Fetal Diagnosis and Therapy

Research Article

Fetal Diagn Ther 2023;50:480–490
DOI: 10.1159/000533203

Received: April 12, 2023
Accepted: July 11, 2023
Published online: August 11, 2023

Automatic Deep Learning-Based Pipeline for Automatic Delineation and Measurement of Fetal Brain Structures in Routine Mid-Trimester Ultrasound Images

David Coronado-Gutiérrez[a, b] Elisenda Eixarch[a, c, d] Elena Monterde[a]
Isabel Matas[a] Paola Traversi[a] Eduard Gratacós[a, c, d]
Elisenda Bonet-Carne[a, c, e] Xavier P. Burgos-Artizzu[a]

[a]BCNatal | Fetal Medicine Research Center (Hospital Clínic and Hospital Sant Joan de Déu, University of Barcelona), Barcelona, Spain; [b]Transmural Biotech S. L., Barcelona, Spain; [c]Institut d'Investigacions Biomèdiques August Pi I Sunyer (IDIBAPS), Barcelona, Spain; [d]Centre for Biomedical Research on Rare Diseases (CIBERER), Barcelona, Spain; [e]Barcelona Tech, Universitat Politècnica de Catalunya, Barcelona, Spain

Mini-Summary

What does this study add to current knowledge?

- This study proposes a novel pipeline to automatically delineate fetal head and brain structures and obtain automatic measures of each anatomical standard plane acquired during routine fetal ultrasound (US) examination. The results of this study also support further research to develop automated artificial intelligence (AI) methods to improve the visualization, delineation, and classification tasks in maternal-fetal US imaging.

What are the main clinical implications?

- This study demonstrates that the application of AI could improve clinical practice in fetal US. If the proposed techniques were implemented in software integrated in the US machine, they would help in the visualization of important brain structures in real time during the US examination. This application could also measure automatically other relevant parameters standardizing them, thus avoiding operator bias which depends on their experience and other factors.

Keywords
Deep learning · Fetal ultrasound · Fetal brain · Fetal biometry · Brain structures

Abstract
Introduction: The aim of this study was to develop a pipeline using state-of-the-art deep learning methods to automatically delineate and measure several of the most important brain structures in fetal brain ultrasound (US) images. ***Methods:*** The dataset was composed of 5,331 images of the

karger@karger.com
www.karger.com/fdt

Correspondence to:
Elisenda Bonet-Carne, elisenda.bonet@sjd.es

fetal brain acquired during the routine mid-trimester US scan. Our proposed pipeline automatically performs the following three steps: brain plane classification (transventricular, transthalamic, or transcerebellar plane); brain structures delineation (9 different structures); and automatic measurement (from the structure delineations). The methods were trained on a subset of 4,331 images and each step was evaluated on the remaining 1,000 images. ***Results:*** Plane classification reached 98.6% average class accuracy. Brain structure delineation obtained an average pixel accuracy higher than 96% and a Jaccard index higher than 70%. Automatic measurements get an absolute error below 3.5% for the four standard head biometries (head circumference, biparietal diameter, occipitofrontal diameter, and cephalic index), 9% for transcerebellar diameter, 12% for *cavum septi pellucidi* ratio, and 26% for Sylvian fissure operculization degree. ***Conclusions:*** The proposed pipeline shows the potential of deep learning methods to delineate fetal head and brain structures and obtain automatic measures of each anatomical standard plane acquired during routine fetal US examination.

Introduction

Ultrasound (US) is routinely used to assess fetal development during pregnancy, due to its non-invasive and radiation-free nature, low cost, and real-time visualization capabilities [1]. Fetal anatomy and fetal growth evaluation by means of fetal biometries are included in the mid-trimester scan protocol [2]. Within anatomical evaluation, the identification and measurement of specific fetal brain structures is also an essential part of basic prenatal screening for malformations or abnormalities [3, 4]. Evaluating these structures requires clinical training and could be hindered due to many conditions like the US machine and probe characteristics, increased maternal fat tissue, poor sonographer experience, or fetus orientation [3]. Therefore, it would be extremely beneficial to develop automatic methods for detection and measurement of these brain structures to help the sonographer during examination while reducing inter-observer variability.

The search of automatic delineations and measures for fetal US has attracted the researchers' attention for years. Carneiro et al. [5] were one of the first to automatically measure the main biometries using a probabilistic boosting tree classifier. More recently, several publications have used more sophisticated methods to solve the same task [6]. However, the automatic measurement of fetal brain structures is still an open research area. Yaqub et al. [7, 8] and Sofka et al. [9] were part of the first to automatically delineate anatomical structures but using 3D-US images, whose implementation in clinical practice is complicated since 2D-US are still the most widely used modality.

Most recently, some publications have proposed deep learning methods to automatically delineate different fetal brain structures in 2D-US. Chen et al. [10] presented a technique to delineate and measure lateral ventricles, and Singh et al. [11] proposed different methods to delineate the cerebellum. Lately, Lin et al. [12] were able to identify 9 intracranial malformations in different brain structures in standard sonographic reference planes. Still, none of these publications claimed to delineate several brain structures at the same time and make measurements on them.

The goal of this study was to develop a novel pipeline using state-of-the-art deep learning methods to automatically delineate and measure several of the most important fetal brain structures assessed during mid-trimester routine fetal brain US examination (Fig. 1).

Materials and Methods

Study Design

Patients and Images

All images of this study were taken from a dataset composed of 12,400 US images obtained in our center and publicly available from Zenodo [13, 14]. This dataset is composed of US fetal images acquired from pregnant women who attended routine pregnancy screening during second and third trimester (18–40 weeks of gestation). Images were acquired at BCNatal (Hospital Clínic de Barcelona and Hospital Sant Joan de Déu, Barcelona, Spain) during standard clinical practice between January 2018 and April 2019 by 6 different operators. Multiple pregnancies, congenital malformations, or aneuploidies were excluded. Images were acquired with different equipments including Voluson E6 (GE Medical Systems, Zipf, Austria), Voluson S8, Voluson S10 (GE Medical Systems, Zipf, Austria), and Aloka (Aloka Co., Ltd.) with convex transducers (3–7.5 MHz). Operators were instructed to avoid using any type of post-processing or artifacts (noise, calipers, pointers, etc.), leaving the remaining image settings parameters such as gain, frequency, and gain compensation to their discretion. All images of the dataset were manually classified by clinicians into their plane category. The study was approved by the coordinator's Institutional Review Board (HCB 2018/0031), and all patients provided written informed consent to use US images for research purposes. For more information about image acquisition please refer to Burgos-Artizzu et al. [13].

