produced a ROC curve with an AUC (Area Under the Curve) of 0.96 and a specificity of 88.9% at a sensitivity level of 90.4%.

Conclusion:

The integration of AI technology with automated cardiac biometric measurements conducted during the 18-22-week anomaly scan in fetal medicine has the potential to substantially enhance the screening for fetal CoA and subsequently the rate of CoA detection. Future research should clarify how AI technology can be used to aid in screening and detection of congenital heart anomalies to improve neonatal outcomes.

INTRODUCTION

Congenital heart disease (CHD) contributes to 30% of infant mortality resulting from congenital malformations[1]. Coarctation of the aorta (CoA), makes up 5-8% of CHD in children. Approximately 60% of isolated CoA cases go undetected prenatally[2,3], risking circulatory collapse and death when the arterial duct closes without timely intervention.

Worldwide, women are encouraged to attend the 18-22-week-scan performed by sonographers adhering to international guidelines. Risk stratification for CoA relies on subjective evaluation of the symmetry of the four-chamber and three-vessel views without performing any cardiac biometric measurement, which would be infeasible due to time constraints in most settings. Fetal echocardiography takes a long time to master[4], consequently, the detection rate of CHDs largely depends on the clinician's experience level[5,6].

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 septum defect and tetralogy of Fallot[7–10]. These conditions have more evident anatomical characteristics that differentiate them from healthy fetuses that, unlike CoA cases, result in higher detection rates[11]. Indeed, a gap exists in the literature concerning AI-assisted screening of less apparent CHDs, like CoA, which have lower detection rates. Several previous studies have evaluated manual methods for cardiac biometric measurements to predict CoA, however, they have mainly included 3rd trimester fetal echocardiographies performed due to a prior suspicion of ventricular disproportions[3,12–18]. Furthermore, 3rd trimester fetal echocardiographies differ from the 18-22-week-scan by including additional cardiac planes and are often performed at specialized centers.

We have 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 for postnatal development of CoA to develop an AI screening tool at the 18-22-week-scan.

We hypothesize that performing automated biometric measurements during screening examinations will 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[19]. 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, the model performance scores per plane basis and the 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 the 1st of January, 2008 and 31st of December, 2018. We collected ultrasound images, 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 at 12 and 18-22-weeks of gestation. The pregnancy and outcome data were obtained from the Danish Fetal Medicine Database (DFMD)[20] and the image data was collected from regional servers. Fetal ultrasound examinations were conducted using General Electrics logiq 7, E6, E8 or E10 machines.

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 guideline [21].

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 fetuses at risk of developing CoA postnatally during the second-trimester anomaly scan, so we only used cardiac images from the 18-22 -week examination. If a CoA case had missing four-chamber view (4CV) or three-vessel view (3VV) images, video sweeps were reviewed from GA 18-22 weeks 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 18-22 weeks of gestation. The healthy cohort comprised singleton pregnancies with no fetal malformations, a birthweight between the 10th and 90th percentile 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[22]. Additionally, we only included 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 in a ratio 1:5 randomly selected control cases. Any incorrect segmentations were manually corrected by one annotator (CAT), and cases of doubt were discussed with a fetal medicine expert (MGT). 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 impact of image quality on the segmentation accuracy. The quality of the CoA images during the ten-year study period varied considerably and could potentially impact 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 outcome.

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 detection of CoA[23]. A significant majority of three vessel 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 included were for the 4CV; atrioventricular-valves, right and left atrial and ventricular dimensions (area, length and width) measured end-diastolic, and for the 3VV; diameters of the descending aorta, the ascending aorta and the main pulmonary artery. Additionally, we evaluated the ratio between the areas of the ventricles and the diameters of the main pulmonary artery and ascending aorta. To ensure accurate measurements, ventricular widths 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[22,24], besides the width of the atrioventricular valves. Figure 1 provides segmentation examples of the two standard planes for one CoA case and one control.

