Automated Fetal and Placenta Localization with Fetal Biometry Estimation using Obstetric Volume Sweep Imaging in Rural Areas

Authors: Emilio J. Ochoa [1], Stefano E. Romero [2], Luis C. Revilla [2], Jose Zapana [2], Berta Ramos[3], Gloria Rios[4], Marika Toscano [5], Thomas J. Marini [6,7], Benjamín Castaneda [2]

[1]*Department of Biomedical Engineering, University of Rochester, New York, USA;* [2]*Departamento de Ingeniería, Pontificia Universidad Católica del Perú, Lima, Perú;* [3]*Centro de Salud, Pontificia Universidad Católica del Perú, Lima, Perú;* [4]*Medical Innovation and Technology, Lima, Peru;* [5] *Department of Obstetrics, John Hopkins University, Baltimore, USA;* [6]*Department of Imaging Sciences, University of Rochester Medical Center, Rochester, New York, USA;* [7]*University of Rochester Medical Center, Rochester, New York, USA;*

Keywords: telemedicine, ultrasound, fetal research, fetal monitoring.

Corresponding Author: Stefano E. Romero. Departamento de Ingeniería, Pontificia Universidad Católica del Perú, Lima, Perú. Email: sromerog@pucp.edu.pe

Abstract

Background

Prenatal care disparities significantly impact rural areas due to limited access to advanced medical technology, shortages of trained healthcare professionals, and inadequate infrastructure for comprehensive prenatal services. Ultrasound imaging is a critical component of prenatal monitoring but typically requires specialized personnel to acquire and interpret the imaging, limiting its accessibility in low-resource settings. Volume Sweep Imaging (VSI), a standardized ultrasound acquisition protocol independent of region and operator, presents an opportunity to bridge this gap. Yet, VSI still necessitates expert interpretation. To overcome this limitation, we propose integrating VSI with computer vision algorithms and artificial intelligence (AI) to fully automate fetal monitoring tasks, including placenta localization, fetal presentation detection, fetal biometry, and gestational age (GA) estimation.

Methods

This study assessed the epidemiological generalizability and computational efficiency of an AI-enabled VSI protocol across distinct regional contexts in Peru. A dataset consisting of ultrasound videos from 30 patients in Lima was used for model training and hyperparameter tuning via Leave-One-Out Cross-Validation. A second independent dataset comprising ultrasound videos from 86 patients across three diverse Peruvian regions was employed for hold-out testing to validate algorithmic performance and robustness. The AI system was evaluated based on sensitivity, specificity, accuracy, precision, and correlation with manual annotations provided by healthcare professionals.

Results

The developed AI model achieved a sensitivity of 98.6% and specificity of 85.7% for fetal presentation detection, indicating high diagnostic reliability. Placenta localization demonstrated an accuracy of 83.7% and a precision of 74.4%. Automated fetal biometry showed robust correlation with physician manual measurements, specifically for biparietal diameter (BPD) and head circumference (HC), with Spearman correlation coefficients indicating no significant difference from manual annotations. Furthermore, automated GA estimation correlated strongly with physician assessments, confirming its clinical validity.

Conclusions

This investigation highlights the potential of an AI-driven VSI system as an effective and scalable solution to prenatal care disparities in remote and resource-limited regions. By automating fetal monitoring tasks through integration of computer vision and AI, the proposed system reduces dependency on specialist interpretation and improves access to high-quality prenatal diagnostics. The study demonstrates that such technology can reliably support clinical decision-making, enhancing prenatal care delivery in underserved areas.

Introduction

Monitoring of the fetus is important for ensuring a successful birth and the well-being of both mother and child. World health organization suggests at least a trimester evaluation should happen for a correct monitoring of the fetus (1). In developed countries, access to fetal monitoring is commonly available due to the presence of numerous obstetric clinics and healthcare professionals. However, in low-middle income countries (LMICs), this essential service is often compromised by the lack of available sonographers and the sparse medical facilities(2). This discrepancy highlights the critical need for enhanced healthcare support and resources in rural communities to ensure equitable access to prenatal care for all expecting mothers. In addition, it is difficult for people in these areas to access hospitals as they may feel that going to a health facility is unwelcoming, ineffective and time-consuming (3).

Obstetrics ultrasound imaging is widely used, particularly for monitoring fetal growth and detecting potential complications during pregnancy or assessing characteristics of the fetus (4). In recent advancements, obstetrics ultrasound can be benefited from Computer Vision Algorithms powered by Artificial Intelligence (AI). These developments can revolutionize prenatal diagnostics, particularly in monitoring fetal growth and identifying pregnancy complications. For instance, Jing Lin et al. accurately detected cephalic and breech presentation (Figure 1) with AI to aid in choosing the right delivery method (AUC = 0.89, C.I. 0.77 - 1.00), as abnormal positions often result in cesarean sections (5). Crucial metrics such as fetal head circumference and placental location (Figure 2) are now more accurately determined, assisting in critical decision-making for delivery methods. Particularly, AI-driven techniques have shown promise in fetal biometry, enhancing the accuracy of measurements such as head circumference (HC) and biparietal diameter (BPD)(6-8). Van den Heuvel et al. developed Deep Learning (DL) models for classification of fetal segmentation and subsequent HC calculation with a mean difference, mean absolute error and Dice similarity coefficient (DSC) of 0.6 ± 5.9, 4.8 ± 3.4 and 97.2 ± 1.2, respectively for the third trimester(8). In the same line, Schilpzand et al. and our previous work have used U-Net for placenta segmentation with post-processing, while other researchers use 3D convolutional neural networks (CNN) architectures, as Yang et al(9-11). These works have the potential to aid in diagnosing conditions such as hydrocephalus and placenta previa, thereby significantly contributing to maternal and fetal health. The ongoing evolution of AI in obstetrics US underscores its potential to improve outcomes in maternal-fetal medicine, especially in resource-constrained settings where access to advanced diagnostic tools is limited (9-14).

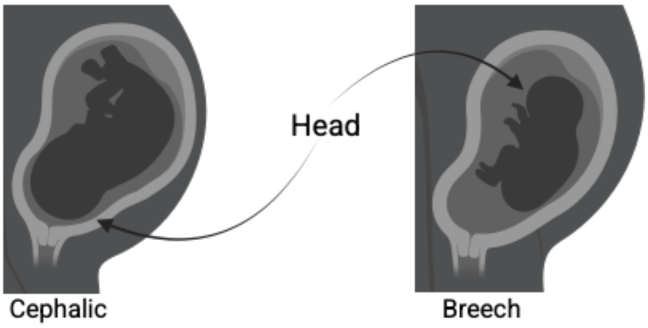


Figure 1. Head presentation. Breech position is a risk pathology. Cephalic: normal case. Created in https://BioRender.com

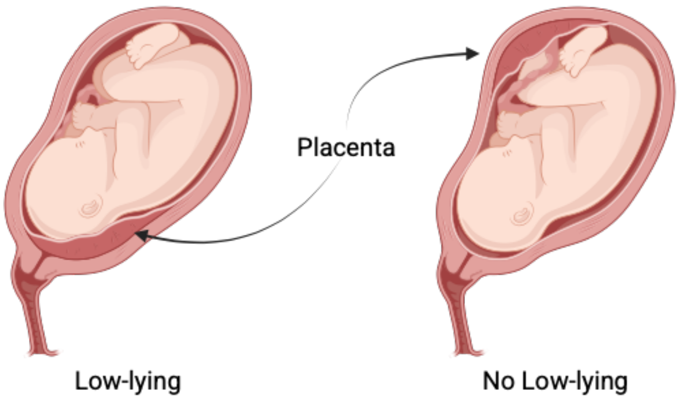


Figure 2. Placenta localization. Low-lying placenta is considered a risk pathology. No low-lying: normal case. Created in https://BioRender.com

Asynchronous ultrasound acquisition protocols have been developed for the rural areas scenarios, and previous studies underscore the feasibility of technologies in scenarios characterized by limited resources, where Volume Sweep Imaging (VSI) stand out significantly. In VSI for obstetric applications (VSI-OB), a simplified process as described in Figure 3 is employed where the ultrasound probe is methodically moved across the region of interest, where no acquisition expertise is needed(2). However, while the acquisition removes the need for a trained sonographer, the interpretation of the medical images requires a trained physician

AI algorithms are often tested on high-end scanners, potentially compromising their performance in real-world implementations, especially in rural areas with limited access to advanced equipment. Not only VSI-OB have been tested clinically in rural areas using a low-cost equipment(15), but similar protocols made by other groups were also enhanced by AI(11). Gomes et al. showed high AUC-ROC (0.98, C.I. 0.95-1.00) in a model to predict fetal malpresentation and gestational age (GA) by just using BUS for women in USA and Zambia, responding to shortage of adequately trained healthcare workers in low-to-middle-income countries(16). Comparatively, an AI for a VSI-OB protocol was implemented in Peru. A U-Net model matched specialist diagnosis perfectly for fetal presentation (100% agreement, Cohen κ=1) and showed 76.7% agreement for placental location (Cohen κ=0.59), with 100% sensitivity and specificity for fetal presentation and 87.5% sensitivity and 85.7% specificity for anterior placental location(10).

