Adaptive Frame Selection for Gestational Age Estimation from Blind Sweep Fetal Ultrasound Videos

Tanya Akumu[1], Marawan Elbatel[2], Victor M. Campello[1], Richard Osuala[1], Carlos Martin-Isla[1], Ignacio Valenzuela[1,3], Xiaomeng Li[2], Bishesh Khanal[4], and Karim Lekadir[1,5]

[1] Departament de Matemàtiques i Informàtica, Universitat de Barcelona, Spain
tanya.akumu@ub.edu
[2] Department of Electronic and Computer Engineering, The Hong Kong University of Science and Technology, Hong Kong SAR, China
[3] Fetal Medicine Research Center, BCNatal, Barcelona, Spain
[4] Nepal Applied Mathematics and Informatics Institute (NAAMII)
[5] Institució Catalana de Recerca i Estudis Avançats (ICREA), Barcelona, Spain

Abstract. The blind sweep ultrasound protocol, coupled with artificial intelligence (AI), offers promising solutions for expanding ultrasound availability in low-resource settings. However, existing AI approaches for gestational age (GA) prediction using bind sweeps face challenges like reliance on manual segmentation, computational inefficiency from high frame volume, and suboptimal sampling strategies that compromise performance, particularly with smaller datasets. We propose SelectGA, a novel framework for automated blind sweep analysis that enables effective fine-tuning of pretrained models through adaptive frame selection for GA prediction. Our approach identifies the most informative and least redundant frames, enhancing both training efficiency and prediction accuracy. Validated on data collected from ultrasound devices in diverse resource environments, SelectGA improves gestational age prediction accuracy by 27% on mean absolute error metrics. These results demonstrate substantially improved generalizability, establishing foundations for sustainable AI adoption in prenatal care across resource-constrained settings. Code is available at: https://github.com/tanya-akumu/selectGA

Keywords: Gestational Age · Fetal Ultrasound · Blind sweep.

1 Introduction

Accurate estimation of gestational age (GA) is critical for monitoring fetal development and ensuring timely clinical interventions in antenatal care. Ultrasound imaging is the gold standard for GA prediction, usually estimated from standard planes defined by international guidelines [23]. Even so, the quality of the acquisition relies heavily on expert sonographers and high-end equipment, limiting its accessibility in low-resource settings. A recent scanning protocol, the blind

sweep protocol, offers a promising alternative for resource-constrained environments where clinical expertise is scarce [6]. This protocol consists of performing a set of pre-defined sweeps over the maternal abdomen without a real-time visualization, that can be easily performed by minimally trained healthcare workers. However, blind sweeps commonly result in a large number of uninformative or redundant frames and are unlikely to contain clear views of the anatomical standard planes [11], thus posing significant challenges for automated analysis. Therefore, we aim to develop an Artificial Intelligence (AI) driven framework that optimizes blind sweeps analysis for accurate and efficient GA prediction, thus addressing the limitations of existing methods.

While existing approaches for GA prediction from blind sweep ultrasound videos have made significant strides, notable limitations remain. *Arroyo et al* [1] as well as *Van Den Heuvel et al* [11] proposed methods that rely on manual segmentation of fetal structures in ultrasound sweeps, using AI to estimate fetal biometry from the individual segmentations and further using the estimated fetal biometry to estimate GA using the Hadlock formula [9]. While effective, this approach is labor-intensive and requires expert intervention, making it unsuitable for scalable applications. Processing all the frames from a video, which can contain hundreds to thousands of frames, as input to an AI model is computationally intensive and necessitates sampling techniques to reduce resource demands while maintaining model efficacy. *Pokaprakarn et al* [20] introduced a method that randomly samples frames from sweep videos to predict GA. Although their large dataset (n=109,806 videos) reduces the risk of missing informative frames, this randomness is suboptimal, in particular for smaller datasets, which would need much more iteration during training thereby causing the model to over-fit on the training samples. Conversely, *Gomes et al* [8] and *Lee et al* [15] employed uniform sampling of frames in the sweep videos which, while systematic, likely still includes frames irrelevant to GA prediction, especially in the presence of limited dataset sizes. Therefore, there is a need for a robust frame selection strategy that maximizes the information passed to the model, ensuring reliable GA prediction without relying on manual intervention or large-scale datasets. To the best of our knowledge, to date no publicly available methodology has addressed this challenge.