Statistics

The distribution of cardiac measurements used in this study is expressed in terms of mean value and standard deviation (SD). Welch's t-test was used to account for unequal variances for the statistical comparison between CoA and control measurements. Moreover, the Z-scores were calculated using pooled SD of the two groups and their respective means.

Logistic regression models were fitted with the 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 utilizes all available measurements/features, and one additional model using 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 performance for a given number of features[25]. One of the primary motivations for using backward feature selection was to reduce the risk of overfitting, and to improve generalization 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 ROC

curves. Moreover, due to the limited number of CoA cases, we utilized 5-fold cross-validation during training and testing to ensure all cases were included in estimating the model performance.

Our consideration was to ultimately arrive at the most effective model; however, recognizing the gradual integration of AI into all equipment, we also value the merits of simplicity. Hence, we opted to present the three most informative measures that can presently be employed through manual measurement, along with the composite model post 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.

RESULTS

Ninety-nine fetuses with postnatal diagnosis of CoA born between 2008 and 2018 were identified. Of these 26 were excluded from further analysis; one due to a diagnosis of hypoplastic left heart syndrome, three cases were excluded due to inadequate fetal imaging, four cases had only a 3VV image available and 18 images had only a 4CV image available in our dataset. Nine images were retrieved from video sweeps. See Figure 2 for a flow of participants through the study. The mean GA at the time of fetal echocardiography was 140.2 (±4.7) days for CoA cases and 140.5 (±4.7) for healthy matched controls with mean estimated fetal weights at 312.8 (±50.5) and 326.9 (±38.2) respectively. The mean year of scan for CoA fetuses were 2014 (±2) and for controls 2016 (±2). Table 1 lists the background characteristics of the cases and controls.

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

Fetuses that have been diagnosed with CoA postnatally displayed significant deviations from healthy controls in terms of their cardiac structures. Specifically, these fetuses had smaller left cardiac structures such as the mitral valve width, atrium and ventricle dimensions, while also exhibiting larger right cardiac structures in terms of right ventricle dimensions (Table 2). Moreover, the CoA fetuses showed a significantly larger diameter of the MPA, a smaller diameter of the Aao, and larger ratios of RV/LV area and MPA/Aao diameter when compared to controls (Table 2). The cardiac structures that exhibited the most significant differences with z-scores above 0.7 were the RV area and length, LV, Aao and MPA diameter, and the ratios of RV/LV areas and MPA/AAo diameters.

In the logistic regression model, all parameters (listed in Table 2) from the group-wise comparison were considered, leading to the creation of a ROC curve with an AUC of 0.96 (as depicted in Figure 3). Setting the sensitivity to 90.3% (95%CI; 83.4 to 97.1) yielded a specificity of 84.8% (84.0-85.6), positive predictive value (PPV) 0.24% (95%CI; 0.1 to 0.4), negative predictive value (NPV) 99.9% (95%CI; 99.9-

100), positive likelihood ratio (LR+) 5.94 (5.4-6.5) and negative likelihood ratio (LR-) 0.11 (0.06-0.23).

After performing feature selection, the best performing model comprised seven features. The multicollinearity check revealed two highly related features, the RV length and area, with high variance inflation factors at 4.18 and 4.09, respectively. Consequently, one of the redundant features was removed, selected based on a lower magnitude of the coefficient. Therefore, the final model consisted of six features and resulted in a ROC curve with an AUC of 0.96. Setting the sensitivity to 90.4 (95%CI; 83.7 to 97.2) resulted in specificity 88.9 (95%CI; 88.2-89.6), PPV 0.33% (95%CI; 0.2-0.5), NPV 99.9% (95%CI; 99.9 to 100), LR+ 8.17 (95%CI; 7.4 to 9.02), and LR- 0.11 (0.05 to 0.2). Note that setting sensitivity to the same value in both cases was impossible due to finite dataset.

Lastly as shown in Table 1 there was a significant difference in estimated fetal weight 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 requires fewer measures. Missing data was imputed with mean values from the dataset for the logistic regression model.