AI models trained with a particular dataset may be biased by the operator and the acquisition protocol used(17). In this study, we hypothesized that an AI model trained with VSI-OB data from one clinical site can be directly used in images/sweeps acquired at a different clinic demonstrating that operator training and performance have little impact in AI models applied to VSI-OB. In each site, operators were trained independently, and data was collected by different teams. The AI model is designed to detect fetal biometry (i.e. HC and BPD) to compute gestational age, as well as estimate fetal presentation and placenta localization. Additionally, and as an improvement to our previous work that required two separate AI models to segment the head and the placenta independently(10), we utilize a unique AI model responsible for head and placenta segmentation simultaneously with the inherent use of less computational resources which is key in rural settings.

This work aims to build upon our previous proof-of-concept study, where we demonstrated the feasibility of using AI algorithms trained on VSI-OB data collected from a single clinical site(10). In this study, we specifically evaluated the generalizability and transferability of these AI algorithms across three distinct Peruvian regions, irrespective of the operator who performed the ultrasound acquisition. We developed and validated an automated method for measuring fetal biometry (HC and BPD), detecting fetal presentation, localizing the placenta, as well as estimating gestational age (GA) in third-trimester pregnancies. Initially, a dataset of thirty patients was employed for hyperparameter tuning and model development, considered sufficient for robust algorithmic generalization. Subsequently, we tested the algorithm's performance on a separate dataset of eighty-six patients from geographically diverse locations. Additionally, the algorithm's practical applicability was assessed measuring the pipeline inference time using a low-cost tablet, further reinforcing its feasibility and potential for broad deployment in rural healthcare environments.

Material and Methods

Dataset

Two datasets previously collected for a telemedicine study(2) by the Peruvian startup Medical Innovation & Technology were used after approval by Cayetano Heredia National Hospital Ethics Committee. Each case comprises video clips of the VSI-OB protocol, acquired from women in the third trimester of pregnancy. The protocol follows eight video acquisitions presented in Figure 3. The ultrasound sweeps were performed in a standardized upward direction, sequentially progressing from the right side of the abdomen towards the left. During each sweep, the transducer captured a comprehensive three-dimensional dataset of the targeted anatomical region. The sweep protocol was systematically defined using easily identifiable anatomical landmarks, designed specifically to minimize the requirement for advanced technical proficiency or detailed anatomical knowledge.

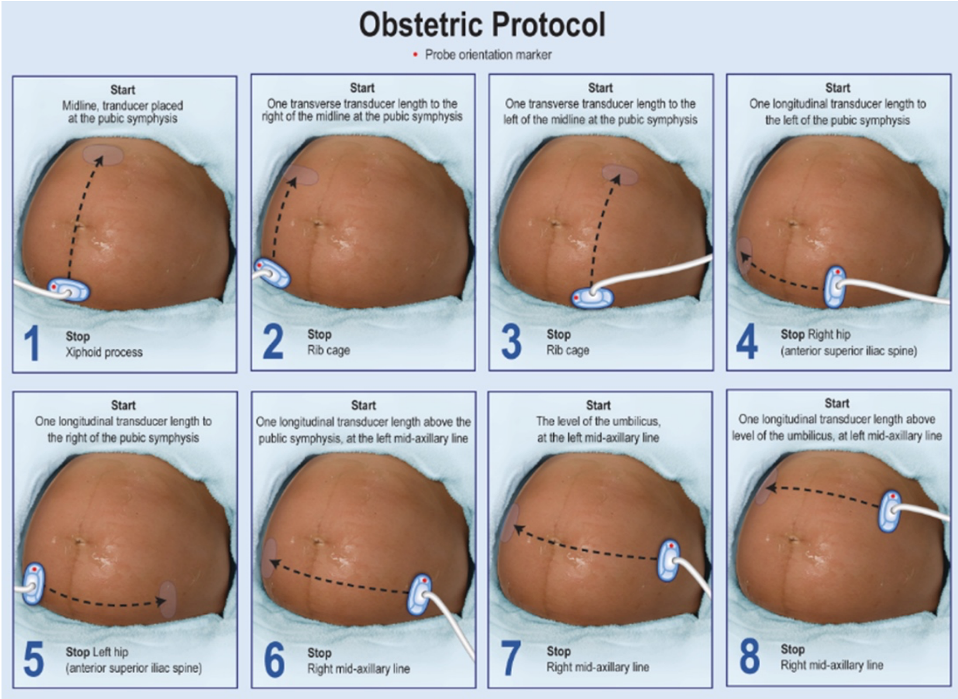
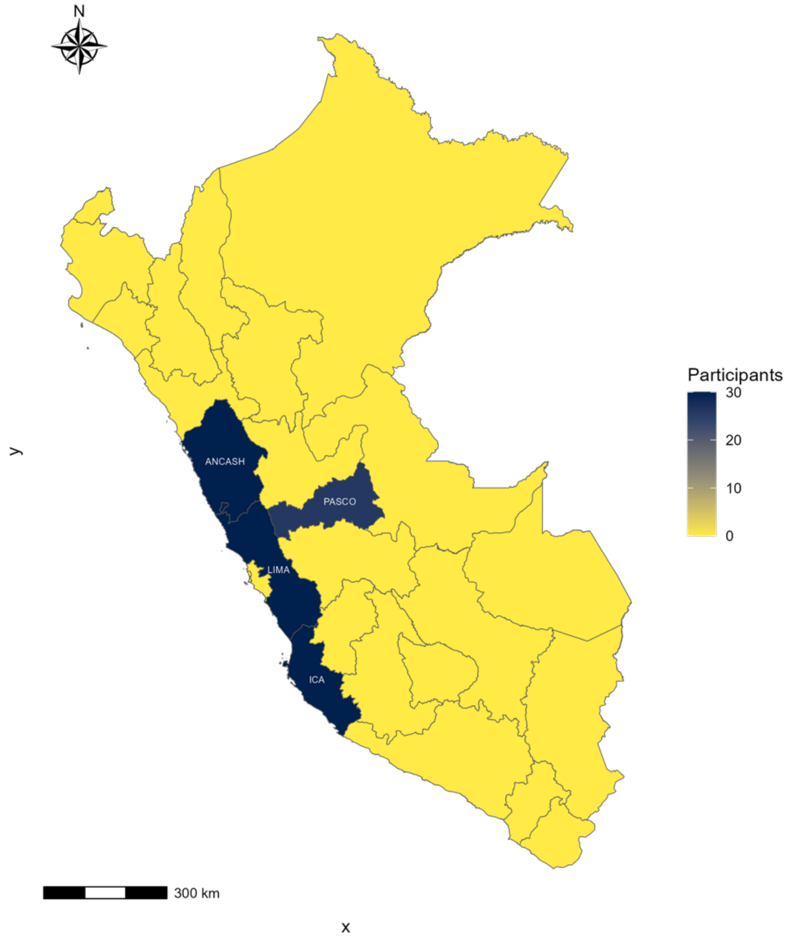


Figure 3. Volume Sweep Imaging (VSI) Protocol for Obstetric Ultrasound Assessment. The protocol begins with the transducer positioned at the midline, near the pubic symphysis, and proceeds with a series of longitudinal and transverse sweeps across the abdomen to cover key anatomical landmarks. These sweeps include movements toward the xiphoid process, the rib cage, and the anterior superior iliac spines on both sides, as well as along the mid-axillary lines.