From the original dataset, we selected all the axial brain images from the mid-trimester scan (18–24 weeks of gestation) [4]. All these brain images were also classified in one of the three standard anatomical planes, according to ISUOG guidelines [15]:

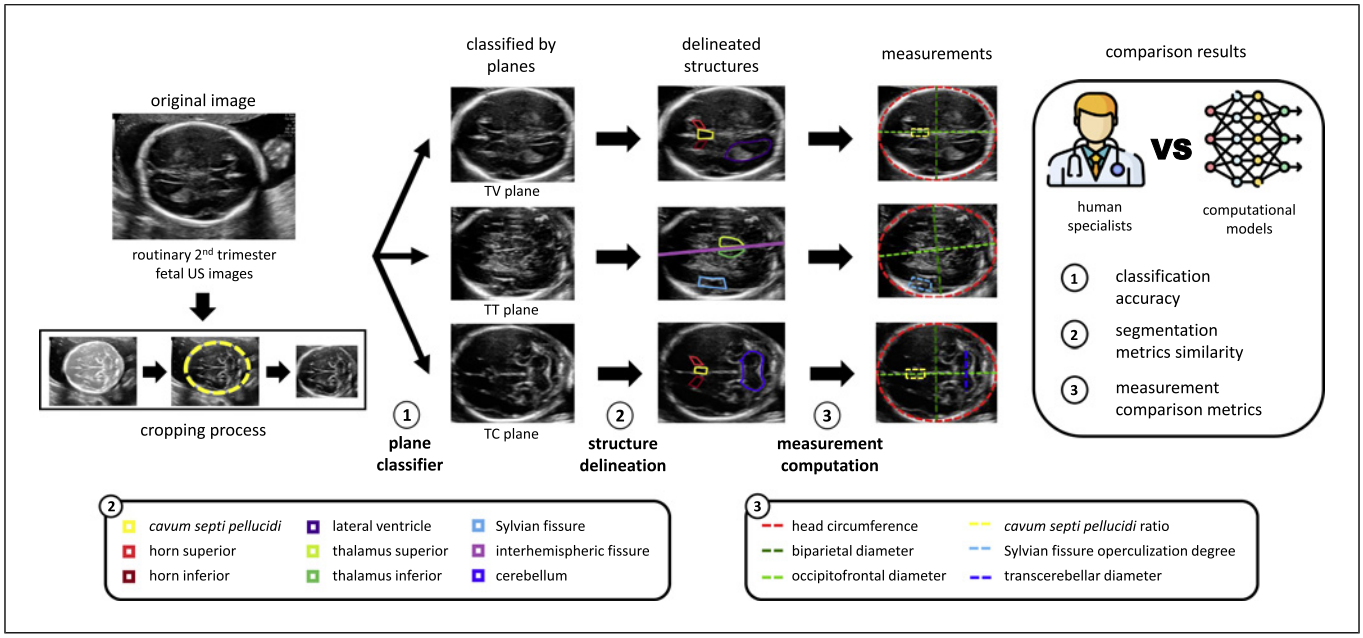

Fig. 1. Full pipeline of the study composed of the three main steps (brain plane classification, structure delineation, and measurement computation), the previous cropping process, and the comparison results. In the bottom the corresponding legend of structures of step 2 and measurements of step 3. US, ultrasound; TV, transventricular; TT, transthalamic; TC, transcerebellar.

transventricular (TV), transthalamic (TT), and transcerebellar (TC) (Fig. 2). The final dataset used in this study was therefore composed of 5,331 images from 3,554 patients: 2,253 TV planes, 1,456 TT planes, and 1,622 TC planes.

Data Preparation: Manual Delineations and Measurements

In order to train and validate the deep learning models used, we previously created the ground-truth for each image of the dataset. Three different fetal medicine specialists manually delineated and measured the proposed structures and measurements. To do this task, the clinicians used an online graphical user interface created by Transmural Biotech (Barcelona, Spain) [16]. This graphical user interface allows to delineate structures point-to-point, rectangularly or with ellipsoids, and allows measuring distances inside the US image.

First, two clinicians delineated the selected structures of each plane: *cavum septi pellucidi*, anterior horn superior, anterior horn inferior, inferior lateral ventricle, thalamus superior, thalamus inferior, Sylvian fissure (SF), interhemispheric fissure, and cerebellum (Fig. 1, "delineated structures"). All these structures were delineated point-to-point, leaving the definition of the delineation (number of points) to the discretion of the clinicians.

Second, a third clinician measured head biometries (the same way they would do on the US machine) including head circumference (HC), biparietal diameter (BPD), occipitofrontal diameter (OFD), and cephalic index (CI) (ratio between BPD and OFD). In addition, other measurements to identify brain anomalies were also measured including transcerebellar diameter (TCD) [17], CSP ratio (ratio between length and width) [18], and SF operculization degree (ratio between SF width and uncovered insula width) [19] (Fig. 1, "measurements"). All these measurements were calculated with a point-to-point line, except HC which was measured with an ellipsoid.

Finally, the images were randomly separated into train and test subsets, preventing by putting images of the same patient in both sets. One-thousand images (500 TV, 200 TT, and 300 TC) from 803 patients were selected to use as a test set and the other 4,331 images (2,751 patients) were used to train and validate the different methods detailed below (10% of the training images were used as internal validation split).

Pipeline Development

The proposed pipeline (Fig. 1) consists of three steps: (1) automatic plane classification; (2) automatic structure delineation; and (3) automatic measurements. Each one of these steps is explained in more detail below. The entire pipeline was implemented in MATLAB® R2017a (MathWorks, Inc., Natick, MA, USA).

Automatic Plane Classification

The first step is to classify the brain image-specific plane into the three possible classes: TV, TT, and TC (Fig. 1, step 1). Previously, to discard irrelevant information and help the following classification methods to have better performance, each image was cropped over the brain perimeter limits (Fig. 1, "cropping process"). This cropping was done automatically, using the method proposed in a previous study of our group [20]. This method uses a deep segmentation convolutional neural network (CNN) to