To this end, we propose SelectGA, a novel framework for adaptive frame selection and GA prediction from blind sweep ultrasound videos. SelectGA incorporates a pretrained object detector, identifying frames containing fetal structures, and further applies a selection algorithm to determine the most dissimilar representative frames for GA prediction. This approach ensures that only informative frames are used, addressing the challenges of redundancy and variability inherent in blind sweeps. Our contributions can be summarized as follows:

- We introduce a novel method for adaptive frame selection in blind sweep ultrasound videos thus maximizing the information for training.
- We implement a full GA prediction framework based on this method with data from diverse geographical centers.

– Our framework establishes a new benchmark for fetal ultrasound analysis, outperforming existing methods in the low data regime.

Table 1: Comprehensive dataset summary detailing the multi-country fetal ultrasound dataset used in this study.

Site	Patients	Scans	Videos	Frames	Resolution	Train	Val	Test
Spain center	114	114	871	157,173	735×975	543 vids	150 vids	178 vids
Kenya center	36	48	443	87,875	768×1024	244 vids	71 vids	128 vids

2 Methodology

2.1 Dataset

For this study, we collected a new fetal ultrasound dataset that consists of blind sweep ultrasound videos acquired from two centers, one from Barcelona, Spain (high resource), and the other from Rabai, Kenya (low-resource). Blind sweeps were performed using a standardized protocol involving a fixed number of vertical and horizontal freehand sweeps over the maternal abdomen. We used the Symphysio Fundal Height (SFH) [19] to determine the number of sweeps to collect from a patient. 6 (3 horizontal and 3 vertical) sweep videos were collected for a SFH of between 16 to 24 cm, while 10 (5 horizontal and 5 vertical) sweeps were collected for a SFH$\geq$25 cm. In the Spain center, the sweeps were conducted by a trained sonographer using the Philips Lumify (Philips, The Netherlands) ultrasound device, while in the rural Kenya center, the Voluson v8 (General Electric, USA) device was used. Each loop video was set to 10 seconds. The ground truth GA label was established from the first ultrasound scan before the 14th week of pregnancy based on the Crown Rump Length (CRL) [7]. The data totaled 162 study scans and 1,314 blind sweep videos. The data was split into training, validation, and test sets at patient level using a 60-20-20 ratio, ensuring no data leakage. Detailed statistics of the dataset are provided in Table 1.

2.2 Overall Framework

Figure 1 illustrates an overview of the proposed framework. It consists of two stages: 1)**Adaptive Frame Selection** and 2) **gestational age prediction**. In stage 1) we apply the **anatomically guided (AG)** selector to filter frames that contain fetal structures and use the **diversity-guided selector (DS)** to select the most diverse optimal frames. In stage 2), the chosen frames are fed to the model to predict the gestational age.

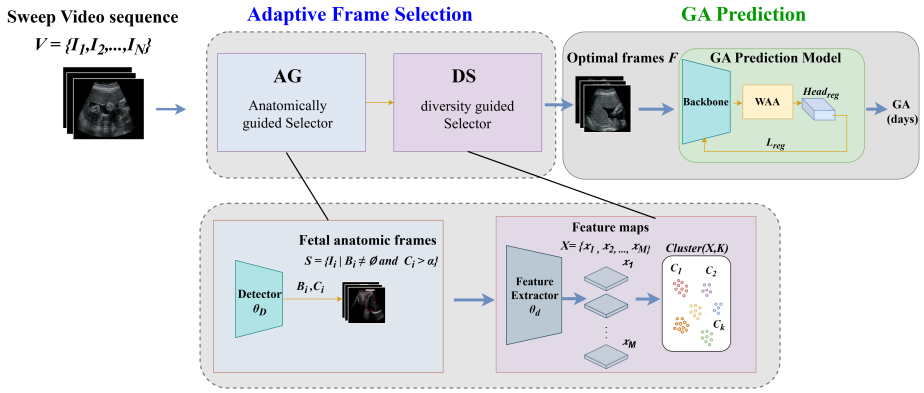


Fig. 1: Overview of our proposed framework which predicts gestational age through a sequential process of Adaptive Frame Selection followed by GA Prediction. The adaptive frame selection employs anatomically guided selector to filter frames containing anatomical features and an adaptive diversity guided selector to chose the most diverse frames based on clustering by Euclidean distance between feature embeddings. The chosen optimal frames are then processed through a ResNet-50 backbone with weighted attention module and regression head to produce final gestational age estimates.

