

RESEARCH ARTICLE

Using deep-learning in fetal ultrasound analysis for diagnosis of cystic hygroma in the first trimester

Mark C. Walker^{1,2,3,4,5,6,*}, Inbal Willner^{1,5}, Olivier X. Miguel², Malia S. Q. Murphy², Darine El-Chaâr^{1,2,4,5}, Felipe Moretti^{1,5}, Alysha L. J. Dingwall Harvey², Ruth Rennicks White^{2,5}, Katherine A. Muldoon^{1,2}, André M. Carrington^{7,8}, Steven Hawken^{2,4}, Richard I. Aviv^{8,9,10}

1 Department of Obstetrics and Gynecology, University of Ottawa, Ottawa, Canada, **2** Clinical Epidemiology Program, Ottawa Hospital Research Institute, Ottawa, Canada, **3** International and Global Health Office, University of Ottawa, Ottawa, Canada, **4** School of Epidemiology and Public Health, University of Ottawa, Ottawa, Canada, **5** Department of Obstetrics, Gynecology & Newborn Care, The Ottawa Hospital, Ottawa, Canada, **6** BORN Ontario, Children's Hospital of Eastern Ontario Research Institute, Ottawa, Canada, **7** Department of Systems Design Engineering, University of Waterloo, Waterloo, Canada, **8** Department of Radiology and Medical Imaging, University of Ottawa, Ottawa, Canada, **9** Department of Radiology and Medical Imaging, The Ottawa Hospital, Ottawa, Canada, **10** Neuroscience Program, Ottawa Hospital Research Institute, Ottawa, Canada

* mwalker@toh.ca



OPEN ACCESS

Citation: Walker MC, Willner I, Miguel OX, Murphy MSQ, El-Chaâr D, Moretti F, et al. (2022) Using deep-learning in fetal ultrasound analysis for diagnosis of cystic hygroma in the first trimester. PLoS ONE 17(6): e0269323. <https://doi.org/10.1371/journal.pone.0269323>

Editor: Ruxandra Stoean, University of Craiova, ROMANIA

Received: December 24, 2021

Accepted: May 19, 2022

Published: June 22, 2022

Peer Review History: PLOS recognizes the benefits of transparency in the peer review process; therefore, we enable the publication of all of the content of peer review and author responses alongside final, published articles. The editorial history of this article is available here: <https://doi.org/10.1371/journal.pone.0269323>

Copyright: © 2022 Walker et al. This is an open access article distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Data Availability Statement: The ultrasound images used in this study are not currently available to upload to PLOS ONE. All images belong to The Ottawa Hospital. A formal request can be

Abstract

Objective

To develop and internally validate a deep-learning algorithm from fetal ultrasound images for the diagnosis of cystic hygromas in the first trimester.

Methods

All first trimester ultrasound scans with a diagnosis of a cystic hygroma between 11 and 14 weeks gestation at our tertiary care centre in Ontario, Canada were studied. Ultrasound scans with normal nuchal translucency were used as controls. The dataset was partitioned with 75% of images used for model training and 25% used for model validation. Images were analyzed using a DenseNet model and the accuracy of the trained model to correctly identify cases of cystic hygroma was assessed by calculating sensitivity, specificity, and the area under the receiver-operating characteristic (ROC) curve. Gradient class activation heat maps (Grad-CAM) were generated to assess model interpretability.

Results

The dataset included 289 sagittal fetal ultrasound images; 129 cystic hygroma cases and 160 normal NT controls. Overall model accuracy was 93% (95% CI: 88–98%), sensitivity 92% (95% CI: 79–100%), specificity 94% (95% CI: 91–96%), and the area under the ROC curve 0.94 (95% CI: 0.89–1.0). Grad-CAM heat maps demonstrated that the model predictions were driven primarily by the fetal posterior cervical area.

made to the IRB, who would then consult a privacy officer at The Ottawa Hospital. Requests can be sent to the corresponding author (Dr. Mark Walker) or the Ottawa Health Sciences Network Research Ethics Board (rebadministration@ohri.ca)/(<http://www.ohri.ca/ohsn-reb/contacts.htm>).

Funding: This study was supported by a Canadian Institutes of Health Research Foundation Grant (FDN 148438)(MCW). The funding agency played no role in the study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing interests: The authors have declared that no competing interests exist.

Conclusions

Our findings demonstrate that deep-learning algorithms can achieve high accuracy in diagnostic interpretation of cystic hygroma in the first trimester, validated against expert clinical assessment.

Introduction

Artificial intelligence (AI) and machine learning models are increasingly being applied in clinical diagnostics, including medical imaging [1]. Deep-learning is a class of machine learning models inspired by artificial neural networks which can process large amounts of data to identify important features that are predictive of outcomes of interest. Importantly, model performance can be continuously and incrementally improved as data accrues. Deep-learning models perform well in pattern recognition, making them particularly useful in the interpretation of medical images [2]. A systematic review and meta-analysis of studies reporting on the diagnostic performance of deep-learning models in the identification of disease features from medical images found that they performed equivalently to trained clinicians, however, few studies make direct comparisons between deep-learning models and health care professionals [2].

AI has many potential uses in ultrasonography, including automating standardized plane detection, and extracting quantitative information about anatomical structure and function. Inter-operator heterogeneity in ultrasound image acquisition poses unique challenges for the use of AI in this space, and are further impeded by a lack of standardized imaging planes [3]. Still, AI has been successfully tested in a range of organ systems, and there is growing interest for its application in obstetric ultrasonography for fetal structure identification, automated measurement of fetal growth parameters, and for the diagnosis of congenital anomalies [4]. Uptake of AI applications into clinical practice is beginning, however, ongoing investigation is warranted to demonstrate the feasibility, validity and reliability of such tools.

Cystic hygroma is a congenital lymphangioma documented in approximately 1 in 800 pregnancies and 1 in 8 000 live births [5]. It is commonly associated with chromosomal abnormalities including Trisomy 21, Turner Syndrome, and anatomical malformations [6–8]. Cystic hygroma is diagnosed based on the assessment of nuchal translucency (NT) thickness between the fetal skin and the subcutaneous soft tissue at the neck and cervical spine [4]. Diagnosis is straightforward and possible through routine first- or second-trimester ultrasonography. Affected pregnancies require extensive antenatal and postpartum management which may include early cytogenetic testing for suspected aneuploidy, comprehensive assessment of fetal anatomy, and postpartum surgical intervention [8, 9]. Some studies have investigated automated and semi-automated systems for measuring NT thickness, however, no studies to date have investigated AI and deep-learning models to assess diagnoses associated with NT thickness, including cystic hygroma [10, 11].