Figure 4 displays the ROC curves for the three single most informative measures, while Table 3 provides the AUC, specificity**,** sensitivity, PPV, NPV, LP+/- and thresholds for all the ROC curves depicted in Figure 3 and 4.

The MPA/Aao ratio emerged as the most crucial feature, on its own, 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 to 97.1) and a specificity of 61.9% (95% CI; 60.8 to 63.1) for identifying fetuses at risk of developing CoA postnatally. Figure 5 illustrates the relation between the sensitivity and specificity as a function of the threshold.

DISCUSSION

We developed a predictive AI 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 the 18-22-week-scan, which aligns with previous research focused on later gestations[16]. In particular, CoA fetuses exhibited significantly larger right ventricular dimensions, Aao diameter, as well as significantly larger ratios of RV/LV and MPA/Aao, when compared to healthy controls.

Prior research has shown results consistent with our findings, where ROC curves were displayed with AUCs ranging from 0.82-0.98[16] the best of which involved technically challenging ultrasound planes, including the sagittal plane of the aortic arch, and were conducted by experts[3,16,23]. Previous studies have centered on constructing diagnostic prediction models based on CoA cases already suspected of having ventricular imbalance and, consequently, receiving fetal echocardiographies in the third trimester[2,3,12,14–18,23,26]. In contrast, the primary objective in this study was to refine the CoA screening procedure ensuring at-risk fetuses 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 CoA diagnostics[23].

This study involving AI measurements on screening images has yielded results that are 20-40% better than current detection rates[2,3] and comparable to existing expert-based predictions in a pre-selected group[16]. The comparable results can be attributed to the accurate and consistent measurements carried out by AI, the more extensive dataset compared to previous studies[16] , as well as, 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 exceed the performance of those focusing on single features[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 technician) 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 enhance the detection of other cardiac lesions with similar findings.

Previous research has demonstrated that measuring the isthmus in the 3VT is a strong predictor of CoA development[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[3]. For a prudent approach to also reduce the false positive rate, sonographers can acquire the 3VT plane and subsequently the aortic arch in sagittal plane if the prediction model, based on the 4CV and 3VV, flags CoA-risk. This preliminary step may eliminate the need for a comprehensive echocardiography performed by a fetal medicine expert, improving efficiency and cost-effectiveness. However, this approach is only feasible in settings where 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 where 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. The exclusion hampers the direct applicability to current screening programs where the 3VT is an integral component.

Another limitation is that images in our sample spanned over a 10-year period, 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, and the healthy control images did not have the same problem since they were more recent and partly chosen based on image quality. Moreover, significant differences in background characteristics were observed between the CoA cases and the control group

concerning higher BMI of the CoA mothers compared to the controls, which partly can explain the impaired image quality in the case group. The study groups' 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 where 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 ensures 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[28,29].

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 offer not just explanations to clinicians, but also streamline workflow and enable cardiac assessments during routine screenings 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 assessments of health resources is beyond the scope of our research. Nonetheless, identifying at-risk fetuses is important for 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.

Sources of Funding

The Project is supported by The Novo Nordisk Foundation through grant NNFSA170030576, the Danish Regions' AI Signature Project, through the Centre for Basic Machine Learning Research in Life Science (NNF20OC0062606), and the Pioneer Centre for AI, DNRF grant no. P1.

Disclosures

The authors have no conflicts of interest to report.