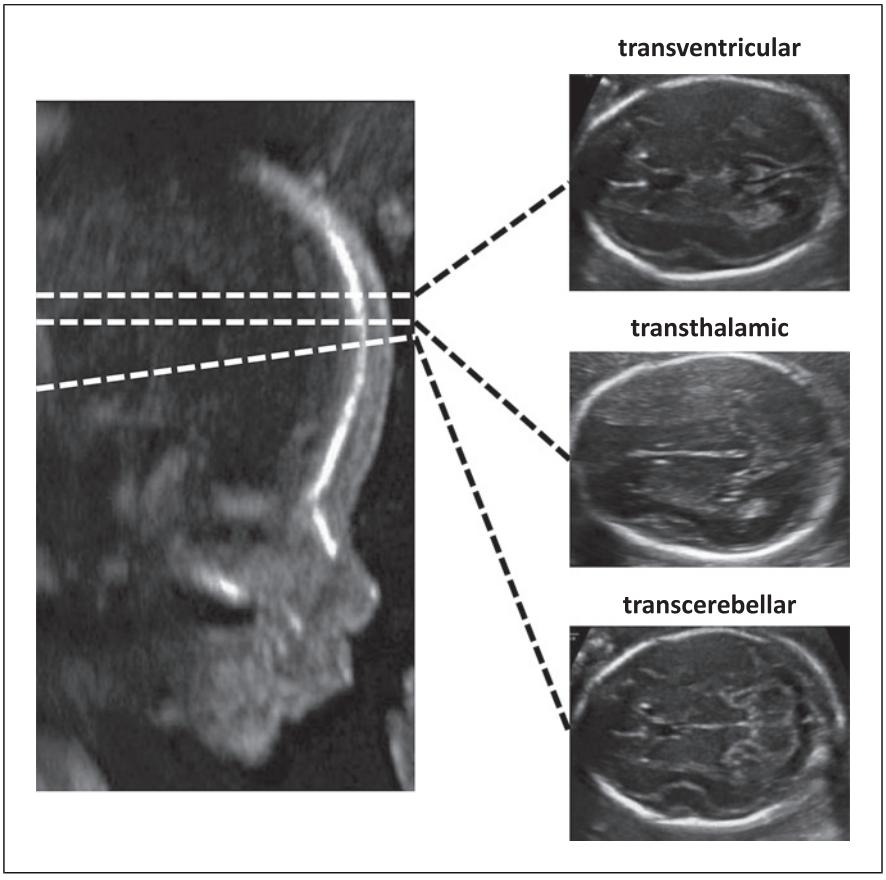

Fig. 2. Examples of the three different brain planes (TV, TT, and TC).

estimate the brain region of interest (ROI). Once the ROI is detected, we estimated the ellipse which creates less error with respect to all their points, and then, the image was rectangularly cropped around the ellipse.

Subsequently, to perform the classification task, a ResNet-18 (Residual Neural Network) [21] was trained with the images and the labels (planes) of the training subset and later applied to the testing subset. ResNets are CNNs pre-trained with ImageNet dataset [22]. For our purpose, we trained the network during 50 complete passes through the dataset (epochs) with batches of 32 samples (batch size), and we used a softmax loss function [23], Adam optimization algorithm [24], and we also applied batch normalization [25].

Automatic Fetal Brain Structure Delineation

To automatically delineate the brain structures (Fig. 1, step 2), we developed an automatic brain structure delineation algorithm. The backbone of the approach is the DeepLab [26] CNN which uses an ImageNet [22] pre-trained VGG [27] as its main feature extractor network. DeepLab proposes a new residual block for multi-scale feature learning which will be useful to delineate the proposed brain structures. In our case, we used two different DeepLab CNNs: one was trained only with TT images and the other with TV and TC images together (as both had 3 same structures: CSP and anterior horns). We also used data augmentation to increase the amount of data to train the model and get better results, modifying the contrast, brightness, and saturation of the original images and mirroring the images horizontally. Both CNNs were trained during 600 complete passes through the dataset (epochs) with batches of 64 samples (batch size) and we used a softmax loss function [23] and gradient descent optimization algorithm [28].

Automatic Measurements

Once all the structures were delineated by the various deep learning CNNs, we developed different equations and methods to estimate from the ROIs the proposed biometric parameters and measures automatically. To measure the 4 head biometries, we used the estimated ellipse obtained from the ROI of the brain perimeter (Fig. 1, "cropping process"). We obtained the BPD and OFD measurements by computing the double of the short axis and the double of the long axis parameters of the ellipse, respectively [29]. Then, we can calculate the HC and the CI with the following equations:

$$HC \approx 1.62 \cdot (BPD - OFD) \quad (1)$$

$$CI = \frac{BPD}{OFD} \quad (2)$$

To obtain the remaining three measures, we used the three ROIs estimated by the automatic structure segmenter CNNs: CSP from TV and TC planes, SF from TT planes, and cerebellum from TC planes. First, we estimated the bound rectangle which better fits each ROI, computing the convex hull of all ROI

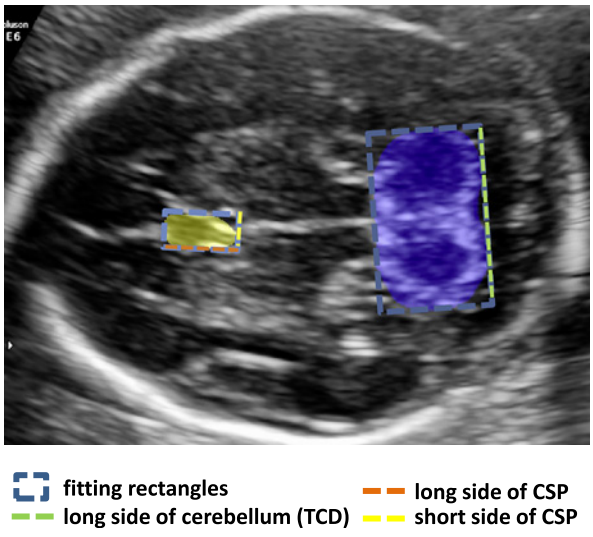

Fig. 3. Example of an automatic region of interest (ROI) of *cavum septi pellucidi* (yellow region) and cerebellum (blue region), the fitting rectangles of each one (dashed rectangles) and the proposed measurements (in legend).

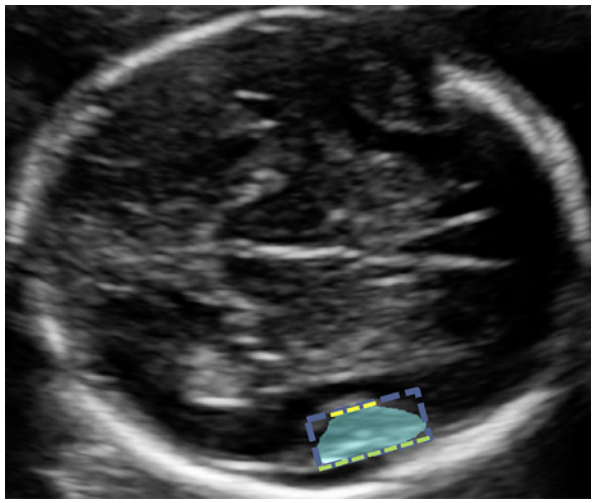

Fig. 4. Example of an automatic region of interest (ROI) of SF (cian region), the fitting rectangle (dashed rectangle), and the proposed measurements (in legend).

points and finding the minimal area for the rectangle. Once we had the bound rectangle, the CSP shape could resemble that rectangle. Then, to obtain the CSP ratio, we computed the relation between the short side and the long side of the rectangle. In the case of the cerebellum, when we had their fitting rectangle, we assigned the long side of the rectangle to the TCD. Figure 3 shows an example of the fitting rectangles of CSP and cerebellum delineations with the corresponding measurements.