Given the unique challenges in applying AI in fetal ultrasonography, cystic hygroma is an ideal condition to investigate the feasibility of using deep-learning models in the interpretation of ultrasound images. We sought to develop a deep-learning model that could analyze fetal ultrasound images and correctly identify cases of cystic hygroma compared to normal controls.

Methods

Study setting and design

This was a retrospective study conducted at The Ottawa Hospital, a multi-site tertiary-care facility in Ottawa, Canada, with a catchment area of 1.3 million people. First trimester

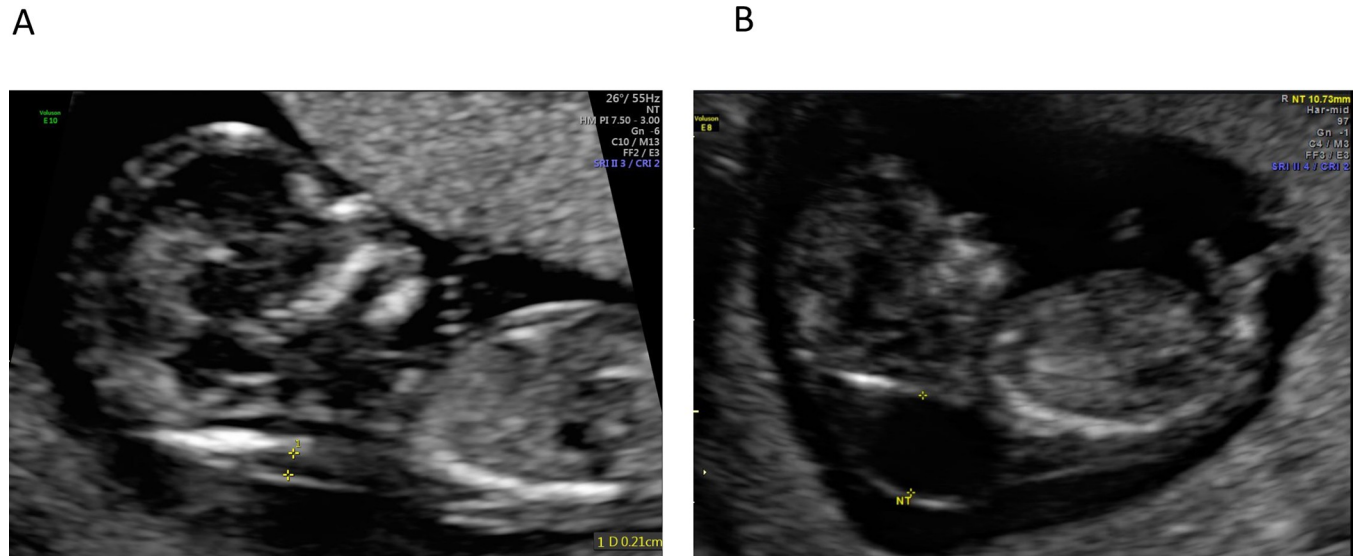


Fig 1. Fetal ultrasound images of normal (A) and cystic hygroma (B) scans.

<https://doi.org/10.1371/journal.pone.0269323.g001>

ultrasound images taken between March 2014 and March 2021 were retrieved from the institutional Picture Archiving and Communication System (PACS) and saved in Digital Imaging and Communications in Medicine (DICOM) format. Eligible images were those that included a mid-sagittal view of the fetus taken between 11 and 14 weeks gestation. Cases were identified if there was a final clinical diagnosis of a cystic hygroma in the ultrasound report (Fig 1A). A set of normal ultrasound images from NT screens were used for controls and were retrieved between March 2021 and June 2021 (Fig 1B). Cases and normal images were reviewed and verified by a clinical expert (IW). Patients were not contacted and patient consent was not required to access the images. Data were de-identified and fully anonymized for the model training. This study was reviewed and approved by the Ottawa Health Sciences Network Research Ethics Board (OHSN REB #20210079).

A 4-fold cross-validation (4CV) design was used, whereby the same deep-learning architecture was tested and trained four different times using randomly partitioned versions (folds) of the image dataset. For each fold, 75% of the dataset was used for model training, and 25% was used for model validation. The 4CV design was chosen instead of the more commonly used 10-fold cross-validation (10CV) design to optimize the performance of the deep-learning models within our small dataset. With 4CV, each prediction error affects the accuracy of the model by 1.4%, and the sensitivity by 3.1%. Had a 10CV approach been used, each prediction error would have affected model accuracy by 3.4% and the sensitivity by 7.8%.

Data preparation

DICOM images included coloured annotations such as calipers, text, icons, profile traces (Fig 2A) and patient personal health information (PHI), which were removed prior to analysis. PHI was handled by cropping the images to remove the identifying information contained in the image borders. Coloured annotations were removed by first converting image data from the Red Green Blue (RGB) colour space to the Hue Saturation Value (HSV) colour space. Image pixels belonging to the grey ultrasound image were empirically identified and ranged from 0–27, 0–150 and 0–255 for the H, S and V values, respectively (Fig 2B).