References

1. Rosamond W, Flegal K, Friday G, Furie K, Go A, Greenlund K, Haase N, Ho M, Howard V, Kissela B, Kittner S, Lloyd-Jones D, McDermott M, Meigs J, Moy C, Nichol G, O'Donnell CJ, Roger V, Rumsfeld J, Sorlie P, Steinberger J, Thom T, Wasserthiel-Smoller S, Hong Y. Heart Disease and Stroke Statistics—2007 Update. *Circulation*. 2007;115(5).
2. Zwanenburg F, Ten Harkel ADJ, Snoep MC, Bet BB, Linskens IH, Knobbe I, Pajkrt E, Blom NA, Clur SAB, Haak MC. Prenatal detection of aortic coarctation in a well-organized screening setting: Are we there yet? *Prenat Diagn* 2023; 45: 620-628.
3. Arya B, Bhat A, Vernon M, Conwell J, Lewin M. Utility of novel fetal echocardiographic morphometric measures of the aortic arch in the diagnosis of neonatal coarctation of the aorta. *Prenat Diagn*. 2016;36(2):127-134.
4. Tegnander E, Eik-Nes SH. The examiner's ultrasound experience has a significant impact on the detection rate of congenital heart defects at the second-trimester fetal examination. *Ultrasound Obstet Gynecol* 2006;28(1).
5. Tegnander E, Eik-Nes SH. OC105: Congenital heart defects and associated extracardiac anomalies in a non-selected population of 42 381 fetuses. *Ultrasound Obstet Gynecol* 2007;30(4).
6. Corcoran S, Briggs K, O'Connor H, Mullers S, Monteith C, Donnelly J, Dicker P, Franklin O, Malone FD, Breathnach FM. Prenatal detection of major congenital heart disease - Optimising resources to improve outcomes. *Eur J Obstet Gynecol Reprod Biol* 2016;203.
7. Budd S, Sinclair M, Day T, Vlontzos A, Tan J, Liu T, Matthew J, Skelton E, Simpson J, Razavi R, Glocker B, Rueckert D, Robinson EC, Kainz B. Detecting Hypo-plastic Left Heart Syndrome in Fetal Ultrasound via Disease-Specific Atlas Maps. In: *Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*. Vol 12907 LNCS. ; 2021.

8. Tan J, Au A, Meng Q, FinesilverSmith S, Simpson J, Rueckert D, Razavi R, Day T, Lloyd D, Kainz B. Automated Detection of Congenital Heart Disease in Fetal Ultrasound Screening. In: *Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*. Vol 12437 LNCS. ; 2020.

9. Arnaout R, Curran L, Zhao Y, Levine JC, Chinn E, Moon-Grady AJ. An ensemble of neural networks provides expert-level prenatal detection of complex congenital heart disease. *Nat Med*. 2021;27(5):882-891.

10. Nurmaini S, Rachmatullah MN, Sapitri AI, Darmawahyuni A, Jovandy A, Firdaus F, Tutuko B, Passarella R. Accurate detection of septal defects with fetal ultrasonography images using deep learning-based multiclass instance segmentation. *IEEE Access* 2020;8.

11. Vedel C, Tabor A, Jørgensen FS, Rode L, Sundberg K, Vejlstrup N, Zingenberg H, Ekelund CK. Prenatal detection rate of major congenital heart defects in Copenhagen from 2015 to 2018. *Ultrasound Obstet Gynecol* 2021;58(2).

12. Freeman K, Kronmal R, Clouse M, Conwell J, Bhat A, Young L, Lewin M, Arya B. Validation of Prenatal Aortic Arch Angle Measurements in the Diagnosis of Neonatal Coarctation of the Aorta. *Pediatr Cardiol* 2021;42(6):1365-1371.

13. Toole BJ, Schlosser B, McCracken CE, Stauffer N, Border WL, Sachdeva R. Importance of Relationship between Ductus and Isthmus in Fetal Diagnosis of Coarctation of Aorta. *Echocardiography*. 2016;33(5):771-777.

14. Hermida U, van Poppel MPM, Lloyd DFA, Steinweg JK, Vigneswaran T V., Simpson JM, Razavi R, De Vecchi A, Pushparajah K, Lamata P. Learning the Hidden Signature of Fetal Arch Anatomy: a Three-Dimensional Shape Analysis in Suspected Coarctation of the Aorta. *J Cardiovasc Transl Res* 2023; 16: 738-747.