2.3 SelectGA: Adaptive Optimal Frame Selection

The goal of our selection Algorithm 1 is to choose a subset of the most informative frames from an ultrasound sweep video V for GA prediction. Let $V = \{I_1, I_2, \ldots, I_N\}$ represent the ultrasound sweep video, where I_i denotes the i-th frame and N is the total number of frames. A pretrained object detector, θ_D, is applied to each frame I_i to detect fetal structures. This pretrained detector is based on Faster R-CNN [22] and fine-tuned on a small subset (500 images) of open-source fetal ultrasound standard plane images [4,24]. The detector outputs a set of bounding boxes B_i and a confidence score C_i for each frame. Frames with detected fetal structures and a confidence score above a predefined threshold α are retained. We set $\alpha = 0.25$ based on the lowest mean of the classes in our dataset. Formally, the set of selected frames S is defined as:

$$S = \{I_i \mid B_i \neq \emptyset \text{ and } C_i > \alpha\}, \tag{1}$$

where $B_i \neq \emptyset$ indicates the presence of fetal structures in frame I_i, and $C_i > \alpha$ ensures that only frames with high-confidence detections are included.

Feature Embedding Extraction Next, a pretrained feature extractor θ_d, consisting of only the hidden layers of θ_D, is used to compute feature embeddings for each frame in S. Let $M = |S|$ be the number of selected frames. The feature embeddings are represented as $X = \{x_1, x_2, \ldots, x_M\}$, where $x_j \in \mathbb{R}^d$ is the d-dimensional feature vector corresponding to the j-th frame in S.

Algorithm 1 Adaptive Frame Selection for Gestational Age Prediction

Require: Ultrasound sweep video $V = \{I_1, I_2, \ldots, I_N\}$, pretrained object detector θ_D, confidence threshold α, number of clusters K
Ensure: Selected informative frames F for GA prediction
$S \leftarrow \emptyset$ ▷ Set to store frames with detected fetal structures
for each frame $I_i \in V$ **do**
  $(B_i, C_i) \leftarrow \theta_D(I_i)$ ▷ Detect fetal structures, where B_i is the set of bounding boxes and C_i are a set of representative confidence scores
  if $B_i \neq \emptyset$ **and** $C_i > \alpha$ **then**
    $S \leftarrow S \cup \{I_i\}$ ▷ Retain frames with fetal structures and confidence above α
  end if
end for
Extract feature embeddings $X = \{x_1, x_2, \ldots, x_M\}$ from S using a pretrained feature extractor θ_d
Cluster X into K clusters: $\{C_1, C_2, \ldots, C_K\} \leftarrow$ K-means(X, K)
Select one representative frame f_k closest to the centroid of each cluster C_k to form $F = \{f_1, f_2, \ldots, f_K\}$
return F

Clustering and Representative Frame Selection To reduce redundancy and select the most informative frames, the feature embeddings X are clustered into K groups using the K-means algorithm [16]. Let $\{C_1, C_2, \ldots, C_K\}$ denote the resulting clusters, where each cluster C_k contains a subset of feature embeddings closest in Euclidean distance. From each cluster C_k, a single representative frame f_k is selected. The final set of informative frames $F = \{f_1, f_2, \ldots, f_K\}$ is given by sampling the frame corresponding to the feature embedding closest to the centroid of cluster C_k. The algorithm returns the set of the most dissimilar informative frames F, which can be used as input for gestational age prediction models. This approach ensures that the selected frames are both representative of the ultrasound sweep while having minimal overlap, thus allowing for efficient prediction in resource-limited settings.

2.4 Gestational Age Prediction Model

The GA prediction model consists of three main components: a ResNet-50 feature extractor [10], a Weighted Average Attention module (WAA) as implemented in [20], and a regression head. The ResNet-50 backbone is initialized with weights pretrained on the ImageNet dataset [5], providing a robust feature extraction capability. The ResNet-50 backbone processes each selected frame $f_k \in F$ to extract high-dimensional feature representations $\boldsymbol{x}_t \in \mathbb{R}^{2048}$, where t is the frame index. The WAA module [20], inspired by the additive Bahdanau attention mechanism [2], assigns a contribution score $\boldsymbol{w}_t$ to each frame based on its relevance to the task. The module consists of three trainable parameters: V, W, and Q. The attention weights $\boldsymbol{w}_t$ and scores $\boldsymbol{s}_t$ are computed as shown equation 2. Q reduces the dimensionality of $\boldsymbol{x}_t$ to produce the feature representation $\boldsymbol{a}$.