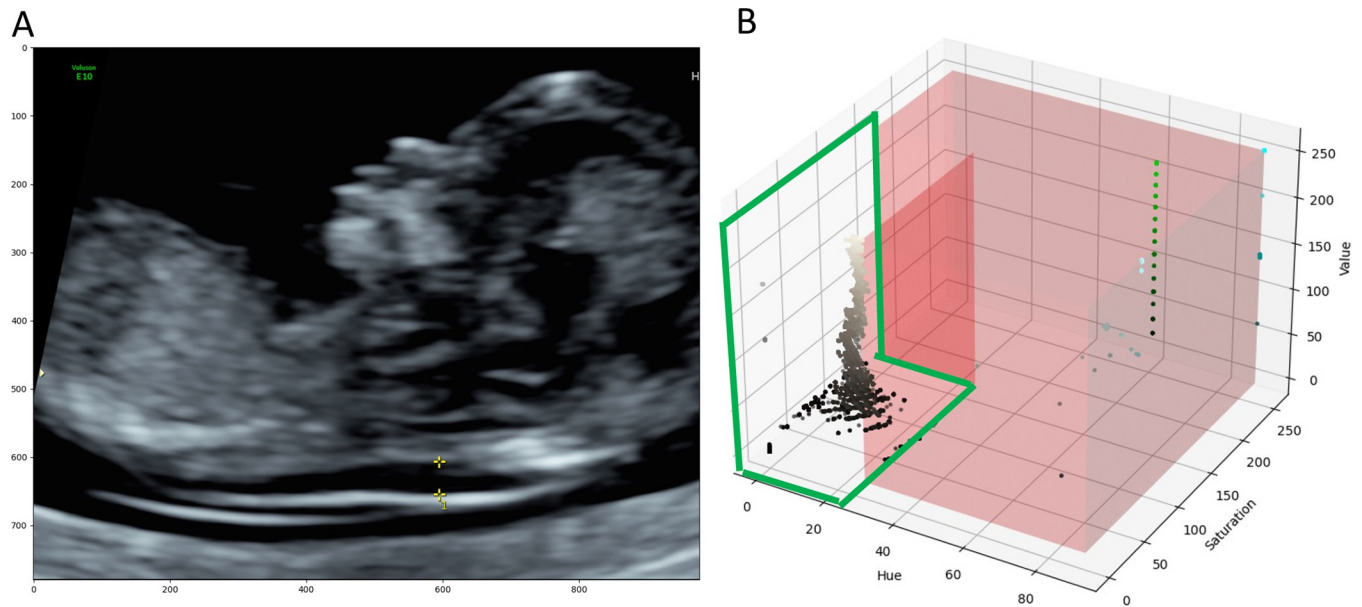


Fig 2. Identifying image annotations on a normal NT scan. (A) Image annotations included calipers, text, icons, and profile traces, all of which were removed prior to model training. (B) 3D Scatter Plot of HSV image data. Each point represents one image pixel and its associated HSV values. The red region highlights the range of values which do not belong to the grayscale ultrasound image. The area encircled in green shows pixel values that belong to the grayscale ultrasound image. Grayscale images had H, S and V values ranging from 0–27, 0–150 and 0–255, respectively.

<https://doi.org/10.1371/journal.pone.0269323.g002>

Pixels outside of these ranges were determined to be part of the coloured annotations. Third, a binary mask of the image was created where annotation pixels were labelled ‘1’ and ultrasound image pixels were labelled ‘0’. The binary mask was then dilated with a 5x5 kernel to include contours surrounding the annotations. Last, the Navier-Stokes image infill method [12] was used to artificially reconstruct the ultrasound image without annotations (Fig 3).

After DICOM images were cleaned, they were converted to grayscale (1 channel image). Intensities were standardized to a mean of zero and a standard deviation of one for better stability during the neural network training [13]. Finally, the images were resized to 256 x 256 pixels.

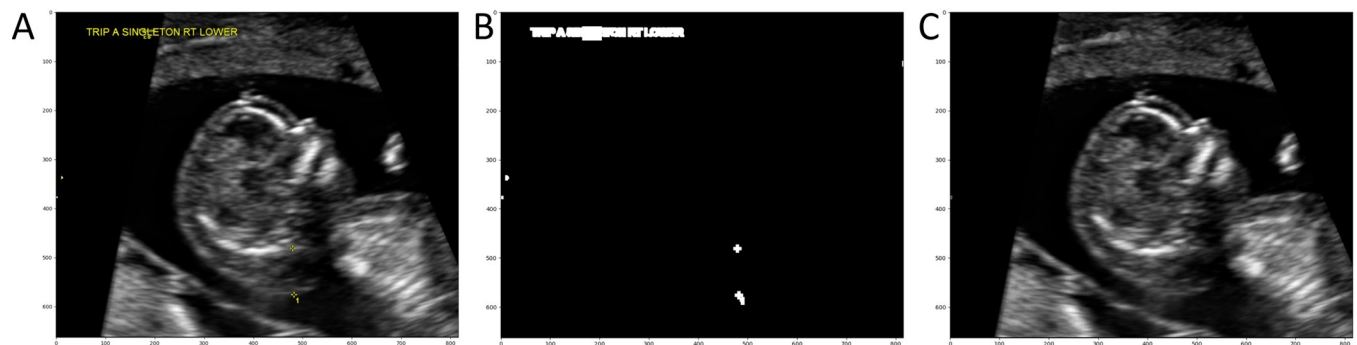


Fig 3. Removal of image annotations on a scan with cystic hygroma diagnosis. (A) Ultrasound image before annotations were removed. Yellow calipers (bottom middle) are visible, along with text annotations (top left). (B) The binary mask of the image which was generated to define the region of the image that need to be infilled (white pixels). (C) Result of the Navier-Stokes image infill method; all image annotations have been removed.

<https://doi.org/10.1371/journal.pone.0269323.g003>

Neural network model

A DenseNet convolutional neural network (CNN) model architecture [14] was used to classify images as “normal” or those with “cystic hygroma”. Specifically, we used the DenseNet169 PyTorch model [15], with the input layer modified to have a shape of 256 x 256 x 1 pixels and the output layer modified to produce two possible outcome features: 1) normal and 2) cystic hygroma. DenseNet169 was chosen because of its advantages over the ResNet architecture as shown in a publication by Gao et al. (2017) demonstrating that DenseNet achieves high performance while requiring less computation [12]. On some image classification benchmarks DenseNet demonstrated better performance and efficiency than the ResNet model.

Model training

Model training was performed from scratch with random weights initialization for 1000 epochs (i.e., step-in optimization) using the cross-entropy loss function [16], the Adam optimizer [17, 18] (epsilon = 1e-8, beta1 = 0.9, beta2 = 0.999) and a batch size of 64. The available pretrained DenseNet models were trained on ImageNet. ImageNet and fetal ultrasound images are significantly different, therefore CNN models for fetal US applications should not be pre-trained on ImageNet. We trained the architecture using our dataset of ultrasound images from The Ottawa Hospital with random weights initialization. The learning rate was set to 1e-2 and reduced by a factor of 0.72 at every 100 epochs (gamma = 0.72, step size = 100) using a learning rate step scheduler [19].

Data augmentation was used during model training to increase the generalizability of the CNN model. Augmentations included random horizontal flip of the ultrasound images (50% probability), random rotation in the range of $[-15, 15]$ degrees, random translation ($x \pm 10\%$ and $y \pm 30\%$ of image size) and random shearing in the range of $[-0.2, 0.2]$ degrees. These augmentations were performed dynamically at batch loading. Due to the randomness of these operations, a single training image underwent different augmentations at each epoch.

To address imbalance in the number of normal NT and cystic hygroma images in the training dataset, cystic hygroma images were randomly up sampled, with replacement, to match the number of normal NT images.

Model validation

For each epoch, the validation data were used to assess the total number of true and false positives, and true and false negatives and used to calculate accuracy, sensitivity, specificity and the area under the receiver-operating characteristic curve (AUC). For each fold, the performance metrics for the epoch with the highest level of accuracy were reported. The mean, standard deviation and 95% confidence intervals (CI) of the performance metrics across all folds were then computed.