15. Gómez-Montes E, García IH, Abad DE, Calvo JR, González CV, Izquierdo AG. Application of a global multiparameter scoring system for the prenatal prediction of coarctation of the aorta. *J Clin Med* 2021;10(16).

16. Familiari A, Morlando M, Khalil A, Sonesson SE, Scala C, Rizzo G, Del Sordo G, Vassallo C, Elena Flacco M, Manzoli L, Lanzone A, Scambia G, Acharya G, D'Antonio F. Risk Factors for Coarctation of the Aorta on Prenatal Ultrasound: A Systematic Review and Meta-Analysis. *Circulation*. 2017;135(8):772-785.

17. Anuwutnavin S, Satou G, Chang RK, DeVore GR, Abuel A, Sklansky M. Prenatal sonographic predictors of neonatal coarctation of the aorta. *J Ultrasound Med*. 2016;35(11):2353-2364.

18. Fricke K, Liuba P, Weismann CG. Fetal Echocardiographic Dimension Indices: Important Predictors of Postnatal Coarctation. *Pediatr Cardiol*. 2021;42(3):517-525.

19. Ronneberger O, Fischer P, Brox T. U-net: Convolutional networks for biomedical image segmentation. In: *Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*. Vol 9351. ; 2015.

20. Ekelund CK, Kopp TI, Tabor A, Petersen OB. The Danish fetal medicine database. *Clin Epidemiol*. 2016;8.

21. Collins GS, Reitsma JB, Altman DG, Moons KGM. Transparent reporting of a multivariable prediction model for individual prognosis or diagnosis (TRIPOD): The TRIPOD statement. *BMJ (Online)*. 2015;350.

22. García-Otero L, Gómez O, Rodriguez-López M, Torres X, Soveral I, Sepúlveda-Martínez Á, Guirado L, Valenzuela-Alcaraz B, López M, Martínez JM, Gratacós E, Crispi F. Nomograms of Fetal Cardiac Dimensions at 18-41 Weeks of Gestation. *Fetal Diagn Ther*. 2020;47(5):387-398.

23. Matsui H, Mellander M, Roughton M, Jicinska H, Gardiner HM. Morphological and physiological predictors of fetal aortic coarctation. *Circulation*. 2008;118(18):1793-1801.

24. Luewan S, Yanase Y, Tongprasert F, Srisupundit K, Tongsong T. Fetal cardiac dimensions at 14-40 weeks' gestation obtained using cardio-STIC-M. *Ultrasound Obstet Gynecol*. 2011;37(4).

25. Pedregosa F, Varoquaux G, Gramfort A, Michel V, Thirion B, Grisel O, Blondel M, Prettenhofer P, Weiss R, Dubourg V, Vanderplas J, Passos A, Cournapeau D, Brucher M, Perrot M, Duchesnay É. Scikit-learn: Machine learning in Python. *JMLR* 2011;12.

26. Contro E, Cattani L, Balducci A, Prandstraller D, Perolo A, Larcher L, Reggiani MLB, Farina A, Donti A, Gargiulo GD, Pilu G. Prediction of neonatal coarctation of the aorta at fetal echocardiography: a scoring system. *J Matern Fetal Neonatal Med*. 2022;35(22):4299-4305.

27. Gómez-Montes E, Herraiz I, Gómez-Arriaga PI, Escribano D, Mendoza A, Galindo A. Gestational age-specific scoring systems for the prediction of coarctation of the aorta. *Prenat Diagn*. 2014;34(12).

28. Tolsgaard MG, Svendsen MBS, Thybo JK, Petersen OB, Sundberg KM, Christensen AN. Does artificial intelligence for classifying ultrasound imaging generalize between different populations and contexts? *Ultrasound Obstet Gynecol*. 2021;57(2):342-343.