Figure 4 shows the location of all the Peruvian regions from where the datasets were collected. Dataset 1 included thirty cases from Conde de la Vega Health Center in Lima, Peru, with videos annotated at a pixel level of the head and placenta, acquired by portable Mindray DP-10 ultrasound scanner (Mindray, China) with a 4.5 MHz transducer. Dataset 2 includes eighty-six cases acquired in in the Peruvian regions of Ancash, Cerro de Pasco and Ica, with the same ultrasound scanner. The purpose of each dataset consists of the following activities: Dataset 1 was used for grid search for hyperparameter tuning, Leave-One-Out-Cross-Validation (LOO-CV), and final model for semantic segmentation; Dataset 2 was used for testing of fetal malpresentation (80 cephalic and 6 podalic presentations), placenta location (86 no low-lying cases), gestational age and fetal biometry. Table 1 summarizes the distribution of the Dataset, information, and purpose of each dataset.



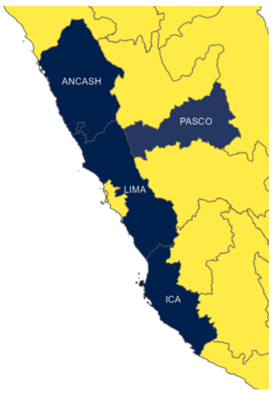


Figure 4: Map of Peru showing the regions from which the datasets were collected. The regions with data from the study are highlighted in dark blue, including Ancash, Ica, Lima, and Pasco.

Table 1: Distribution of the provided datasets. Dataset 1 was used for tuning hyperparameters and model validation. Dataset 2 was used for cephalic or breech presentation, placenta location and fetal - biometry such as BPD and HC

	Number of Cases	Clinical Information	Pixel-Level Annotations	Purpose
Dataset 1	Thirty	Yes	Yes	Training: Head and Placenta Segmentation
Dataset 2	Eighty-six	Yes	No	Testing: Cephalic or Breech Presentation, Placenta Location, Fetal Biometry, Gestational Age

Pre-Processing

A pre-processing for each frame of the videos was performed. The Contrast Limit Adaptive Histogram Equalization (CLAHE) algorithm(18) with a clip limit of 2.5 with a kernel size of 4x4 was used for each image of the video for enhancement purpose.

Multi-Input Attention U-Net

A multi-class semantic segmentation architecture was performed to identify head circumference and placenta. Multi-Input Attention U-Net(19, 20) was used. In comparison with U-Net, Multi-Input Attention U-Net allows more attention to specific areas or locations to integrate multiple data inputs. The detail of the architecture is shown in Figure 5. As a starting point, the selected static hyperparameters were learning rate: 0.01, momentum: 0.90, and decay 10^{-6} using a stochastic gradient decanter (SGD). Based on the results, it was decided to keep these hyperparameters.

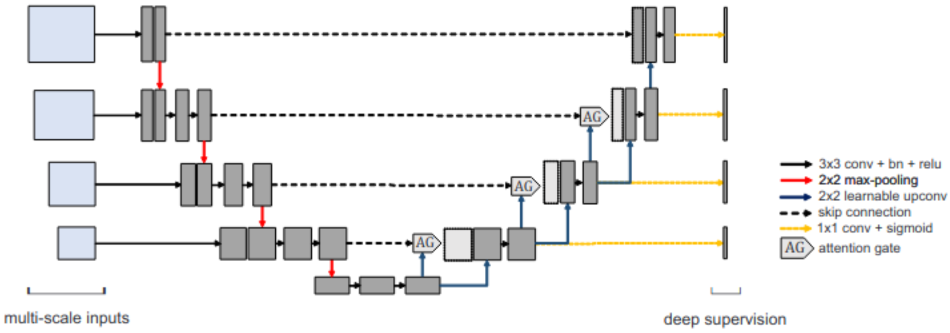


Figure 5. Multi-Input Attention U-Net architecture. Multi-scale inputs that are passed through a series of 3x3 convolutional layers and 2x2 max-pooling operations, represented by red connections. Skip connections (dashed lines) are used to retain spatial information across the network. The upsampling path consists of learnable 2x2 up-convolutions, and attention gates (AG) are applied at different stages to refine the feature maps, focusing on relevant regions in the input data. The final layers utilize 1x1 convolutions followed by a sigmoid activation for output prediction.

Training strategy

a. Grid Search

A grid search strategy was applied to obtain the best configuration of two hyperparameters: filter size and type of activation function. Dataset 1 was used for this purpose, following a distribution of 80% and 20% of the thirty patients for training and validation, respectively. In addition, a data augmentation strategy was considered by horizontal flipping and rotation of 5°. ReLU, Leaky ReLU activation functions, as well as Categorical Focal Jaccard Loss and Categorical Cross Entropy loss functions were tested(21, 22). The selection of the best configuration was performed by the DSC.

b. Leave-One-Out-Cross-Validation

After selection of hyperparameters in the previous section, a LOO-CV strategy was followed. The validation process was performed by training twenty-nine patients and extracting the model. Thirty models were evaluated with unseen test data.

Cephalic-Breech Presentation and Placenta Location

Heat maps were constructed to determine “cephalic” or “breech” presentation and placenta location (“low lying”, “no low-lying” and “undetermined”). For each class, two matrices were created projecting the object along the width and detecting the presence of any non-zero element for every sweep of the VSI-OB protocol. The heat map is generated by the summation of two matrices. The first matrix is created by performing a horizontal concatenation of sweep 2, sweep 1, and sweep 3 using a resize algorithm with the nearest neighbor approach with the largest value of the frames from these sweeps. The second matrix is generated by a vertical concatenation of sweep 8, sweep 7, and sweep 6 with a horizontal concatenation of sweep 5 and sweep 4 resizing with the largest value of sweep 8,7 or 6. After obtaining both matrices, a summation is performed. This procedure is separately applied to each of the two classes. Finally, a Gaussian filter is applied to improve the visualization of the segmented object. The resultant object in the heat map is used to determine the object's centroid position. In the case of Cephalic-Breech Presentation, if the centroid is in the heatmap's upper section, it is labeled as 'cephalic'; otherwise, it is designated 'breech'. In the case of the placenta, if the centroid is located within the top three-quarters is classified as 'low-lying', if it is in the bottom quarter is classified as 'no low-lying'. Figure 6 shows this process to identify the spatial location in the mother's womb. This process is the same for both head and placenta but using their corresponding labels.