Finally, to measure the SF operculization degree, we assigned the bottom side of the SF's fitting rectangle to the uncovered insula width, and for the SF width, we assigned only the part of the upper side of the rectangle which is in contact with the obtained ROI (Fig. 4). Then, to obtain the SF operculization degree, we computed the relation between SF width and uncovered insula width.

Statistical Analysis

We evaluated the performance of the automatic pipeline by comparing results of each step with the manual ground-truth on the test set images. Statistical analysis was performed using MATLAB R2017a.

To evaluate the plane classification, we computed its global accuracy with respect to the labels produced by clinicians in our previous study [13]. Then, full confusion matrices were also computed, and the mean and standard deviation of the diagonal of the confusion matrix was reported to analyze average accuracy on each class.

To evaluate the structure delineation, we computed segmentation metrics to compare the obtained automated ROIs and the manual ROIs made by our clinicians. The metrics used are global pixel accuracy (percent of total pixels correctly classified), Jaccard index [30] (intersection over union of total pixels), class accuracy (average accuracy obtained for all ROIs without taking into account the number of pixels), and failures (images with less than 50% of Jaccard index). Pixel accuracy and Jaccard index was also computed for each structure individually to see the performance on each one.

Finally, to evaluate the measurements, we computed basic concordance statistical measures between the obtained automatic measurements and the manual measurements made by our clinician. The measures are as follow: average manual measurement, average automatic measurement, mean of the absolute difference between manual and automatic measurements, root mean square error (RMSE), and Pearson correlation. Scatterplots of manual and automatic measurements were also computed to observe the differences.

Results

Automatic Plane Classification

Figure 5 shows the confusion matrix between the three brain planes manually classified by the clinicians and the same planes automatically classified by the CNN on the test images set. The global accuracy achieved by the model was 98.1% and the top-1 error was 1.2%.

Automatic Fetal Brain Structure Delineation

Table 1 shows the performance on the test images set of the automatic structure delineations created by the two proposed CNN models (TT and TV/TC planes) with respect to the manual delineations made by clinicians.

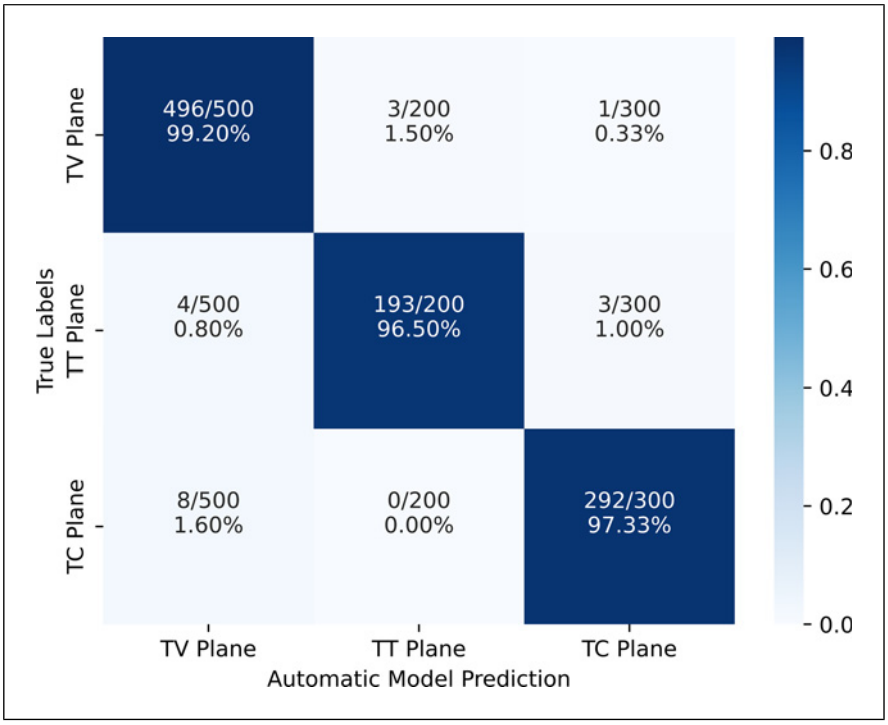

Fig. 5. Confusion matrix obtained on the test subset images to classify the brain images in planes. Matrix rows show the true class labeled by our expert maternal-fetal clinician and columns are the prediction from our automatic model. TV, transventricular; TT, transthalamic; TC, transcerebellar.

Table 1. Segmentation metrics to compare manual and automatic delineations on the two different CNN models (TT and TV/TC plane) and also for each proposed brain structure

	Pixel accuracy, %	Jaccard index, %	Class accuracy, %	Failures (<50% Jacc.), %
TT plane	**95.9**	**72.8**	**88.8**	**5.0**
Thalamus superior	89.4	75.4		
Thalamus inferior	91.3	75.1		
SF	84.7	65.1		
Interhemispheric fissure	81.8	52.9		
TV/TC planes	**97.6**	**70.3**	**85.1**	**4.9**
Cavum septi pellucidi	86.5	65.2		
Anterior horn superior	63.3	44.4		
Anterior horn inferior	79.8	54.7		
Ventricle inferior (TV)	87.2	76.3		
Cerebellum (TC)	95.4	83.7		

TT, transthalamic; TV, transventricular; TC, transcerebellar; SF, Sylvian fissure.

Results are shown for each CNN and also for each proposed brain structure, using the described delineation metrics. For more information about the results of Table 1, online supplemental Table 1, and online supplemental Figure 1 show the results with incremental train images sets and online supplemental Table 2 shows the results divided by US machines (for all online suppl. material, see https://doi.org/10.1159/000533203).

Figure 6 shows six examples of comparisons between manual and automatic delineations, highlighting each structure in a different color. The first row shows one of the best automated delineations for each plane in terms of Jaccard index and the second row shows one of the worst for each one.