Explainability

The Gradient-weighted Class Activation Mapping (Grad-CAM) method was used to improve the interpretability of our trained DenseNet models, and visually contextualize important features in the image data that were used for model predictions [20]. Grad-CAM is a widely used technique for the visual explanation of deep-learning algorithms [21, 22]. With this approach, heat maps were generated from 8x8 feature maps to highlight regions of each image that were the most important for model prediction (Fig 4).

Pr(class=Cystic Hygroma) = 1.00

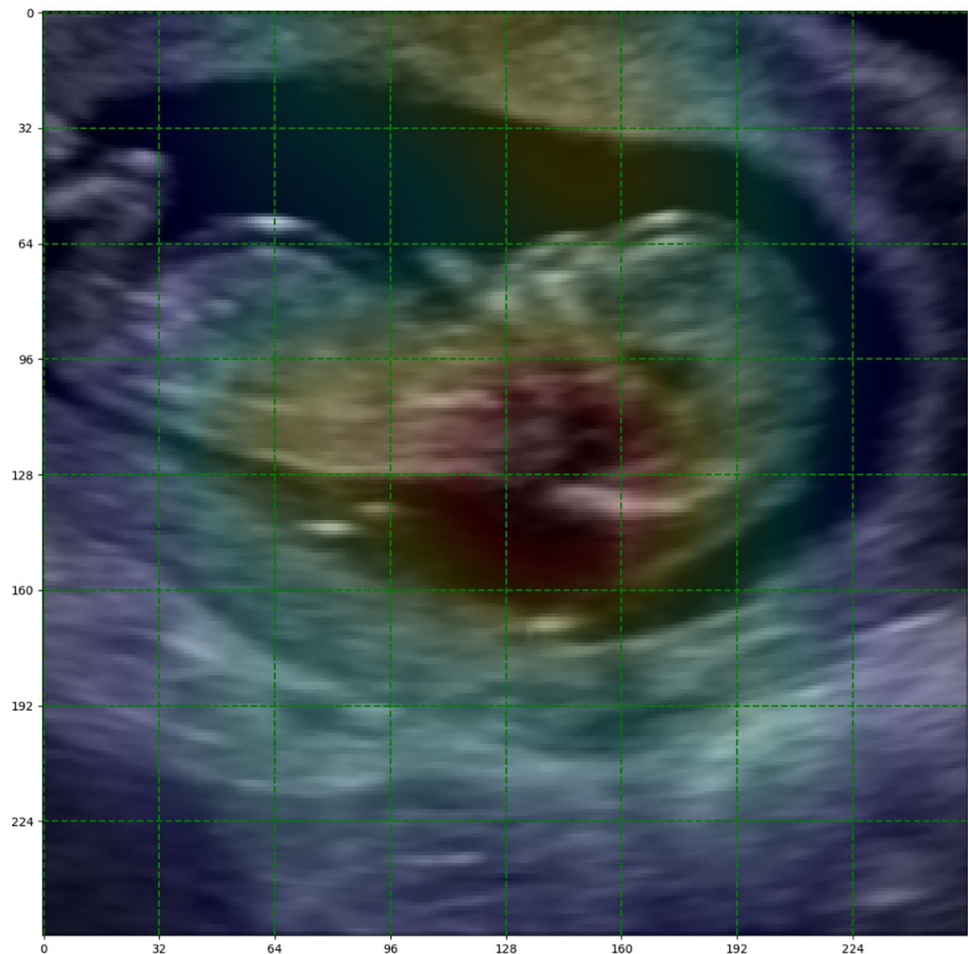


Fig 4. Grad-CAM image of a cystic hygroma case. The green gridlines indicate the size of the feature maps (8x8) used to generate the heat maps. The red highlights the region of the image that influenced the model's prediction the most.

<https://doi.org/10.1371/journal.pone.0269323.g004>

Results

The image dataset included 289 unique ultrasound images; 160 control images with normal NT measurements, and 129 cases with cystic hygroma. A total of 217 images were used for model training, and 72 images were used for model validation (Table 1).

Table 2 shows the results for all 4 cross validation folds. All 4 models performed well in the validation set. The overall mean accuracy was 93% (95% CI: 88–98%), and the area under the receiver operating characteristic curve was 0.94 (95% CI: .89–1.0) (Fig 5). The sensitivity was 92% (95% CI: 79–100%) and the specificity was 94% (95% CI: 91–96%).

Most of the Grad-CAM heat maps highlighted the fetal head and neck (Fig 6). Although some heat maps specifically highlighted the posterior cervical region, in the area used for NT measurement (Fig 7A and 7B), poor localization did occur (Fig 7C and 7D).

There were 10 false negatives and 10 false positives. The misclassified images were reviewed with a clinical expert (MW) and it was determined that misclassifications commonly happened

Table 1. Partitioning of data across training and validation datasets^a.

	Overall, n (%)	Normal NT images, n (%)	Cystic hygroma images, n (%)
Total dataset	289 (100%)	160 (100%)	129 (100%)
Training dataset	217 (75.1%)	120 (75%)	97 (75.2%) ^b
Validation dataset	72 (24.9%)	40 (25%)	32 (24.8%)

NT, nuchal translucency.

^aColumn statistics are provided.

^b97 original images; 23 cystic hygroma images were randomly resampled to reduce imbalance between the two groups in the training dataset, to produce in a final cystic hygroma training dataset of 120 images.

<https://doi.org/10.1371/journal.pone.0269323.t001>

when the fetus was close to the placental membrane leading the heat map to focus on another part of the brain.

Discussion

Our findings demonstrate the feasibility of using deep-learning models to interpret fetal ultrasound images and identify cystic hygroma diagnoses with high performance in a dataset of first trimester ultrasound scans. The model achieved excellent prediction of cystic hygroma with a sensitivity of 92% (95% CI: 79–100%) and specificity of 94% (95% CI: 91–96%). This study contributes to the literature on AI and medical diagnostics, and more specifically to the use of AI in fetal ultrasonography where there are scant data.

Ultrasound is critical in the observation of fetal growth and development, however, small fetal structures, involuntary fetal movements and poor image quality make neonatal image acquisition and interpretation challenging. Our study has shown that deep-learning and Grad-CAM heat maps can correctly identify the fetal head and neck region to identify abnormalities. There have been several investigations focusing on AI-based localization of standard planes in fetal ultrasonography, suggesting that AI-models perform as well as clinicians in obtaining reasonable planes for image capture and diagnosis [23–27]. Building on this literature, cystic hygroma has not been evaluated yet it is an ideal fetal diagnostic condition to assess the accuracy of AI-based models because it is a clearly visible diagnosis to the trained expert.

More recently, others have sought to apply machine learning methods to the identification of fetal malformations, with promising results. Xie et al. developed and tested a CNN-based

Table 2. 4-fold cross validation results^a.