29. Sendra-Balcells C, Campello VM, Torrents-Barrena J, Ahmed YA, Elattar M, Botwe BO, Nyangulu P, Stones W, Ammar M, Benamer LN, Kisembo HN, Sereke SG, Wanyonyi SZ, Temmerman M, Mikolaj K, Tolsgaard MG, Lekadir K. Generalisability of fetal ultrasound deep learning models in low-resource imaging settings in five African countries. Sci Rep 2023; 13: 2728.

FIGURE LEGENDS

Figure 1 Segmentation examples.

Left (green): Healthy control with and without AI segmentations.

Right (red): CoA case, without and with AI segmentations.

Three-vessel trachea view (3VV), Four chamber view (4CV), Coarctation aortae (CoA)

Figure 2: Flowchart of the study.

Coarctatio aortae (CoA), three vessel view (3VV), four chamber view (4CV), hypoplastic left heart syndrome (HLHS)

Figure 3: ROC curves for all measures and after feature selection.

All; measures listed in table 2. RV; Right ventricle, LV; Left ventricle, Dao; Descending aorta, Aao; Ascending aorta, MPA; Main pulmonary artery, LA; Left atrium

d; diameter, a; area. AUC; Area under the curve.

Figure 4: ROC curves for single measures.

MPA; Main pulmonary artery, Aao; Ascending aorta, RV; Right ventricle, LV; Left ventricle.

d; diameter, a; area. AUC; Area under the curve.

Figure 5: Relationship between sensitivity and specificity as a function of the threshold for the MPA/Aao diameter ratio measure.

MPA; Main pulmonary artery, Aao; Ascending aorta. d; diameter.

Table 1. Background characteristics of the cohort.

	CoA cases	**Healthy controls**	**p-value**
N	73	7300	
Maternal age (y), *mean(SD)*	30.2 (5.1)	30.5 (4.7)	0.62
BMI, *mean (SD)*	24.6 (5.3)	21.7 (3.4)	<0.001
Parity, *mean (SD)*	0.8 (0.9)	0.8 (0.8)	1
GA at scan (d), *mean (SD)*	140.2 (4.7)	140.5 (4.7)	0.59
EFW* (g), mean SD	**312.8 (50.5)**	**326.9 (38.2)**	**0.0027**
Year of scan, *mean (SD)*	2014 (2)	2016 (2)	N/A
GA at birth (d), *mean (SD)*	271.9 (18.7)	279.8 (10.8)	<0.001
Birthweight (g), *mean (SD)*	3217.7 (871.4)	3527 (413.3)	0.003
Boys, n (%)	41 (56.1%)	3715 (50.9%)	N/A
Conception:			
Spontaneous, n (%)	64 (87.6%)	7300 (100%)	N/A

Differences calculated with welch's t-test.

Y, years; SD, standard deviation; BMI, body mass index; GA, gestational age; d, days; EFW, estimated fetal weight; CoA, coarctation aortae

Some cases lacked the necessary measurements for inclusion in the 4-parameter Hadlock formula.

Missing N (cases)=6, N (controls)=362.

Table 2. Anatomical measures and Z-scores.

Anatomy	**CoA mean (SD)**	**Control mean (SD)**	**p-value**	**Z-score**	**Coefficients (All)**	**Coefficients (Feature Selection)**
N (4CV)	73	7300				
RV diameter	0.61 (0.10)	0.56 (0.08)	<.0001	0.63	-0.031	-
RV length	1.15 (0.18)	1.03 (0.16)	<.0001	0.75	0.825	-
RV area	0.45 (0.13)	0.38 (0.09)	<.0001	0.78	1.781	1.160
LV diameter	0.54 (0.11)	0.61 (0.10)	<.0001	-0.7	-0.014	-
LV length	1.06 (0.16)	1.10 (0.18)	0.0249	-0.22	0.046	-
LV area	0.38 (0.12)	0.46 (0.13)	<.0001	-0.62	-2.749	-0.970
RV/LV area ratio	1.22 (0.26)	0.87 (0.44)	<.0001	0.80	-1.017	-
RA diameter	0.58 (0.12)	0.59 (0.10)	0.3232	-0.1	0.701	-
RA length	0.76 (0.13)	0.79 (0.13)	0.1043	-0.23	0.136	-
RA area	0.32 (0.11)	0.34 (0.10)	0.1366	-0.2	-0.726	-