Automatic Measurements

Table 2 shows the difference between measurements obtained by clinicians and the same measurements obtained by our automatic pipeline on the test images set. To observe the differences, the table shows average values of the automatic and manual measurements, mean difference between them, RMSE, and Pearson

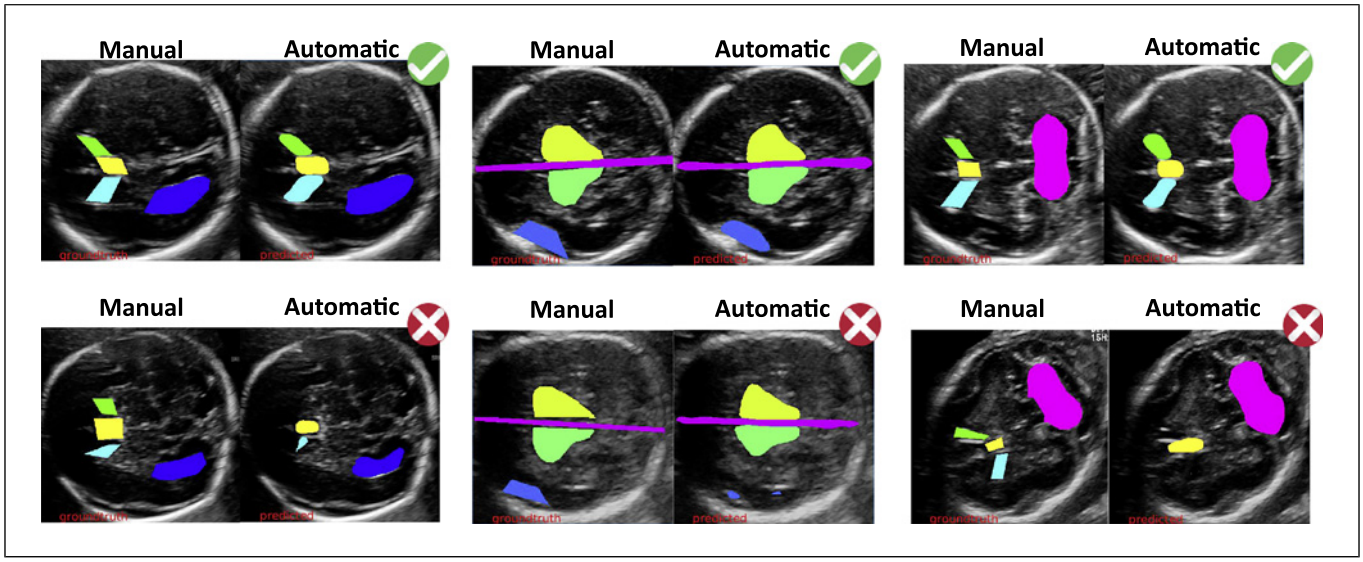

Fig. 6. Six delineation examples: the first line shows three properly automatic delineated examples of each plane and the second line shows three of the worst automatic delineated examples of each plane. Each brain structure is highlighted in a different color.

correlation values. For more detail, the results are shown both for the entire test set and for 3 different gestational age intervals.

Figures 7 and 8 show the scatterplots of manual and automatic measures of Table 2 for the entire test set. Figure 7 shows the four head biometries and Figure 8 shows the other three defined measurements. For more detail, the estimated regression line and the corresponding Pearson correlation value are also shown.

Discussion

In this study, we have proposed a complete pipeline to automatically measure the main fetal brain structures relevant to mid-trimester routine US fetal brain examination. The novel pipeline includes several different deep learning neural networks and mathematical methods. Results showed that the proposed pipeline can perform all the needed tasks, validating the main goal of this study.

First, we developed a CNN to automatically classify the brain images into the three standard planes (TV, TT, and TC). The proposed technique reached 98.1% average accuracy, far superior than the 74.0% presented in our previous study [13]. These results are also better than the results obtained manually by different clinicians of the same previous study which was 81.7% and 83.2% average class accuracy, thus demonstrating the potential of this automatic technique.

Second, we developed two CNNs to automatically delineate several brain structures and compared results with that of maternal-fetal specialists. The pixel accuracy and Jaccard index obtained in each structure individually showed that these CNN were capable of properly delineating the 9 proposed fetal brain structures. Still, some of them did not get optimal results (less than 60% Jaccard for anterior horns and interhemispheric fissure) due to having less manual delineations available to train the networks. In the future, this could be improved by increasing the training images. In general, the results obtained by our pipeline on several structures at the same time are in line with previous published articles which aimed on detecting only one specific structure. Delineation of lateral ventricles obtained 96% accuracy [10] compared with our 87.2%, and cerebellum delineation obtained an 84.6% Jaccard index [11] while we got 83.7%.

Third, we developed different mathematical methods to automatically measure four head biometries and three specific brain structures. On one hand, the obtained results showed that the proposed methods are useful to automatically measure all fetal head biometries (HC, BPD, OFC, and CI). The comparison between all these parameters measured by clinicians and those obtained with automatic methods showed a difference lower than 3.5% of the absolute measurement values (RMSE divided by average measure). On the other hand, the differences between manual and automatic methods in the rest of measurements were larger with an RMSE of 9%, 12%, and 26% of the total measurement (average) for TCD, CSP ratio SF operculization

Table 2. Metrics to compare measurements obtained by clinicians and the same measurements obtained by our automatic pipeline

Measures	Manual average	Automatic average	Mean difference	RMSE	Pearson correlation
HC, mm	**185.57±15.85**	**185.40±15.75**	**3.47±2.90**	**4.52**	**0.96**
18–20 weeks	175.40±7.04	176.51±7.65	3.03±2.82	4.14	0.85
21–22 weeks	187.51±7.86	188.17±8.17	3.34±3.00	4.48	0.85
23–24 weeks	216.95±9.08	218.61±8.50	4.12±3.49	5.38	0.83
BPD, mm	**50.04±4.56**	**50.27±4.64**	**1.03±0.89**	**1.36**	**0.96**
18–20 weeks	47.50±2.42	47.87±2.51	0.87±0.78	1.17	0.90
21–22 weeks	50.69±2.51	51.12±2.52	1.02±0.77	1.28	0.88
23–24 weeks	58.55±2.78	59.60±2.45	1.26±1.16	1.70	0.87
OFD, mm	**64.52±5.59**	**64.18±5.46**	**1.60±1.35**	**2.09**	**0.93**
18–20 weeks	60.77±2.58	61.08±2.87	1.38±1.26	1.86	0.78
21–22 weeks	65.05±2.99	65.03±3.23	1.64±1.39	2.15	0.76
23–24 weeks	75.37±3.63	75.35±3.64	2.04±1.70	2.64	0.73
Cephalic index, %	**77.60±3.77**	**78.37±3.78**	**2.18±1.73**	**2.78**	**0.75**
18–20 weeks	78.21±3.66	78.44±3.78	1.93±1.51	2.45	0.78
21–22 weeks	77.99±3.53	78.70±3.80	2.22±1.78	2.84	0.72
23–24 weeks	77.35±3.68	79.20±3.69	2.75±1.97	3.37	0.65
TCD, mm	**20.46±2.43**	**22.34±2.37**	**2.14±1.24**	**2.40**	**0.78**
18–20 weeks	18.76±0.83	21.05±0.97	2.29±0.84	2.44	0.55
21–22 weeks	20.50±1.18	22.14±1.50	1.91±0.96	2.13	0.49
23–24 weeks	23.90±1.47	25.97±1.55	2.07±1.49	2.40	0.52
CSP ratio, %	**41.03±8.53**	**46.97±5.44**	**8.18±5.80**	**10.0**	**0.40**
18–20 weeks	38.82±8.30	47.65±5.73	9.49±6.44	11.45	0.33
21–22 weeks	41.97±8.1	46.48±5.34	6.96±4.97	8.54	0.47
23–24 weeks	44.51±8.06	42.98±3.22	7.96±3.99	8.77	0.22
SF op. degree, %	**43.58±6.25**	**48.76±11.56**	**9.76±8.35**	**12.82**	**0.24**
18–20 weeks	43.33±6.39	47.47±10.4	8.84±6.59	11.00	0.33
21–22 weeks	43.09±5.39	46.71±10.58	8.61±6.00	10.45	0.37
23–24 weeks	44.17±6.16	51.51±8.56	13.34±8.91	16.18	0.16