	FOLD NUMBER				OVERALL PERFORMANCE	
	Fold 0	Fold 1	Fold 2	Fold 3	Mean±SD	95% Confidence Interval
True positives, n	27	29	32	31	29.75±1.92	26.0–33.5
True negatives, n	38	38	37	37	37.50±0.50	36.5–38.5
False positives, n	2	2	3	3	2.50±0.50	1.5–3.5
False negatives, n	6	3	0	1	2.50±2.29	0.0–7.0
Accuracy	0.89	0.93	0.96	0.94	0.93±0.03	0.88–0.98
Sensitivity	0.82	0.91	1.00	0.97	0.92±0.07	0.79–1.0
Specificity	0.95	0.95	0.93	0.93	0.94±0.01	0.91–0.96
AUC	0.91	0.93	0.98	0.95	0.94±0.03	0.89–1.0

AUC, area under the curve; SD, standard deviation.

^aDecision threshold in the receiver-operating characteristic (ROC) curve was set at 0.5 in the [0,1] range of predicted probability (or class membership).

<https://doi.org/10.1371/journal.pone.0269323.t002>

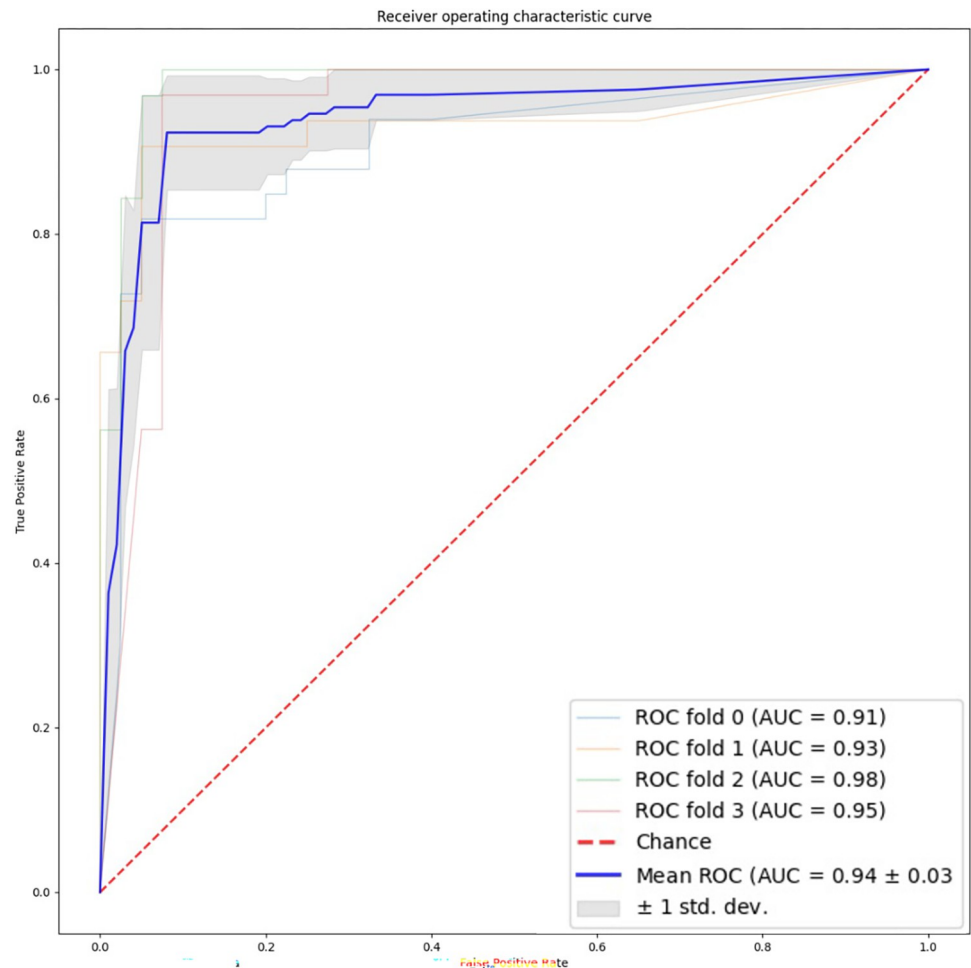


Fig 5. Receiver operating characteristic plot summarizing performance of all four cross validation folds.

<https://doi.org/10.1371/journal.pone.0269323.g005>

deep-learning model in a dataset of nearly 30,000 fetal brain images, including over 14,000 images with common central nervous system abnormalities [28]. Although, the final model was able to discriminate well between normal and abnormal images (sensitivity and specificity were 96.9% and 95.9%, respectively) in a hold out test set, the models were not trained to distinguish between specific brain abnormalities and did not have information on cystic hygroma diagnoses. In a large sample of 2D ultrasounds, fetal echocardiograms and anatomy scans from second trimester pregnancies, Arnaout et al. trained CNN models to accurately localize the fetal heart and detect complex congenital heart disease [29]. Their models achieved high sensitivity (95%, 95%CI, 84–99), and specificity (96%, 95%CI, 95–97) and were validated in several independent datasets. Finally, Baumgartner et al. [25] demonstrated the potential for such systems to operate in real-time. Their findings suggest that deep-learning models could be used “live” to guide sonographers with image capture and recognition during clinical practice. The model proposed by Baumgartner et al. was trained for detection of multiple fetal standard views in freehand 2D ultrasound data. They achieved a 91% recall (i.e., sensitivity) on the profile standard views which is close to the evaluation metrics obtained in our study. Although published data on deep-learning models for the identification of specific malformations are generally lacking, our findings, combined with those of others in this space demonstrate the feasibility of deep-learning models for supporting diagnostic decision-making.