$$\boldsymbol{w}_t = \sigma\left(V\left(\tanh\left(W\boldsymbol{x}_t\right)\right)\right), \boldsymbol{s}_t \quad = \frac{\boldsymbol{w}_t}{\sum_{u=1}^{N} \boldsymbol{w}_u}, \boldsymbol{a} = \sum_{t=1}^{N} \boldsymbol{s}_t \cdot Q(\boldsymbol{x}_t), \tag{2}$$

Finally, a regression head consisting of a dense linear layer predicts the GA in days. We train the model using the L1 loss function, which measures the absolute difference between the predicted GA $\hat{y}$ and the ground truth y:

$$L_{reg} = \frac{1}{N} \sum_{i=1}^{N} |\hat{y}_i - y_i|. \tag{3}$$

3 Experiments and Results

Baselines As prior work [8,20] does not provide accessible model weights or datasets, we re-implement their methods on our dataset for fair comparison. We evaluate our framework against diverse baselines including image-based models (ResNet-50 [10,20], Universal Ultrasound Foundation Model [13]), medical imaging models (EchoNET [17] adapted from LVEF regression), and video understanding models (ViFi-CLIP [21], Qwen-VL [3]). All baselines are adapted by keeping the same backbone architecture and replacing final layers with our GA regression head as described in 2.4 and fine-tuning end-to-end.
Implementation Details Following the original implementations, we employ different sampling strategies for the baseline models. ResNet-50 and USFM use random sampling as in [20,13]. EchoNET, ViFi-CLIP, and Qwen-VL employ uniform sampling as per their original designs [17,21,3]. The baseline models are trained using the AdamW optimizer [14] with a batch size of 16 and a learning rate of 10^{-4}. We reduce the learning rate by 10x after every 45 iterations for a total of 200 iterations. To avoid over-fitting, we incorporate early stopping with patience of 5 iterations on the validation set. All implementations were done with the Pytorch framework [12] and the models were trained on a NVIDIA GeForce RTX 3090 24GB GPU. For all experiments, we sampled $K = 16$ frames per sweep video, resized the images to size (224x224) and applied data augmentation techniques such as cropping, contrast adjustment, brightness, rotation, and blur.
Evaluation Metrics We use the Mean Absolute Error (MAE), Root Mean Squared Error (RMSE), R^2 metrics to evaluate the performance on a held-out test set. Moreover, we assess the percentage proportion of study scans with absolute errors within clinically relevant thresholds (<7 days and <14 days) [18,23].

3.1 Quantitative and Qualitative Results

Table 2 presents the performance comparison of our proposed framework against the selected baselines. Our method outperforms other approaches across all evaluation metrics on the overall test set. SelectGA achieves the lowest MAE

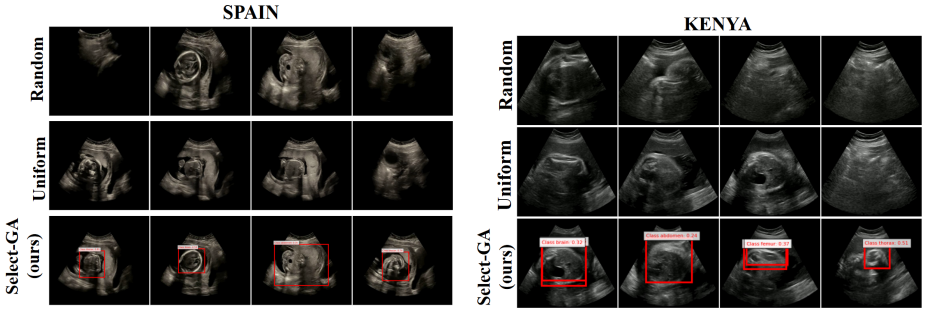
Table 2: Comparison of Model Performance on Test Set.