Results are shown globally for each measurement and also for different gestational ages. HC, head circumference; BPD, biparietal diameter; OFD, occipitofrontal diameter; TCD, transcerebellar diameter; CSP, Cavum septi pellucidi; SF op., Sylvian fissure operculization; RMSE, root mean square error.

degree, respectively. These errors are due to imprecision of the delineations in edges of the structures. To improve this, the models should be trained with more images that would allow a more precise delineation of the structures. Our results on head biometries (BPD, HC, and OFD) and TCD measurement are comparable to those obtained using 3D-US volumes [9]: they had an error of 0.94, 4.06, 2.31, and 1.37 mm, respectively (mean difference) while we got 1.03, 3.47, 1.60, and 2.14 error millimeters.

From a clinical standpoint, this study supports the idea that application of AI could improve clinical practice in fetal US. If the proposed techniques were implemented in software integrated in the US machine, they would help in the visualization of important brain structures in real time during the US examination. This application could also measure automatically other relevant parameters standardizing them thus avoiding operator bias which depends on their experience and other factors.

This study has several strengths and limitations. Among strengths, the images used were real images of the clinical practice not acquired for research purposes. All images were collected by different operators during routine clinical practice, with different machines from different brands and no specific instructions except to avoid post-processing or artifacts when possible [13]. Therefore, we expect the system here described to be fairly robust and reproducible with respect to the acquisition environment (the results of online suppl. Table 2 also support this idea). Second, the whole proposed system is totally automated in each of their parts. This would allow integration of software with these techniques in an US machine where the operator would only have to acquire the US image. Third, while the focus of this study is to present the pipeline as a whole, each of its parts can be used independently and could serve as backbone to many other automated fetal brain US examination applications.

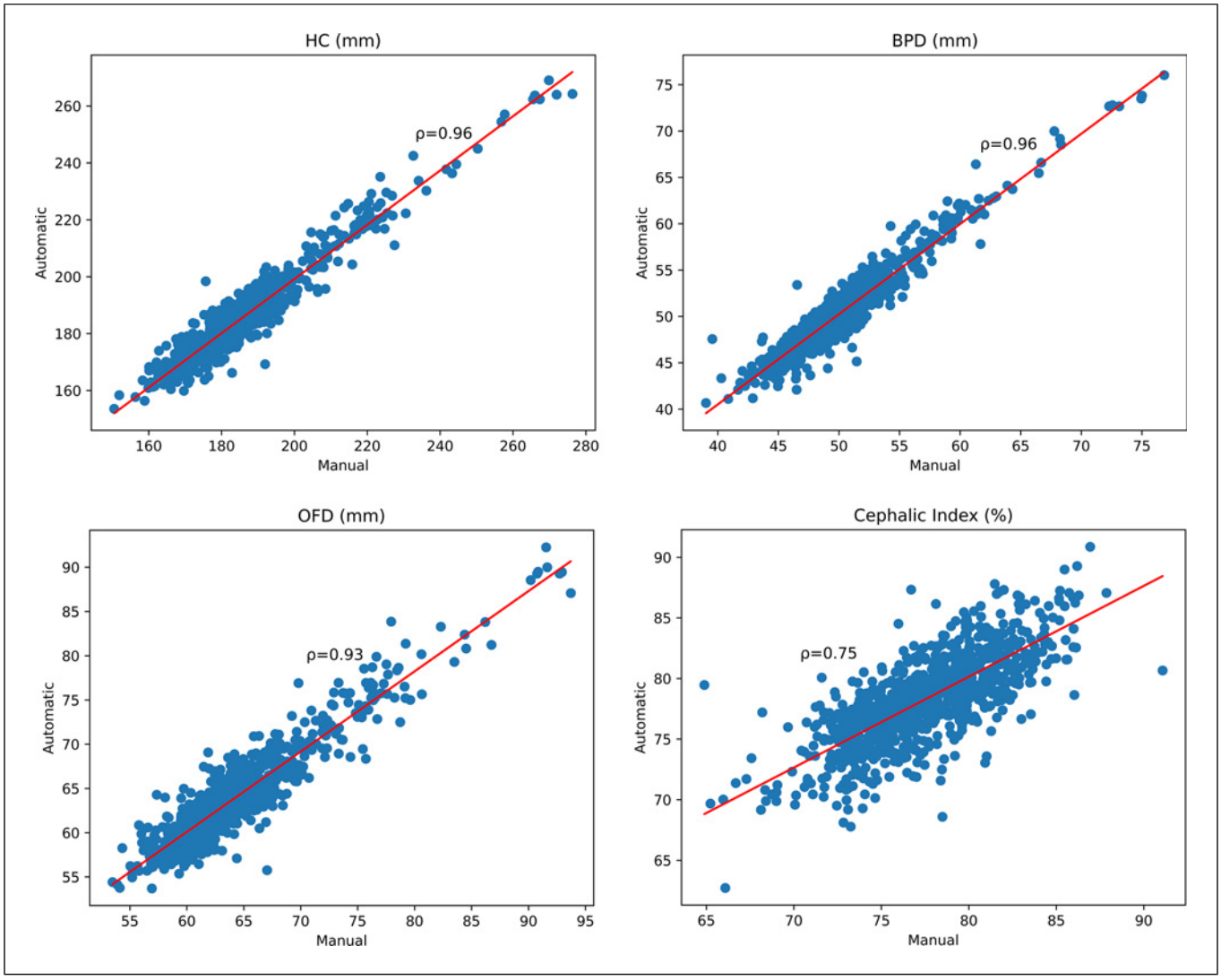

Fig. 7. Scatterplots between manual and automatic measures of the four head biometries obtained on the test set. From left to right and up to down: head circumference (HC), biparietal diameter (BPD), occipitofrontal diameter (OFD), and cephalic index (CI). Red line is the estimated regression line and the ρ value is the corresponding Pearson correlation value.