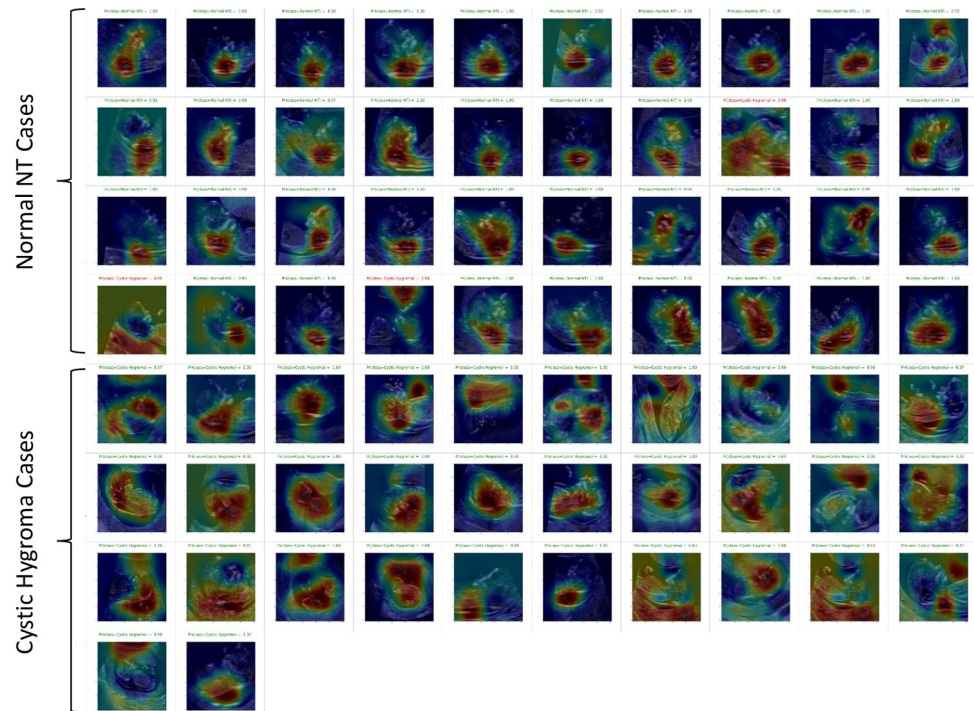


Fig 6. Grad-CAM heat maps for the full validation set of Fold 2. Top 4 rows are normal NT cases and bottom 4 rows are cystic hygroma cases. Red colours highlight regions of high importance and blue colours highlight regions of low or no importance. Therefore, a good model would have Grad-CAM heatmaps that highlight the head and neck area for both normal and cystic hygroma images.

<https://doi.org/10.1371/journal.pone.0269323.g006>

Using the latest data augmentation and image dimensionality techniques, we have developed a deep-learning algorithm with very good predictive accuracy on a relatively small dataset. The strengths of this study include our use of the k-fold cross-validation experiment design. Although computationally intensive, k-fold cross-validation reduces model bias and variance, as most of the data are used for model training and validation. In addition, removal of calipers and text annotations from ultrasound images established a dataset free from the clinical bias on which to develop our models. Annotations on routinely collected ultrasound images have historically limited their utility for medical AI research [30]. Furthermore, our use of Grad-CAM heat maps enabled transparent reporting of how the deep-learning models performed on a case-by-case basis. With this approach, we were able to confirm excellent localization for most of the images used in our dataset. Class-discriminative visualization enables the user to understand where models fail (i.e., why the models predicted what they predicted) and can be used to inform downstream model enhancements. Additionally, all false negative and false positive images were reviewed and it was determined that misclassifications commonly occurred when the fetus was close to the placental membrane. Future work could collect more images where the fetus is close to membrane, up-sample the images that were error prone, and further train the model to incorporate this pattern.

Our study is not without limitations. First, as a single centre study, the sample size available for developing and validating deep-learning model was relatively small. However, use of data augmentation to increase the variability in our dataset, enrichment of cystic hygroma cases in our training set, and use of the k-fold cross validation experiment design are all well-accepted strategies to overcome the limitations of small datasets [31]. Second, although we removed all image annotations, we cannot discount the possibility that the blending and infill methods used to reconstitute the

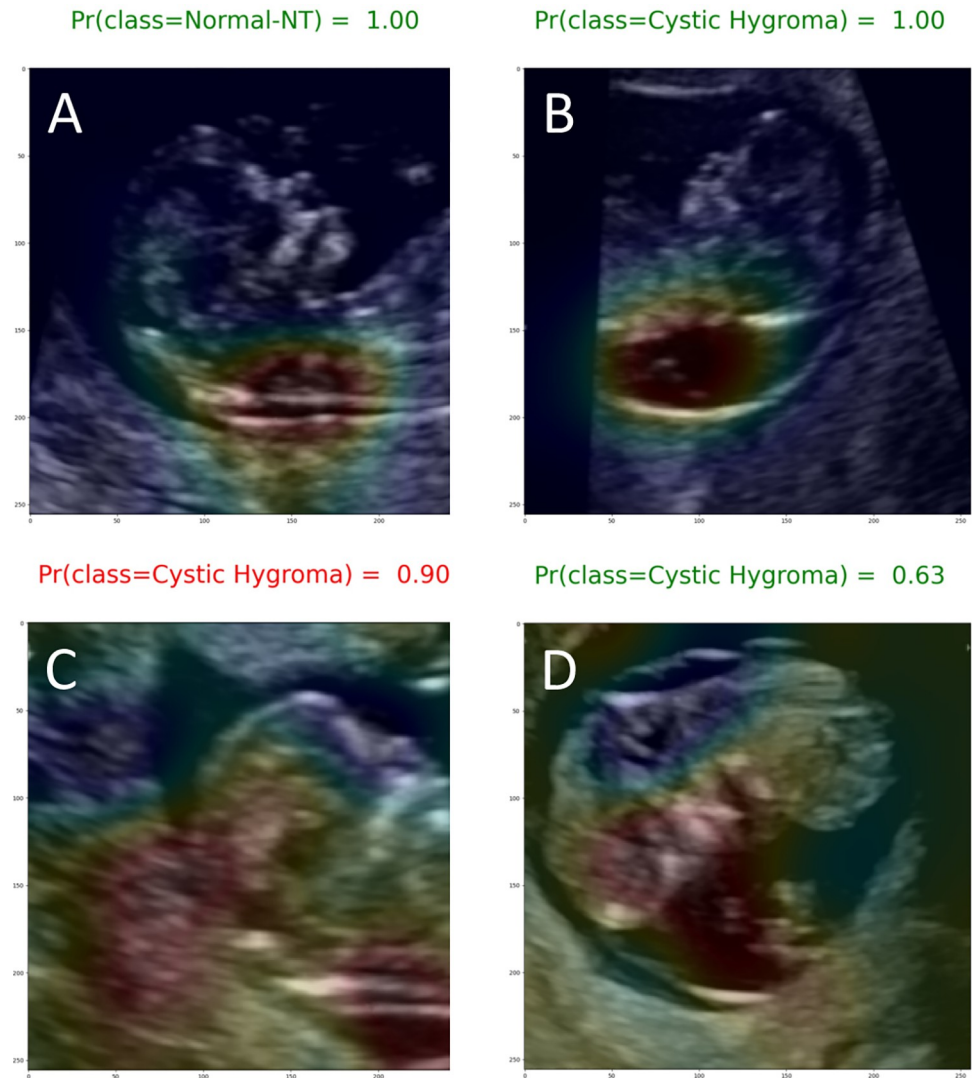


Fig 7. Exemplary Grad-CAM heat maps. (A) Normal NT case with good localization in which the model predicted the correct class with a high (1.00) output probability (true negative). (B) Cystic hygroma case with good localization in which the model predicted the correct class with a high (1.00) output probability (true positive). (C) Normal NT case showing poor localization in which the model predicted this class incorrectly with a 0.90 output probability (false positive). (D) Cystic hygroma case showing poor localization in which the model predicted the correct class, but with an output probability that suggests uncertainty (0.63) (true positive).