Method	MAE ↓	RMSE ↓	R^2 ↑	< 7d (%) ↑	< 14d (%) ↑
2nd trimester					
Resnet50 [20]	10.20 ± 2.29	14.08 ± 3.28	0.440	44.4	77.8
USFM [13]	59.04 ± 5.65	63.72 ± 5.85	−10.467	0.0	0.0
EchoNet [17]	9.53 ± 2.60	14.57 ± 4.50	0.400	55.6	77.8
ViFi-CLIP [21]	$\mathbf{5.22 \pm 1.09}$	$\mathbf{7.04 \pm 1.55}$	**0.881**	78.9	**94.7**
Qwen-VL [3]	11.81 ± 1.99	14.54 ± 1.89	0.403	38.9	55.6
SelectGA (ours)	5.89 ± 2.13	10.80 ± 3.50	0.671	**83.3**	83.3
3rd trimester					
Resnet50 [20]	15.60 ± 3.23	20.76 ± 3.28	0.003	38.9	55.6
USFM [13]	17.34 ± 3.34	22.40 ± 3.49	−0.161	27.8	50.0
EchoNet [17]	13.45 ± 2.47	17.05 ± 2.42	0.327	38.9	**61.1**
ViFi-CLIP [21]	20.46 ± 4.35	27.21 ± 2.42	−1.142	23.5	41.2
Qwen-VL [3]	27.65 ± 5.03	34.93 ± 5.22	−1.822	22.2	33.3
SelectGA (ours)	$\mathbf{13.32 \pm 2.38}$	$\mathbf{16.71 \pm 2.53}$	**0.354**	**44.4**	55.6
Spain Center (high-resource)					
Resnet50 [20]	14.84 ± 2.64	19.52 ± 2.79	0.839	34.8	56.5
USFM [13]	45.12 ± 5.14	51.41 ± 4.90	−0.116	4.3	13.0
EchoNet [17]	11.07 ± 2.11	15.01 ± 2.30	0.905	52.2	**73.9**
ViFi-CLIP [21]	15.42 ± 3.64	23.30 ± 4.69	0.771	47.8	60.9
Qwen-VL [3]	21.78 ± 4.32	30.08 ± 5.00	0.618	34.8	43.5
SelectGA (ours)	$\mathbf{10.18 \pm 2.19}$	$\mathbf{14.62 \pm 2.41}$	**0.910**	**60.9**	65.2
Kenya Center (low-resource)					
Resnet50 [20]	9.47 ± 2.87	14.04 ± 4.49	0.863	53.8	84.6
USFM [13]	25.93 ± 8.62	40.48 ± 12.01	−0.137	30.8	46.2
EchoNet [17]	12.23 ± 3.38	17.27 ± 5.11	0.793	38.5	61.5
ViFi-CLIP [21]	$\mathbf{7.09 \pm 1.50}$	$\mathbf{8.93 \pm 1.96}$	**0.945**	61.5	**84.6**
Qwen-VL [3]	16.10 ± 3.06	19.52 ± 3.40	0.736	23.1	46.2
SelectGA (ours)	8.59 ± 2.72	13.04 ± 3.90	0.882	**69.2**	76.9
Overall					
Resnet50 [20]	12.90 ± 2.03	17.73 ± 2.38	0.851	41.7	66.7
USFM [13]	38.19 ± 4.78	47.76 ± 5.06	−0.082	13.9	25.0
EchoNet [17]	11.49 ± 1.82	15.86 ± 2.45	0.881	47.2	69.4
ViFi-CLIP [21]	12.41 ± 2.48	19.38 ± 3.74	0.822	52.8	69.4
Qwen-VL [3]	19.73 ± 3.01	26.75 ± 3.79	0.661	30.6	44.4
SelectGA (ours)	$\mathbf{9.60 \pm 1.71}$	$\mathbf{14.07 \pm 2.07}$	**0.906**	**63.9**	**69.4**

Table 3: Ablation study of the proposed model framework. We assess the contribution of the Anatomically Guided (AG) selector and Diversity Guided (DG) Selector to the overall performance.

Components	AG	DS	MAE ↓	RMSE ↓	R^2 ↑	< 7d (%) ↑	< 14d (%) ↑
Resnet50	×	×	13.17 ± 2.12	18.31 ± 2.65	−0.061	44.4	66.7
w/ WAA	×	×	12.90 ± 2.03	17.73 ± 2.38	0.851	41.7	66.7
w/ WAA + AG	✓	×	10.96 ± 2.12	16.83 ± 3.10	0.866	55.5	**72.2**
SelectGA	✓	✓	$\mathbf{9.60 \pm 1.71}$	$\mathbf{14.07 \pm 2.07}$	**0.906**	**63.9**	69.4

(9.60 days) and RMSE (14.07 days) while maintaining the highest R^2 value (0.906), representing a 16.4% improvement in MAE over the next best performer, EchoNet (11.49 days). Notably, SelectGA shows consistent performance with 63.9% of predictions within a 7 days error – a 21% relative improvement over ViFi-CLIP's 52.8%. When evaluated across trimesters and geographical centers, SelectGA reveals its robustness in adapting to varying clinical conditions. In the more challenging 3rd trimester predictions, SelectGA outperforms all competitors with an MAE of 13.32 days. In terms of geographical performance, SelectGA displays cross-center generalization, achieving the best performance at the Spain center (10.18 days MAE, R^2=0.910) while remaining competitive at the Kenya center with just slightly higher error than the best performer ViFi-CLIP (8.59 vs. 7.09 days MAE). The qualitative results reported in Figure 2 demonstrate the effectiveness of our method in identifying and prioritizing more frames from the blind sweep videos that contain the fetal anatomical structures that are relevant for GA prediction.