As a main limitation of the study, some standard structures and measurements which were analyzed in routine mid-trimester US were not included. Clinically relevant measurements as lateral ventricle diameter or cisterna magna diameter were not covered in this study due to the difficulty to automatically measure a segment between two points without a clear reference structure. Further improvements in the study could be to delineate more relevant structures, as well as detect singular points which could allow us to later make measurements such as the lateral ventricle diameter or cisterna magna diameter. Another limitation of the study was the sample size of ground-truth images. We speculate that if we had a higher number of images with manual delineation, the automatic delineations, and therefore the automatic measurements, would have been more precise (the results of online suppl. Table 1 and online suppl. Fig. 1 also support this idea).

In conclusion, we have developed a fully automated AI pipeline to delineate fetal head and brain structures and obtain automatic measures of each anatomical standard plane acquired during routine fetal US examination. We proposed novel and efficient deep learning techniques and mathematical methods to solve the different tasks that

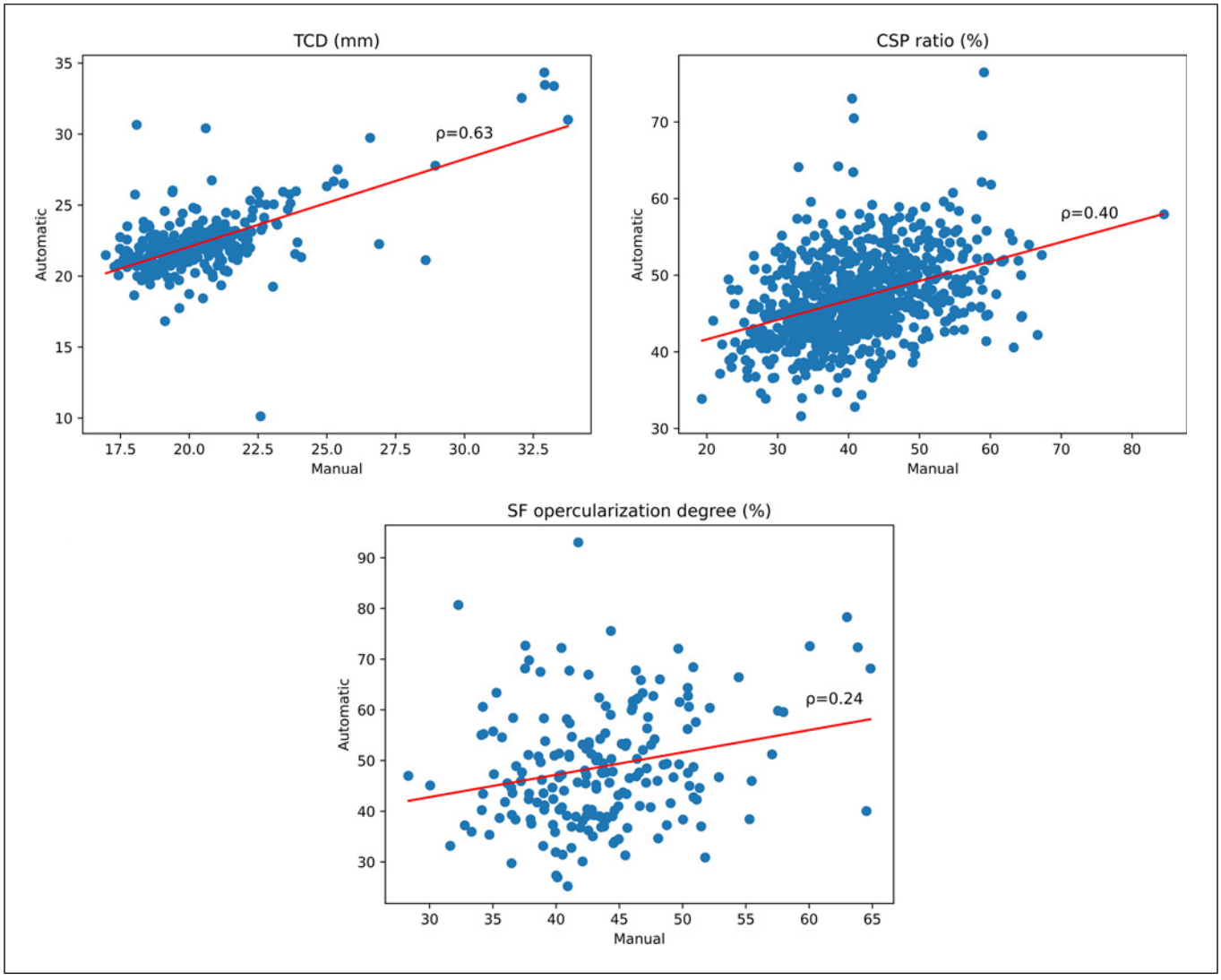

Fig. 8. Scatterplots between manual and automatic measures of brain structure measurements obtained on the test set. From left to right and up to down: transcerebellar diameter (TCD), *cavum septi pellucidi* (CSP) ratio, and Sylvian fissure (SF) operculization degree. Red line is the estimated regression line, and the ρ value is the corresponding Pearson correlation value.

compose our system. The results of this study support further research to develop automated AI methods to improve the visualization, delineation, and classification tasks in maternal-fetal US imaging.

Statement of Ethics

This study protocol was reviewed and approved by the coordinator's Institutional Review Board of Hospital Clínic de Barcelona with approval number HCB 2018/0031. All patients provided written informed consent to use US images for research purposes.

Conflict of Interest Statement

D. Coronado-Gutierrez and X.P. Burgos-Artizzu were employed by Transmural Biotech S.L. when the study was developed. The remaining authors report no conflict of interest.

Funding Sources

The research leading to these results has received funding from the Cerebra Foundation for the Brain Injured Child (Carmarthen, Wales, UK) and ASISA foundation.

Author Contributions

D. Coronado wrote the manuscript with support from X.P. Burgos-Artizzu, E. Bonet-Carné, and E. Eixarch, who also conceived and designed the whole study. D. Coronado also collected the data, performed the experiments, and performed the analysis of the study. E. Monterde, I. Matas, and P. Traversi made the measurements and manual delineations of all images in the database. E. Eixarch and E. Gratacós supervised the study and gave clinical support. Also, all authors revised critically the study and gave the final approval of the version to be published.

Data Availability Statement

The data that support the findings of this study are openly available in "Zenodo" at http://doi.org/10.5281/zenodo.3904280. Further inquiries can be directed to the corresponding author.