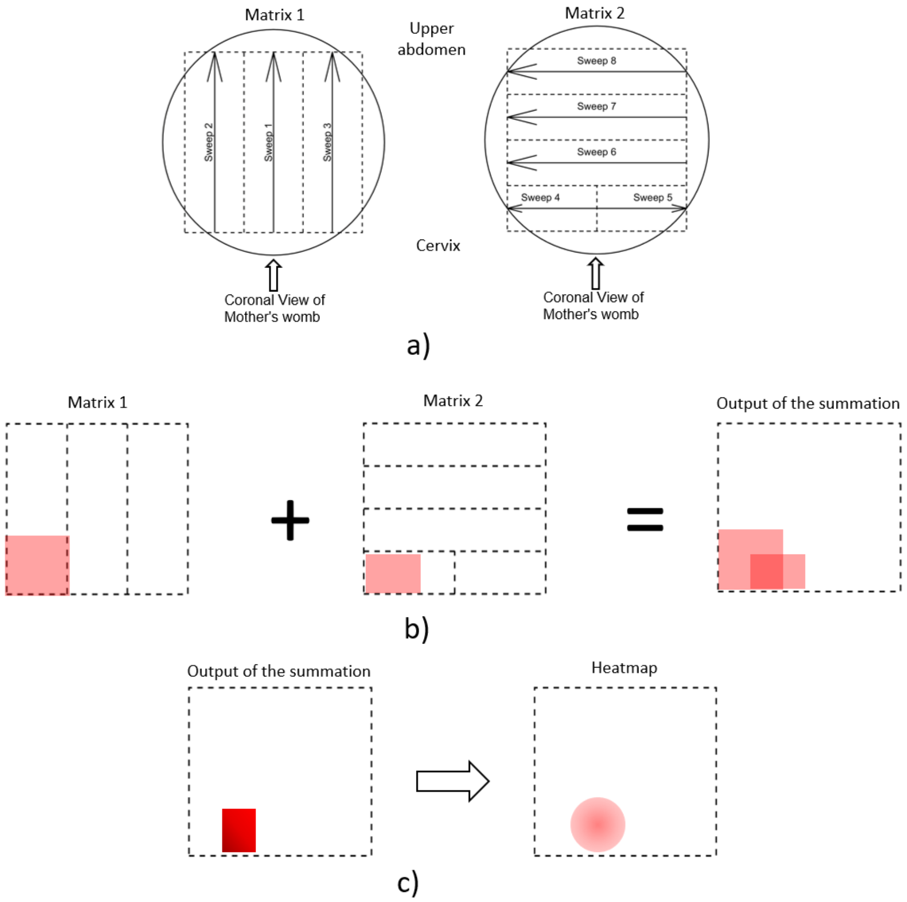


Figure 6. Heat Map Generation Process for Cephalic–Breech Presentation and Placenta Location Using the VSI-OB Protocol. (a) Two matrices, Matrix 1 and Matrix 2, are constructed based on a coronal view of the uterus. Matrix 1 merges sweeps 2, 1, and 3 horizontally, while Matrix 2 first stacks sweeps 8, 7, and 6 vertically, then concatenates sweeps 5 and 4 horizontally. (b) Each matrix is resized via nearest-neighbor interpolation, selecting the maximum response across frames for each sweep. The matrices are then summed to form a combined output matrix. (c) A Gaussian filter is applied to the combined matrix, resulting in a heat map that highlights the spatial location of the relevant anatomical structure (e.g., fetal head or placenta).

Fetus Biometry

Automatic Pixel Spacing

An algorithm for Automatic Pixel Spacing (APS) was designed considering the image characteristics of the ultrasound equipment Mindray DP-10 used for VSI obstetrics applications (Figure 7.a). A manual region of interest (ROI) of the ultrasound image ruler was selected in two parts to identify the dimensions of the image. The first part corresponded to vertical lines uniformly separated to store the amplitude of the lines of the ultrasound image ruler. Indices of elements in the line that are greater than 60% or brightness were kept and sorted in ascending order to calculate the distance between them. The second image corresponds to the numbers of each line in centimeters. For enhanced image resolution, a resize of 500% was applied. The numbers were identified by using the Optical Character Recognition (OCR) algorithm with the EasyOCR library from Python(23).

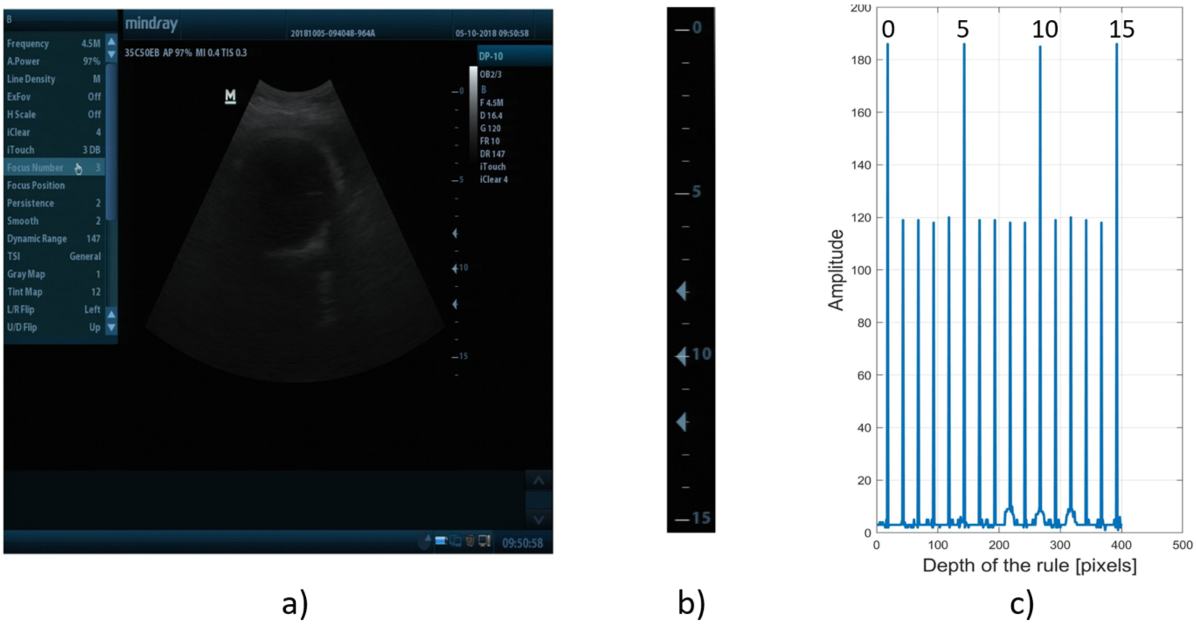


Figure 7. Automatic Pixel Spacing (APS) algorithm applied to ultrasound imaging. Figure 5.a, a manual region of interest (ROI) was selected from the ultrasound image, dividing it first to capture the uniform vertical lines. Figure 5.b shows the second ROI division, containing the corresponding numeric values of the depth ruler in centimeters. The algorithm identified the center column of the image, storing the amplitude values of the detected lines. Figure 5.c, the algorithm computed the pixel spacing by analyzing the amplitude of vertical lines.

Frame Selection for fetus biometry

The eight videos from each case of Dataset 2 were used as input in the algorithm. A selection of the largest area of all the frames from the eight segmented videos was extracted. The selected frame is a representative frame of the case. The prediction of this frame was fitted with an ellipse to obtain the minor and major axes to calculate the BPD and the Occipitofrontal Diameter (OFD). The HC was calculated as proposed by Borges et al. (Eq. 1)(24).

$$HC = 1.62 \times (BPD + OFD) \tag{1}$$

Metrics - Statistical Analysis

The automatic classification of Cephalic-Breech presentation was compared with physician clinical annotations using sensitivity and specificity as key metrics. Placenta location was measured by focusing on low-lying cases but with three possible classifications: low-lying, no low-lying, and undetermined. The presence of placenta was determined by accuracy (Eq.2). The performance in the classification was determined by the True Negative Rate (TNR) - (Eq. 3).

$$Accuracy = \frac{TN}{Total} \tag{2}$$

$$TNR = \frac{TN}{TN+FP} \tag{3}$$

Automated calculations for BPD and HC were assessed against physician measurements using linear regression with Spearman correlation coefficient, and Bland-Altman compared with a theoretical mean of 0 using a one-sample t-test.

Results

The demographic statistics of the participants included in the study are presented in Table 2 (**Supplementary material)**. The patients' age ranged from under 25 to over 45 years, with a mean age of 31.67 ± 5.42 years. In terms of body mass index (BMI), the average BMI was 33.29 ± 7.95, indicating an overweight population. In terms of body mass index (BMI), the average BMI was 33.29 ± 7.95, indicating an overweight population. Regarding geographic distribution, participants were recruited from four regions in Peru: Lima (25.86%), Ica (25.86%), Ancash (25.86%), and Cerro de Pasco (22.42%). The parity data revealed an average of 1.11 ± 1.35 previous births. Regarding amniotic fluid and placental assessments, the average DVP oligo and AFV oligo values were 1.56 ± 0.89 and 2.8 ± 1.76, respectively, while the mean DVP poly was 8.72 ± 1.2. Prevalence of various maternal conditions in the cohort was notable. Among these, chronic hypertension (15.87%) and gastrointestinal disease (20.73%) were prominent, while endocrine diseases affected 14.95% of the patients. Cardiac disease was present in 5.76% of the cases, and 7.45% of the patients had gestational diabetes. Other conditions observed included hematologic disease (15.03%), neurologic disease (18.13%), and pulmonary disease (18.04%). Additionally, 12.17% of patients developed pre-eclampsia, while 7.33% had pregestational diabetes. A small percentage (2.63%) of the cohort experienced preterm labor.

Grid Search

Table 3 shows the results of the Grid Search methodology. For these experiments, based on the Dice coefficient it was found that the best hyperparameter selection was with a Filter Size of 512 using ReLU.