<https://doi.org/10.1371/journal.pone.0269323.g007>

images influenced the deep-learning algorithm. However, the Grad-CAM heatmaps provide reassurance that fetal craniocervical regions were driving the deep-learning algorithm, and that the model appropriately places high importance on regions which are clinically relevant for diagnosis. Given the relatively low incidence of congenital anomalies such as cystic hygroma, a natural extension of this work will be to introduce our models to a larger, multi-centre dataset with more variability in the image parameters and greater feature variety specific to cystic hygroma.

Conclusions

In this proof-of-concept study, we demonstrate the potential for deep-learning to support early and reliable identification of cystic hygroma from first trimester ultrasound scans. We

present a novel application of convolutional neural networks to automatically identify cases of cystic hygroma and localise the relevant fetal structures for clinical decision-making. With further development, including testing in a large multi-site dataset and external validation, our approach may be applied to a range of other fetal anomalies typically identified by ultrasonography.

Acknowledgments

The authors would like to acknowledge that this study took place on unceded Algonquin Anishinabe territory.

Author Contributions

Conceptualization: Mark C. Walker, Olivier X. Miguel, Steven Hawken.

Data curation: Inbal Willner, Olivier X. Miguel.

Formal analysis: Olivier X. Miguel, Steven Hawken.

Funding acquisition: Mark C. Walker.

Investigation: Mark C. Walker, Inbal Willner, Darine El-Chaâr, Felipe Moretti, André M. Carrington, Steven Hawken, Richard I. Aviv.

Methodology: Katherine A. Muldoon, André M. Carrington, Steven Hawken.

Project administration: Alysha L. J. Dingwall Harvey, Ruth Rennicks White.

Resources: Mark C. Walker.

Supervision: Mark C. Walker, Darine El-Chaâr, Felipe Moretti, Alysha L. J. Dingwall Harvey, Ruth Rennicks White, André M. Carrington, Steven Hawken.

Writing – original draft: Olivier X. Miguel, Malia S. Q. Murphy, Katherine A. Muldoon.

Writing – review & editing: Mark C. Walker, Inbal Willner, Olivier X. Miguel, Malia S. Q. Murphy, Darine El-Chaâr, Felipe Moretti, Alysha L. J. Dingwall Harvey, Ruth Rennicks White, Katherine A. Muldoon, André M. Carrington, Steven Hawken, Richard I. Aviv.

References

1. Drukker L, Noble JA, Papageorghiou AT. Introduction to artificial intelligence in ultrasound imaging in obstetrics and gynecology. *Ultrasound Obstet Gynecol* [Internet]. 2020 Oct; 56(4):498–505. Available from: <https://onlinelibrary.wiley.com/doi/10.1002/uog.22122>
2. Liu X, Faes L, Kale AU, Wagner SK, Fu DJ, Bruynseels A, et al. A comparison of deep learning performance against health-care professionals in detecting diseases from medical imaging: a systematic review and meta-analysis. *Lancet Digit Heal* [Internet]. 2019 Oct; 1(6):e271–97. Available from: <https://linkinghub.elsevier.com/retrieve/pii/S2589750019301232> [https://doi.org/10.1016/S2589-7500\(19\)30123-2](https://doi.org/10.1016/S2589-7500(19)30123-2) PMID: 33323251
3. Park SH. Artificial intelligence for ultrasonography: unique opportunities and challenges. *Ultrasonography* [Internet]. 2021 Jan 1; 40(1):3–6. Available from: <http://e-ultrasonography.org/journal/view.php?doi=10.14366/usg.20078> PMID: 33227844
4. Chen Z, Liu Z, Du M, Wang Z. Artificial Intelligence in Obstetric Ultrasound: An Update and Future Applications. *Front Med* [Internet]. 2021 Aug 27; 8. Available from: <https://www.frontiersin.org/articles/10.3389/fmed.2021.733468/full> PMID: 34513890
5. The fetal medicine foundation. Cystic Hygroma [Internet]. [cited 2021 Nov 17]. Available from: <https://fetalmedicine.org/education/fetal-abnormalities/neck/cystic-hygroma>
6. Chen Y-N, Chen C-P, Lin C-J, Chen S-W. Prenatal Ultrasound Evaluation and Outcome of Pregnancy with Fetal Cystic Hygromas and Lymphangiomas. *J Med Ultrasound* [Internet]. 2017 Mar; 25(1):12–5.