LA diameter	0.48 (0.11)	0.54 (0.11)	<.0001	-0.55	0.413	-
LA length	0.66 (0.12)	0.72 (0.14)	<.0001	-0.43	0.045	-
LA area	0.22 (0.08)	0.27 (0.9)	<.0001	-0.56	-0.877	-0.388
Mitral valve diameter	0.46 (0.13)	0.51 (0.14)	0.0007	-0.36	0.073	-
Tricuspid valve diameter	0.59 (0.13)	0.60 (0.13)	0.4552	-0.08	-0.240	-
N (3VV)	73	7300				-
MPA diameter	0.39 (0.07)	0.36 (0.07)	0.0068	0,43	0.922	0.627
Aao diameter	0.23 (0.06)	0.33 (0.06)	<.0001	-1.67	-1.841	-1.411
Dao diameter	0.19 (0.06)	0.22 (0.5)	<.0001	-0.75	-0.317	-0.347
MPA/Aao diameter ratio	1.67 (0.48)	1.09 (0.20)	<.0001	2.9	0.043	-
Intercept	-	-	-	-	-8.031	-6.657

All measurements are reported in cm.

P-value of Welch's t-test between mean of CoA and control. Coefficients from logistic regressions.

CoA, coarctation aortae; 4CV, four-chamber view; 3VV, three-vessel view;

RV, right ventricle; LV, left ventricle; RA, right atrium; LA, left atrium; VS, ventricle septum;

MPA, main pulmonary artery; Aao, ascending aorta; Dao, descending aorta

Table 3. Predictive parameters, AUC, sensitivity of 90.3% and corresponding specificity.

Predictive parameter(s)	AUC	Sensitivity (95%CI)	Specificity (95% CI)	PPV	NPV	LR+	LR-	Threshold
All measures from table 2	0.9565	90.3 (83.4-97.1)	84.8 (84.0-85.6)	0.24 (0.1-0.4)	99.9 (99.9-100)	5.94 (5.41-6.52)	0.11 (0.06-0.23)	0.0051 *
RV_a, LV_a, Dao_d, Aao_d, MPA_d, LA_a	0.9625	90.4 (83.7-97.2)	88.9 (88.2-89.6)	0.33 (0.2-0.5)	99.9 (99.9-100)	8.17 (7.40-9.02)	0.11 (0.05-0.23)	0.0095 *
MPA_d/Aao_d	0.8978	90.3 (83.4-97.1)	61.9 (60.8-63.1)	0.09 (0.02-0.17)	99.9 (99.9-100)	2.37 (2.18-2.57)	0.15 (0.08-0.32)	1.1500
Aao_d	0.8864	90.41 (83.7-97.2)	59.5 (58.4-60.7)	0.09 (0.02-0.16)	99.9 (99.9-100)	2.23 (2.06-2.42)	0.16 (0.08-0.33)	0.3240
RV_a/LV_a	0.8784	90.4 (83.7-97.2)	67.1 (66.0-68.2)	0.11 (0.03-0.19)	99.9 (99.9-100)	2.75 (2.53-2.98)	0.14 (0.07-0.29)	0.8900

RV; Right ventricle, LV; Left ventricle, Dao; Descending aorta, Aao; Ascending aorta, MPA; Main pulmonary artery, LA; Left atrium

d; diameter, a; area l; length. AUC; Area under the curve. PPV; Positive predictive value, NPV; negative predictive value, LR+; Positive Likelihood ratio, LR-; Negative Likelihood ratio

* *indicates that the threshold is applied to the output of the logistic regression; otherwise, the threshold is applied directly to the measurement.*