Table 3: Results of Dice coefficient using Grid Search for hyperparameters tunning

	Categorical Cross-Entropy		Categorical Focal Jaccard	
Filter Size	Leaky_ReLU	ReLU	Leaky_ReLU	ReLU
32	61.46	62.21	64.05	63.45
64	64.04	63.33	66.33	66.56
128	66.51	66.70	72.05	73.96
256	67.35	66.32	73.24	70.22
512	65.49	67.03	75.86	76.33

Leave-One-Out-Cross-Validation

Table 4 shows the results of the sensitivity, specificity, accuracy, Positive Predictive Value (PPV), and Negative Predictive Value (NPV) on the detection of the head and placenta in Dataset 1. After these results, a unique final model was trained using this Dataset for further spatial location of the head and placenta.

Table 4: Results of Sensitivity, Specificity, Accuracy, PPV and NPV of the Leave-One-Out-Cross-Validation

	Sensitivity	Specificity	Accuracy	PPV	NPV
Head	0.93	0.98	0.97	0.80	0.99
Placenta	0.74	0.95	0.91	0.80	0.94

Head Presentation

Figure 8.a and Figure 8.b show a representative result of cephalic and breech presentations. Table 5 shows the results of the sensitivity, specificity and Cohen's kappa coefficient by region in Dataset 2.

Table 5: Results of Sensitivity, Specificity and Cohen's Kappa Coefficient for Cephalic-Breech Presentation by region.

Region	Head Presentation		
	Sensitivity (%)	Specificity (%)	Cohen's Kappa coefficient
All sites	98.6	85.7	0.84
Ancash	96.6	50.0	0.46
Ica	100.0	100.0	1.00
Cerro de Pasco	100.0	100.0	1.00

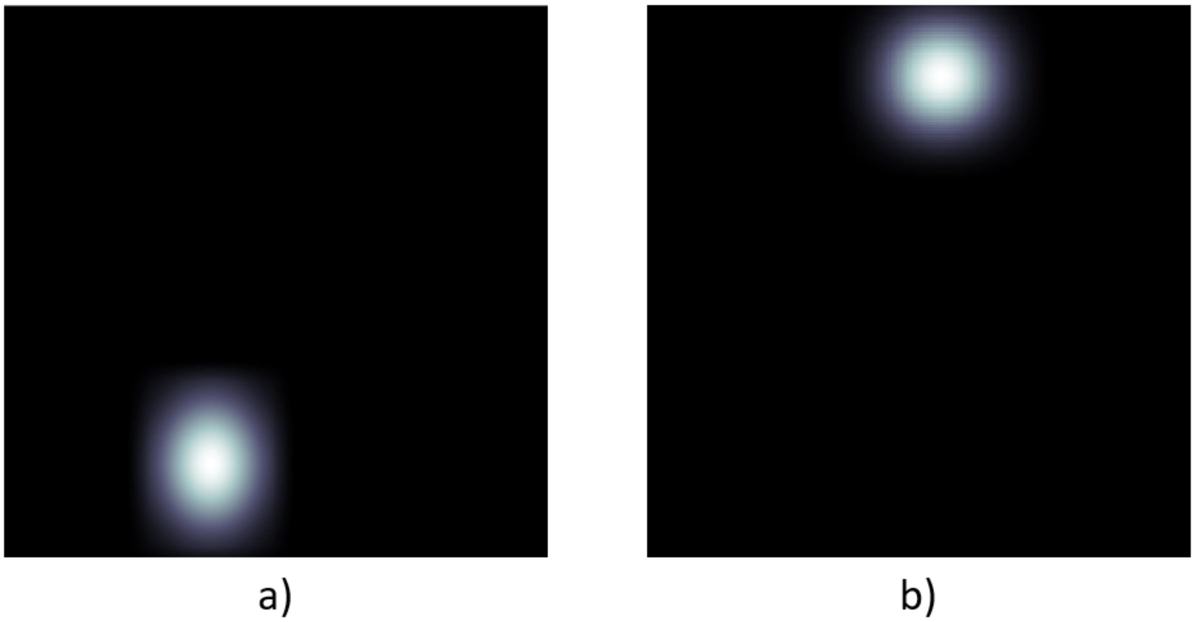


Figure 8. Heat Map Visualization for Cephalic and Breech Presentations in VSI-OB. Figure 6.a, the heat map shows a case of cephalic presentation, where the centroid is in the lower section of the heat map, indicating the fetal head's position near the cervix. In Figure 6.b, the heat map represents a breech presentation, with the centroid in the upper section, corresponding to the fetal head positioned near the upper abdomen.

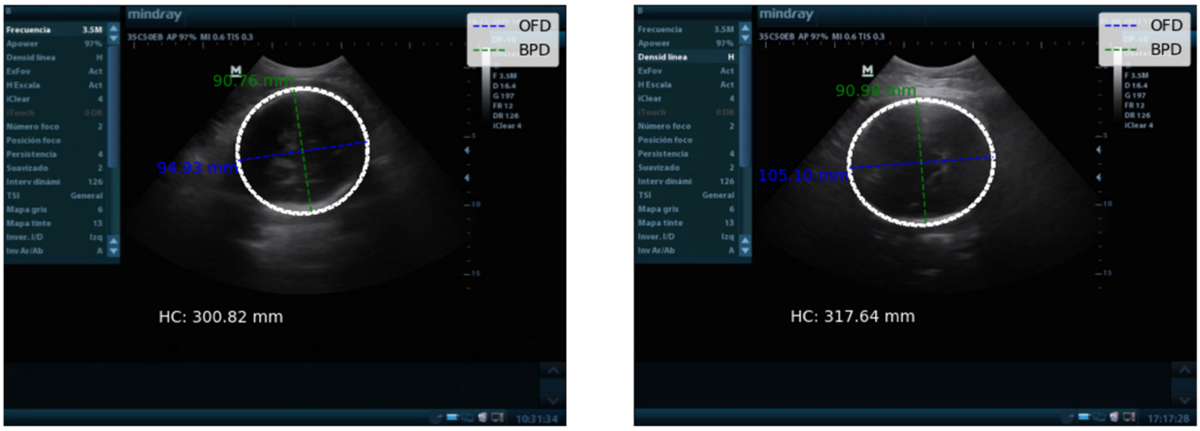
Placenta Location

In the case of the placenta, the classes of low-lying, not low-lying, and undetermined were considered due to the nature of the heatmap. From Dataset 2, only low-lying placenta cases exist. Table 6 shows the results for this dataset.

Table 6: Results of the algorithm using the class low-lying, no low-lying, and undetermined by region. In addition, the accuracy and precision were calculated

Region	Automatic Placenta Location			
	Not low-lying	Low-lying	Undetermined	Accuracy (%)

Peru (Total)	72	1	13	83.72
Ancash	28	0	3	90.32
Ica	21	1	2	87.50
Cerro de Pasco	23	0	8	74.19



a) b)

Figure 9. Representative Ultrasound Results with BPD and OFD Calculations Using Multi-Input Attention U-Net. In both cases, the model's predictions are fitted to an ellipse, outlined in white. The Biparietal Diameter (BPD) is derived from the ellipse's minor axis and is displayed in green, while the Occipitofrontal Diameter (OFD) is calculated using the major axis, shown in blue. Head circumference (HC) is also calculated and displayed at the bottom of each image.

Fetal biometry and gestational age

Figure 9 shows two representative results of two different cases. The prediction of the Multi-Input Attention U-Net was fitted to an ellipse which is represented in white color. The BPD is calculated by using the minor axis of the ellipse and the OFD is calculated by using the major axis. These values are represented in green and blue, respectively. Table 7 shows the results of the metrics in mm of the BPD, HC and GA from manual annotations and automatic annotations. Figure 10, Figure 11 and Figure 12 illustrate the statistical analysis using linear regression, Spearman correlation coefficient, and Bland Altman for BDP, HC, and GA, respectively.