- Available from: <http://linkinghub.elsevier.com/retrieve/pii/S0929644117300061> <https://doi.org/10.1016/j.jmu.2017.02.001> PMID: 30065449
7. Mirza B, Ijaz L, Saleem M, Sharif M, Sheikh A. Cystic hygroma: An overview. *J Cutan Aesthet Surg* [Internet]. 2010; 3(3):139. Available from: <http://www.jcasonline.com/text.asp?2010/3/3/139/74488> <https://doi.org/10.4103/0974-2077.74488> PMID: 21430825
 8. Chen C-P, Liu F-F, Jan S-W, Lee C-C, Town D-D, Lan C-C. Cytogenetic evaluation of cystic hygroma associated with hydrops fetalis, oligohydramnios or intrauterine fetal death: the roles of amniocentesis, postmortem chorionic villus sampling and cystic hygroma paracentesis. *Acta Obstet Gynecol Scand* [Internet]. 1996 Jan; 75(5):454–8. Available from: <http://doi.wiley.com/10.3109/00016349609033353> PMID: 8677770
 9. Alldred SK, Takwoingi Y, Guo B, Pennant M, Deeks JJ, Neilson JP, et al. First trimester ultrasound tests alone or in combination with first trimester serum tests for Down's syndrome screening. *Cochrane Database Syst Rev* [Internet]. 2017 Mar 15;2017(3). Available from: <http://doi.wiley.com/10.1002/14651858.CD012600>
 10. Moratalla J, Pintoff K, Minekawa R, Lachmann R, Wright D, Nicolaidis KH. Semi-automated system for measurement of nuchal transchency thickness. *Ultrasound Obstet Gynecol*. 2010. <https://doi.org/10.1002/uog.7737> PMID: 20617517
 11. Nie S, Yu J, Chen P, Wang Y, Zhang JQ. Automatic Detection of Standard Sagittal Plane in the First Trimester of Pregnancy Using 3-D Ultrasound Data. *Ultrasound Med Biol*. 2017; <https://doi.org/10.1016/j.ultrasmedbio.2016.08.034> PMID: 27810260
 12. Bertalmio M, Bertozzi AL, Sapiro G. Navier-stokes, fluid dynamics, and image and video inpainting. In: *Proceedings of the 2001 IEEE Computer Society Conference on Computer Vision and Pattern Recognition CVPR 2001* [Internet]. IEEE Comput. Soc; p. I-355–I-362. Available from: <http://ieeexplore.ieee.org/document/990497/>
 13. Ziller A, Usynin D, Braren R, Makowski M, Rueckert D, Kaissis G. Medical imaging deep learning with differential privacy. *Sci Rep* [Internet]. 2021 Dec 29; 11(1):13524. Available from: <http://www.nature.com/articles/s41598-021-93030-0> <https://doi.org/10.1038/s41598-021-93030-0> PMID: 34188157
 14. Huang G, Liu Z, Van Der Maaten L, Weinberger KQ. Densely Connected Convolutional Networks. In: *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)* [Internet]. IEEE; 2017. p. 2261–9. Available from: <https://ieeexplore.ieee.org/document/8099726/>
 15. PyTorch DenseNet [Internet]. [cited 2021 Oct 28]. Available from: https://pytorch.org/hub/pytorch_vision_densenet/
 16. CrossEntropyLoss—PyTorch 1.10.0 documentation [Internet]. [cited 2021 Oct 28]. Available from: <https://pytorch.org/docs/stable/generated/torch.nn.CrossEntropyLoss.html>
 17. Kingma DP, Ba J. Adam: A Method for Stochastic Optimization. 2014 Dec 22; Available from: <http://arxiv.org/abs/1412.6980>
 18. Adam—PyTorch 1.10.0 documentation. [Internet]. [cited 2021 Oct 28]. Available from: <https://pytorch.org/docs/stable/generated/torch.optim.Adam.html>
 19. StepLR—PyTorch 1.10.0 documentation [Internet]. [cited 2021 Oct 28]. Available from: https://pytorch.org/docs/stable/generated/torch.optim.lr_scheduler.StepLR.html
 20. Selvaraju RR, Cogswell M, Das A, Vedantam R, Parikh D, Batra D. Grad-CAM: Visual Explanations from Deep Networks via Gradient-Based Localization. *Int J Comput Vis* [Internet]. 2020 Feb 11; 128(2):336–59. Available from: <http://link.springer.com/10.1007/s11263-019-01228-7>
 21. Palatnik de Sousa I, Vellasco MMBR, Costa da Silva. Explainable Artificial Intelligence for Bias Detection in COVID CT-Scan Classifiers. *Sensors* [Internet]. 2021 Aug 23; 21(16):5657. Available from: <https://www.mdpi.com/1424-8220/21/16/5657> <https://doi.org/10.3390/s21165657> PMID: 34451100
 22. Hryniewska W, Bombiński P, Szatkowski P, Tomaszewska P, Przelaskowski A, Biecek P. Checklist for responsible deep learning modeling of medical images based on COVID-19 detection studies. *Pattern Recognit* [Internet]. 2021 Oct; 118:108035. Available from: <https://linkinghub.elsevier.com/retrieve/pii/S0031320321002223> <https://doi.org/10.1016/j.patcog.2021.108035> PMID: 34054148
 23. Burgos-Artizzu XP, Coronado-Gutiérrez D, Valenzuela-Alcaraz B, Bonet-Carne E, Eixarch E, Crispi F, et al. Evaluation of deep convolutional neural networks for automatic classification of common maternal fetal ultrasound planes. *Sci Rep* [Internet]. 2020 Dec 23; 10(1):10200. Available from: <http://www.nature.com/articles/s41598-020-67076-5> <https://doi.org/10.1038/s41598-020-67076-5> PMID: 32576905
 24. Ryou H, Yaqub M, Cavallaro A, Roseman F, Papageorghiou A, Noble JA. Automated 3D Ultrasound Biometry Planes Extraction for First Trimester Fetal Assessment. In 2016. p. 196–204. Available from: http://link.springer.com/10.1007/978-3-319-47157-0_24

25. Baumgartner CF, Kamnitsas K, Matthew J, Fletcher TP, Smith S, Koch LM, et al. SonoNet: Real-Time Detection and Localisation of Fetal Standard Scan Planes in Freehand Ultrasound. *IEEE Trans Med Imaging* [Internet]. 2017 Nov; 36(11):2204–15. Available from: <https://ieeexplore.ieee.org/document/7974824/> <https://doi.org/10.1109/TMI.2017.2712367> PMID: 28708546
26. Gofer S, Haik O, Bardin R, Gilboa Y, Perlman S. Machine Learning Algorithms for Classification of First-Trimester Fetal Brain Ultrasound Images. *J Ultrasound Med* [Internet]. 2021 Oct 28; Available from: <https://onlinelibrary.wiley.com/doi/10.1002/jum.15860> PMID: 34710247
27. Sciortino G, Tegolo D, Valenti C. Automatic detection and measurement of nuchal translucency. *Comput Biol Med* [Internet]. 2017 Mar; 82:12–20. Available from: <https://linkinghub.elsevier.com/retrieve/pii/S0010482517300070> <https://doi.org/10.1016/j.combiomed.2017.01.008> PMID: 28126630
28. Xie HN, Wang N, He M, Zhang LH, Cai HM, Xian JB, et al. Using deep-learning algorithms to classify fetal brain ultrasound images as normal or abnormal. *Ultrasound Obstet Gynecol* [Internet]. 2020 Oct; 56(4):579–87. Available from: <https://onlinelibrary.wiley.com/doi/10.1002/uog.21967>
29. Arnaut 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* [Internet]. 2021 May 14; 27(5):882–91. Available from: <http://www.nature.com/articles/s41591-021-01342-5> <https://doi.org/10.1038/s41591-021-01342-5> PMID: 33990806
30. Prieto JC, Shah H, Rosenbaum A, Jiang X, Musonda P, Price J, et al. An automated framework for image classification and segmentation of fetal ultrasound images for gestational age estimation. In: Landman BA, Išgum I, editors. *Medical Imaging 2021: Image Processing* [Internet]. SPIE; 2021. p. 55. Available from: <https://www.spiedigitallibrary.org/conference-proceedings-of-spie/11596/2582243/An-automated-framework-for-image-classification-and-segmentation-of-fetal/10.1117/12.2582243.full> PMID: 33935344
31. Zhang Y, Yang Y. Cross-validation for selecting a model selection procedure. *J Econom* [Internet]. 2015 Jul; 187(1):95–112. Available from: <https://linkinghub.elsevier.com/retrieve/pii/S0304407615000305>