References

1 Wells PNT, Liang HD. Medical ultrasound: imaging of soft tissue strain and elasticity. J R Soc Interf. 2011 Nov 7;8(64):1521–49.
2 Salomon LJ, Alfirevic Z, Da Silva Costa F, Deter RL, Figueras F, Ghi T, et al. ISUOG Practice Guidelines: ultrasound assessment of fetal biometry and growth. Ultrasound Obstet Gynecol. 2019 Jun 1;53(6):715–23.
3 Paladini D, Malinger G, Monteagudo A, Pilu G, Timor-Tritsch I, Toi A. Sonographic examination of the fetal central nervous system: guidelines for performing the "basic examination" and the "fetal neurosonogram.". Ultrasound Obstet Gynecol. 2007 Jan;29(1):109–16.
4 Salomon LJ, Alfirevic Z, Berghella V, Bilardo CM, Chalouhi GE, Da Silva Costa F, et al. ISUOG Practice Guidelines (updated): performance of the routine mid-trimester fetal ultrasound scan. Ultrasound Obstet Gynecol. 2022 Jun 1;59(6):840–56.
5 Carneiro G, Georgescu B, Good S, Comaniciu D. Detection and measurement of fetal anatomies from ultrasound images using a constrained probabilistic boosting tree. IEEE Trans Med Imaging. 2008 Sep;27(9):1342–55.
6 Rueda S, Fathima S, Knight CL, Yaqub M, Papageorghiou AT, Rahmatullah B, et al. Evaluation and comparison of current fetal ultrasound image segmentation methods for biometric measurements: a grand challenge. IEEE Trans Med Imaging. 2014;33(4):797–813.
7 Yaqub M, Cuingnet R, Napolitano R, Roundhill D, Papageorghiou A, Ardon R, et al. Volumetric segmentation of key fetal brain structures in 3D ultrasound. Lect Notes Comput Sci. 2013;8184. LNCS:25–32.
8 Yaqub M, Napolitano R, Ioannou C, Papageorghiou AT, Noble JA. Automatic detection of local fetal brain structures in ultrasound images. Proc Int Symp Biomed Imaging. 2012;1555–8.
9 Sofka M, Zhang J, Good S, Zhou SK, Comaniciu D. Automatic detection and measurement of structures in fetal head ultrasound volumes using sequential estimation and Integrated Detection Network (IDN). IEEE Trans Med Imaging. 2014;33(5):1054–70.
10 Chen X, He M, Dan T, Wang N, Lin M, Zhang L, et al. Automatic measurements of fetal lateral ventricles in 2D ultrasound images using deep learning. Front Neurol. 2020 Jul 17;11:526.
11 Singh V, Sridar P, Kim J, Nanan R, Poornima N, Priya S, et al. Semantic segmentation of cerebellum in 2D fetal ultrasound brain images using convolutional neural networks. IEEE Access. 2021;9:85864–73.
12 Lin M, He X, Guo H, He M, Zhang L, Xian J, et al. Use of real-time artificial intelligence in detection of abnormal image patterns in standard sonographic reference planes in screening for fetal intracranial malformations. Ultrasound Obstet Gynecol. 2022 Mar 1;59(3):304–16.
13 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. 2020 Dec 1;10(1):10200.
14 FETAL_PLANES_DB. Common maternal-fetal ultrasound images | Zenodo [Internet.] [cited 2023 April 10]. Available from: https://zenodo.org/record/3904280#.YqdniP3P2Um.
15 Malinger G, Paladini D, Haratz KK, Monteagudo A, Pilu GL, Timor-Tritsch IE. ISUOG Practice Guidelines (updated): sonographic examination of the fetal central nervous system. Part 1: performance of screening examination and indications for targeted neurosonography. Ultrasound Obstet Gynecol. 2020 Sep 1;56(3):476–84.
16 Transmural [Internet] [cited 2023 April 10]. Available from: https://gui.transmuralbiotech.com/.
17 Atallah A, Guibaud L, Gaucherand P, Massardier J, des Portes V, Massoud M. Fetal and perinatal outcome associated with small cerebellar diameter based on second- or third-trimester ultrasonography. Prenat Diagn. 2019 Jun 1;39(7):pd.5465–43.
18 Karl K, Esser T, Heling KS, Chaoui R. Cavum septi pellucidi (CSP) ratio: a marker for partial agenesis of the fetal corpus callosum. Ultrasound Obstet Gynecol. 2017 Sep 1;50(3):336–41.
19 Chen X, Li SL, Luo GY, Norwitz ER, Ouyang SY, Wen HX, et al. Ultrasonographic characteristics of cortical sulcus development in the human fetus between 18 and 41 Weeks of gestation. Chin Med J. 2017 Apr 20;130(8):920–8.
20 Burgos-Artizzu XP, Coronado-Gutiérrez D, Valenzuela-Alcaraz B, Vellvé K, Eixarch E, Crispi F, et al. Analysis of maturation features in fetal brain ultrasound via artificial intelligence for the estimation of gestational age. Am J Obstet Gynecol MFM. 2021 Nov 1;3(6):100462.
21 He K, Zhang X, Ren S, Sun J. Deep residual learning for image recognition. Proc IEEE Comput Soc Conf Comput Vis Pattern Recognit. 2016 Dec 9;770–8.
22 ImageNet [Internet] [cited 2023 April 10]. Available from: https://www.image-net.org/about.php.
23 Goodfellow I, Bengio Y, Courville A. 6.2.2.3 Softmax units for multinoulli output distributions. Deep learning. MIT Press; 2016. p. 180–4.
24 Kingma DP, Ba JL. Adam: a method for stochastic optimization. 3rd int conf learn represent ICLR 2015; 2014 Dec 22.
25 Ioffe S, Szegedy C. Batch normalization: accelerating deep network training by reducing internal covariate shift. 32nd Int Conf Mach Learn ICML. 2015 Feb 11;1:448–56.
26 Chen LC, Papandreou G, Kokkinos I, Murphy K, Yuille AL. DeepLab: semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected CRFs. IEEE Trans Pattern Anal Mach Intell. 2018;40(4):834–48.
27 Simonyan K, Zisserman A. Very deep convolutional networks for large-scale image recognition. 3rd Int Conf Learn Represent ICLR 2015 - Conf Track Proc. 2014 Sep 4.
28 Ruder S. An overview of gradient descent optimization algorithms; 2016 Sep 15.
29 Weisstein EW. "Ellipse." From MathWorld: a Wolfram Web Resource. [Internet]. [cited 2023 April 10]. Available from: https://mathworld.wolfram.com/Ellipse.html.
30 Kosub S. A note on the triangle inequality for the Jaccard distance. Pattern Recognit Lett. 2019;120:36–8.