Table 7: Results of mean ± standard deviation of the BPD, HC and GA measured by the manual segmentation realized by a physician and the automatic measurement by the representative frame of the VSI-OB using the proposed algorithm

	Physician Measurement	Automatic Measurement	Bland-Altman (95% CI)	p-value (p>0.01)
BPD (mm)	85.25 ± 8.32	86.93 ± 7.09	-1.97 (-8.47, 4.52)	0.23
Ancash	85.97 ± 10.84	87.86 ± 9.13	-1.89 (-7.99, 4.21)	0.33
Ica	84.71 ± 6.69	86.11 ± 6.35	-1.40 (-7.64, 4.84)	0.44
Cerro de Pasco	84.02 ± 6.56	87.77 ± 9.13	-1.88 (-7.99, 4.21)	0.21
HC (mm)	307.63 ± 28.10	300.69 ± 24.82	5.58 (-26.08, 37.25)	0.05
Ancash	307.80 ± 38.39	299.93 ± 31.89	7.88 (-15.92, 31.68)	0.16
Ica	304.14 ± 21.39	300.17 ± 24.05	4.26 (-19.75, 28.27)	0.27
Cerro de Pasco	307.33 ± 20.57	303.08 ± 17.69	3.97 (-38.54, 46.49)	0.40
GA using only BPD (weeks)	34.67 ± 3.60	35.16 ± 3.14	-0.48 (-4.12, 3.14)	0.42
Ancash	35.16 ± 4.63	35.52 ± 3.91	-0.35 (-3.99, 3.28)	0.84
Ica	34.42 ± 3.09	34.90 ± 2.70	-0.48 (-4.23, 3.26)	0.51
Cerro de Pasco	34.38 ± 2.65	35.04 ± 2.56	-0.67 (-4.26, 2.92)	0.39

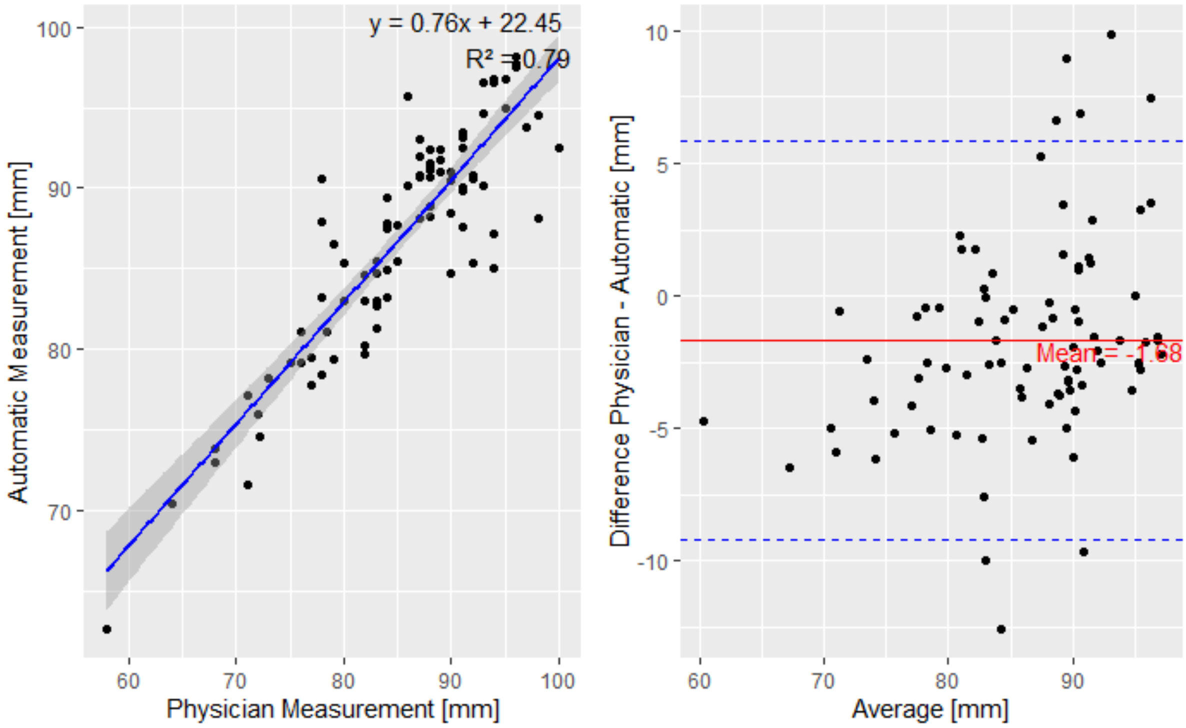


Figure 10. Statistical Analysis of Biparietal Diameter (BPD) Measurements Using Linear Regression, Spearman Correlation, and Bland-Altman Plot

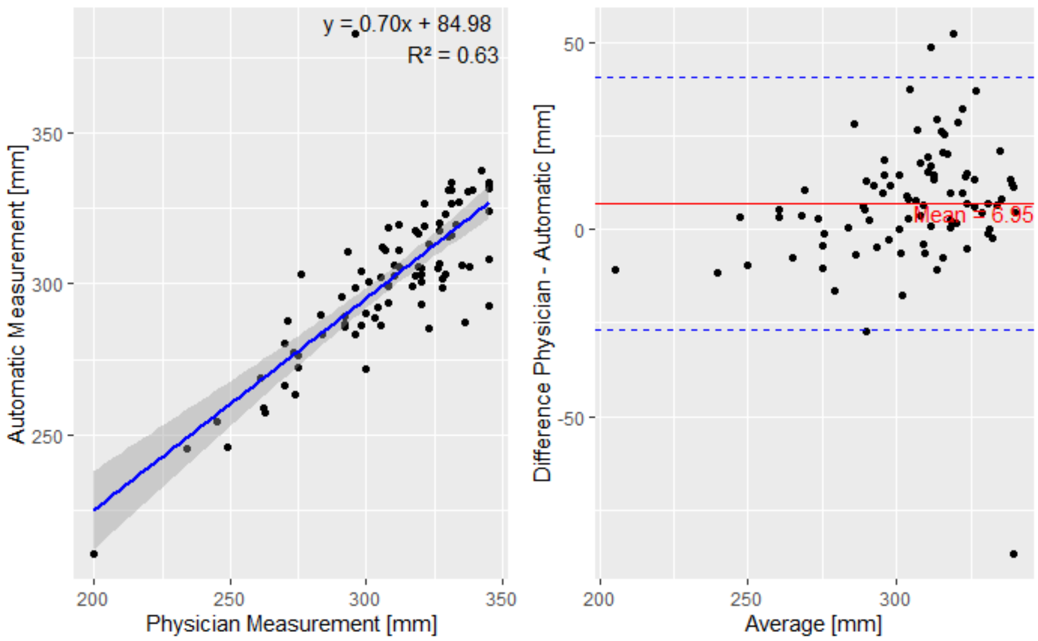


Figure 11. Statistical Analysis of Head Circumference (HC) Measurements Using Linear Regression, Spearman Correlation, and Bland-Altman Plot

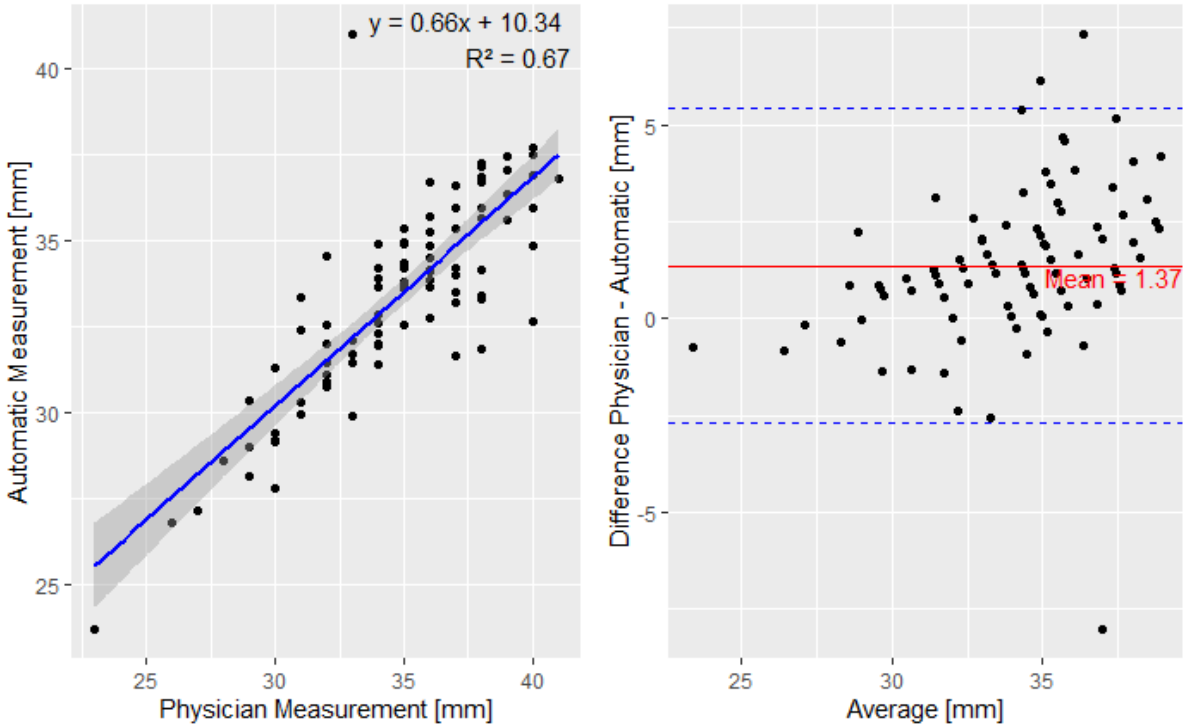


Figure 12. Statistical Analysis of Gestational Age (GA) Measurements Using Linear Regression, Spearman Correlation, and Bland-Altman Plot

Inference time analysis

The processing time was analyzed as a direct result in a clinical environment, using the Medical Box as seen in figure 13 (Medical, Innovation & Technology, Lima, Peru)(25). This device is designed for use in low-resource settings. For an obstetric patient, the prediction times for the models tested in CPU mode (since the Medical Box does not have a GPU, making it a low-cost solution) with a test dataset were as follows: the AI prediction time (8 sweeps, CPU mode, test dataset) was 5.35 ± 1.45 minutes, while the total time, including both pre- and post-processing, was 12.05 ± 3.25 minutes. Therefore, in a real-world environment, we could achieve a complete workflow result in less than 15 minutes in a rural area, without adding significant computational overhead such as the use of a GPU.

Figure 13. Medical Box, proposed device to incorporate in the flowchart. This device can be added in any commercial ultrasound device.

Discussion

The findings of the presented work support the initial hypothesis that an AI model trained with VSI-OB data from one clinical site can be effectively applied to other sites when using similar equipment, reflecting minimal operator dependency and robust generalizability. The high sensitivity and specificity for breech fetal presentation (98.6% and 85.7%, respectively) indicate reliable detection in varied environments. This high sensitivity ensures nearly all breech presentations will be accurately identified. In the case of false positives for cephalic presentations, the clinical risk remains minimal, as subsequent follow-up by trained healthcare personnel is standard practice, even in low-resource settings.

Similarly, the model's ability to accurately localize the placenta (83.72% accuracy) underscores its potential utility for identifying high-risk conditions, such as low-lying placenta. Although these results suggest promising capabilities for risk stratification, a dedicated study specifically focused on the detection and differentiation of placenta previa is required to fully validate clinical applicability and performance in identifying this critical obstetric condition. Furthermore, the close agreement between automatically measured fetal biometry (BPD, HC) and physician annotations reinforces the clinical viability of an integrated VSI-OB + AI approach.

Concerning cephalic and breech presentations, high values of sensitivity and specificity were obtained, as seen in Tables 4 and 5. These values are comparable with the baseline of twenty-eight cases; from Dataset 2, eighty-one cases were cephalic and five were breech, and only one cephalic case was misclassified as breech despite ground truth indicating cephalic. Regarding placental analysis, Dataset 2 contained no placenta previa cases, consistent with its 3–5 per 1 000 incidence worldwide (26), yet overall placenta detection accuracy reached approximately 90 % as seen in Table 4, and no low-lying (normal) placenta detection performed robustly (Table 6), with just thirteen scans deemed indeterminate due to absence of placental views across all sweeps and a single false negative. These outcomes likely reflect the placenta's irregular, heterogeneous echogenicity in the third trimester(27), occasional indistinct placenta–cervix boundaries necessitating transvaginal imaging, and the impact of some sweeps containing fewer than ten useful frames or lacking heatmap overlap, while the consistent visibility of the placenta in key sweeps underscores the pipeline's screening potential. Automated fetal biometry showed high correlation with physician measurements—particularly biparietal diameter versus head circumference (Table 7)—surpassing the baseline and yielding gestational age estimates with a mean difference of 0.48 months (CI –4.12, 3.14), comparable to the 17-day (≈0.56-month) third-trimester error reported by physician measurements(28) without evident regional bias, demonstrating the clinical viability of the VSI + AI approach for comprehensive prenatal assessment.

Collectively, these outcomes suggest that an AI-driven VSI-OB protocol can be deployed successfully across different remote regions, thereby advancing equitable prenatal care in resource-limited settings. This is of particular importance in Peru; for instance, the most populated region in the Peruvian Amazon is Loreto, which has the lowest percentage (74.5%) of hospital deliveries in the country. In fact, in rural areas, it is even lower (48.79%)(29). In these areas, such as the Peruvian Amazon, prenatal care often involves traditional midwives or other healthcare workers. The VSI-OB protocol can be utilized by these caregivers, enhancing their capacity to provide more comprehensive prenatal assessments without the necessity of direct physician involvement. Previous evidence indicates that standardized VSI-OB training, significantly minimizes inter-operator variability. For instance, ultrasound-naive operators trained for only 3 hours achieved diagnostic-quality images in 96.4% of examinations, with protocol deviations affecting diagnostic potential in just 7.7% of cases. Importantly, operator variability improved notably over time (30).

VSI-OB empowered by AI presents a quick and efficient solution to provide relevant information to accurately refer risky pregnant patients to the hospital. This work also technically improves our previous study which is considered a baseline(10). We previously used U-Net as a starting point. However, Parvathavarthini et al.(31) showed that Attention U-Net generated more accurate results for HC calculation. In addition, Multi-input Attention U-net outperformed the traditional U-net on open ultrasound datasets such as BUS 2017 and ISIC 2018(19) in detecting breast nodules and skin lesions respectively. Based on this, using this new model, we obtained an improvement in the metrics in comparison with the baseline results from our previous work in both head (0.97 vs 0.91) and placenta (0.91 vs 0.80) detection(10). The improvement could be achieved due to the deep supervision mechanism and the focal-Tversky loss; characteristics that differ from original U-NET(19).

Finally, these experiments demonstrated the robustness of VSI-OB when conducted on high-performance computers with GPUs. Looking forward, the application of VSI-OB in low-resource settings offers an exciting opportunity for broader impact(2). Based on this it was necessary to deploy these systems in low-cost devices. Our results, measured with the Medical Box, demonstrate that these models can run efficiently on low-compute-power hardware, validating their suitability for resource-limited environments.

Our future work will focus on developing a platform and reporting tool for rapid triage of at-risk patients. To achieve this, we will extend the VSI-OB pipeline by adding new classes—such as heart segmentation, amniotic fluid detection and femur segmentation—to enrich automated fetal monitoring, and by validating the ROI selection and APS on alternative and portable ultrasound systems to support teleultrasound applications beyond Mindray equipment. We also plan to expand our dataset to include first and second trimester scans, increasing case diversity and volume well beyond the current 30 training patients, the small number of breech (six) and the lack of low-lying placenta cases to improve failure-mode analysis and robustness. Placental segmentation in the third trimester remains challenging due to its irregular, heterogeneous echogenicity and sometimes indistinct placenta–cervix boundary, issues that may require transvaginal imaging, while some sweeps contain too few frames or lack heatmap overlap, and potential errors in automated head-circumference calculation demand careful validation and refinement of the processing formulas. By addressing these limitations and broadening both the equipment tested and the gestational stages covered, we aim to enhance the generalizability and clinical utility of VSI + AI for rural and resource-limited settings.

Conclusion

The findings of this work support the initial hypothesis that an AI model trained with VSI-OB data from one clinical site can effectively generalize to other clinical sites, as demonstrated by consistent results across diverse regions using similar equipment. The minimal variability in operator performance further highlights VSI's potential as a standardized data collection method, independent of regional differences or operator experience. High sensitivity (98.6%) and specificity (85.7%) for fetal presentation and reliable placenta localization (83.72% accuracy) confirm robust algorithmic generalizability. Additionally, the algorithm's computational feasibility was validated with the low-cost cpu device, reinforcing its practical applicability in resource-limited rural environments. Overall, our findings provide a proof of concept that AI-driven VSI-OB protocols could standardize obstetric ultrasound imaging across diverse regions, thereby promoting equitable access to accurate prenatal monitoring irrespective of geographic or resource constraints, and lay the groundwork for large-scale deployment of these protocols.

Availability of data and materials statements

The database used for the analysis in this study was provided by the company Medical Innovation & Technology and contains proprietary information. Due to the sensitive nature of the data and confidentiality agreements with the company, the raw data cannot be made publicly available. However, anonymized and aggregated data that support the findings of this study can be shared upon reasonable request, subject to approval by Medical Innovation & Technology. Requests for data access should be directed to the corresponding author and will be reviewed in accordance with the company's data sharing policies to ensure compliance with confidentiality and intellectual property agreements.

References

1. Organization WH. WHO recommendations on antenatal care for a positive pregnancy experience. WHO recommendations on antenatal care for a positive pregnancy experience2016. p. 152-.
2. Toscano M, Marini TJ, Drennan K, Baran TM, Kan J, Garra B, et al. Testing telediagnostic obstetric ultrasound in Peru: a new horizon in expanding access to prenatal ultrasound. BMC Pregnancy and Childbirth. 2021;21(1):328.
3. Yajahuanca A, Rosario S. Concepções sobre saúde reproductiva e sexualidade entre os descendentes Kukamas Kukamirias, Perú. 2009.
4. Whitworth M, Bricker L, Mullan C. Ultrasound for fetal assessment in early pregnancy. Cochrane database of systematic reviews. 2015(7).
5. Lin J, Liu W, Gu W, Zhou Y. A prospective study using an individualized nomogram to predict the success rate of external cephalic version. Scientific Reports. 2022;12(1):11795.
6. Diniz PH, Yin Y, Collins S. Deep learning strategies for ultrasound in pregnancy. European Medical Journal Reproductive health. 2020;6(1):73.
7. Xiao S, Zhang J, Zhu Y, Zhang Z, Cao H, Xie M, et al. Application and progress of artificial intelligence in fetal ultrasound. Journal of clinical medicine. 2023;12(9):3298.
8. van den Heuvel TL, de Bruijn D, de Korte CL, Ginneken Bv. Automated measurement of fetal head circumference using 2D ultrasound images. PloS one. 2018;13(8):e0200412.
9. Schilpzand M, Neff C, van Dillen J, van Ginneken B, Heskes T, de Korte C, et al. Automatic placenta localization from ultrasound imaging in a resource-limited setting using a predefined ultrasound acquisition protocol and deep learning. Ultrasound in Medicine & Biology. 2022;48(4):663-74.
10. Arroyo J, Marini TJ, Saavedra AC, Toscano M, Baran TM, Drennan K, et al. No sonographer, no radiologist: New system for automatic prenatal detection of fetal biometry, fetal presentation, and placental location. PloS one. 2022;17(2):e0262107.
11. Yang X, Yu L, Li S, Wen H, Luo D, Bian C, et al. Towards automated semantic segmentation in prenatal volumetric ultrasound. IEEE transactions on medical imaging. 2018;38(1):180-93.
12. Fiorentino MC, Villani FP, Di Cosmo M, Frontoni E, Moccia S. A review on deep-learning algorithms for fetal ultrasound-image analysis. Medical image analysis. 2023;83:102629.
13. Qi H, Collins S, Noble A, editors. Weakly supervised learning of placental ultrasound images with residual networks. Medical Image Understanding and Analysis: 21st Annual Conference, MIUA 2017, Edinburgh, UK, July 11–13, 2017, Proceedings 21; 2017: Springer.
14. Gleed AD, Chen Q, Jackman J, Mishra D, Chandramohan V, Self A, et al. Automatic image guidance for assessment of placenta location in ultrasound video sweeps. Ultrasound in Medicine & Biology. 2023;49(1):106-21.
15. Marini TJ, Oppenheimer DC, Baran TM, Rubens DJ, Toscano M, Drennan K, et al. New Ultrasound Telediagnostic System for Low-Resource Areas: Pilot Results From Peru. Journal of Ultrasound in Medicine. 2021;40(3):583-95.
16. Gomes RG, Vwalika B, Lee C, Willis A, Sieniek M, Price JT, et al. A mobile-optimized artificial intelligence system for gestational age and fetal malpresentation assessment. Communications Medicine. 2022;2(1):128.

17. Tolsgaard M, Svendsen M, Thybo JK, Petersen OB, Sundberg K, Christensen AN. Does artificial intelligence for classifying ultrasound imaging generalize between different populations and contexts? Ultrasound in Obstetrics and Gynecology. 2021;57(2):342-3.
18. Pizer SM, editor Contrast-limited adaptive histogram equalization: Speed and effectiveness stephen m. pizer, r. eugene johnston, james p. ericksen, bonnie c. yankaskas, keith e. muller medical image display research group. Proceedings of the first conference on visualization in biomedical computing, Atlanta, Georgia; 1990.
19. Abraham N, Khan NM, editors. A novel focal tversky loss function with improved attention u-net for lesion segmentation. 2019 IEEE 16th international symposium on biomedical imaging (ISBI 2019); 2019: IEEE.
20. Fang L, Wang X. Multi-input Unet model based on the integrated block and the aggregation connection for MRI brain tumor segmentation. Biomedical Signal Processing and Control. 2023;79:104027.
21. Agarap AF. Deep learning using rectified linear units (relu). arXiv preprint arXiv:180308375. 2018.
22. Xu B, Wang N, Chen T, Li M. Empirical evaluation of rectified activations in convolutional network. arXiv preprint arXiv:150500853. 2015.
23. Pattanayak BK, Biswal AK, Laha SR, Pattnaik S, Dash BB, Patra SS, editors. A novel technique for handwritten text recognition using easy OCR. 2023 International Conference on Self Sustainable Artificial Intelligence Systems (ICSSAS); 2023: IEEE.
24. Peixoto AB, da Cunha Caldas TMR, Dulgheroff FF, Martins WP, Júnior EA. Fetal biometric parameters: Reference charts for a non-selected risk population from Uberaba, Brazil. Journal of Ultrasonography. 2017;17(68):23.
25. Ferrer J, Chaumont T, Trujillo L, Fernandez I, Guerrero J, Stewart P, et al., editors. New tele-diagnostic model using volume sweep imaging for rural areas. 2017 39th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC); 2017: IEEE.
26. Adere A, Mulu A, Temesgen F. Neonatal and maternal complications of placenta Praevia and its risk factors in Tikur Anbessa specialized and Gandhi memorial hospitals: Unmatched case-control study. Journal of pregnancy. 2020;2020(1):5630296.
27. Bowman ZS, Kennedy AM. Sonographic appearance of the placenta. Current Problems in Diagnostic Radiology. 2014;43(6):356-73.
28. Skupski DW, Owen J, Kim S, Fuchs KM, Albert PS, Grantz KL, et al. Estimating gestational age from ultrasound fetal biometrics. Obstetrics & Gynecology. 2017;130(2):433-41.
29. Del Mastro N I, Tejada-Llacsa PJ, Reinders S, Pérez R, Solís Y, Alva I, et al. Home birth preference, childbirth, and newborn care practices in rural Peruvian Amazon. PloS one. 2021;16(5):e0250702.
30. Erlick M, Marini T, Drennan K, Dozier A, Castaneda B, Baran T, et al. Assessment of a brief standardized obstetric ultrasound training program for individuals without prior ultrasound experience. Ultrasound quarterly. 2023;39(3):124-8.
31. Parvathavarthini S, Sharvanthika K, Sindhu S, Kaviya K, editors. Fetal head circumference measurement from ultrasound images using attention U-net. 2023 Fifth International Conference on Electrical, Computer and Communication Technologies (ICECCT); 2023: IEEE.