(a) Spain Center (b) Kenya Center

Fig. 2: Qualitative results of our SelectGA method on blind sweep videos across two centers.

Ablation Study We perform an ablation study of the different components of our framework as highlighted in Table 3. The integration of **AG** (Anatomically guided selector) which filters frames that contain fetal anatomical structures, significantly boosts performance by 16%, particularly in the proportion of predictions within 14 days (72.2%). Our full SelectGA framework, which incorporates **DS** (diversity guided selection of optimal frames based on dissimilarity), significantly boosts the baseline MAE performance by **27.1**%. This highlights the effectiveness of maximizing the information passed through the model for the GA prediction task, while preventing the selection of uninformative frames. We also assessed the effect of the clustering initialization as a measure of the model's uncertainty for 5 runs. The resulting variance for these runs were $[2.9618, 2.8493, 2.9618, 2.8866, 2.9344]$ with a Spearman correlation of 0.307 between the MAE and the variance.

4 Conclusion

We propose SelectGA, a framework that locates important anatomical regions in frames of blind sweep ultrasound videos and increases the diversity of the chosen subsets.This maximizes the information the model needs for accurate prediction in limited data settings and reduces the redundancy inherent in blind sweep ultrasound. Through extensive experiments, our framework achieved clinically relevant predictions, outperforming existing strategies by 27% on the MAE. Moreover, our framework exhibited competitive performance under varying clinical conditions. Notably, the pretrained Universal Ultrasound Foundation Model (USFM), showed particularly low performance. Future work could develop better generalizable models for automated fetal analysis. Moreover, initial experiments of uncertainty estimation showed that a correlation exists between results for different clustering runs, which will be further investigated in future work. In conclusion, our framework addresses key challenges in blind sweep ultrasound analysis that is practical for low-resource settings.

Acknowledgments. This study was funded by the European Research Council (ERC) under the European Union's Horizon Europe research and innovation programme (AIMIX project - Grant Agreement No. 101044779).

Disclosure of Interests. The authors have no competing interests to declare that are relevant to the content of this article.

References

1. Arroyo, J., Marini, T.J., Saavedra, A.C., Toscano, M., Baran, T.M., Drennan, K., Dozier, A., Zhao, Y.T., Egoavil, M., Tamayo, L., et al.: No sonographer, no radiologist: New system for automatic prenatal detection of fetal biometry, fetal presentation, and placental location. PloS one **17**(2), e0262107 (2022)
2. Bahdanau, D., Chorowski, J., Serdyuk, D., Brakel, P., Bengio, Y.: End-to-end attention-based large vocabulary speech recognition. In: 2016 IEEE international conference on acoustics, speech and signal processing (ICASSP). pp. 4945–4949. IEEE (2016)
3. Bai, J., Bai, S., Chu, Y., Cui, Z., Dang, K., Deng, X., Fan, Y., Ge, W., Han, Y., Huang, F., et al.: Qwen technical report. arXiv preprint arXiv:2309.16609 (2023)
4. Burgos-Artizzu, X.P., Coronado-Gutiérrez, D., Valenzuela-Alcaraz, B., Bonet-Carne, E., Eixarch, E., Crispi, F., Gratacós, E.: Evaluation of deep convolutional neural networks for automatic classification of common maternal fetal ultrasound planes. Scientific Reports **10**(1), 10200 (2020)
5. Deng, J., Dong, W., Socher, R., Li, L.J., Li, K., Fei-Fei, L.: Imagenet: A large-scale hierarchical image database. In: 2009 IEEE conference on computer vision and pattern recognition. pp. 248–255. Ieee (2009)
6. DeStigter, K.K., Morey, G.E., Garra, B.S., Rielly, M.R., Anderson, M.E., Kawooya, M.G., Matovu, A., Miele, F.R.: Low-cost teleradiology for rural ultrasound. In: 2011 IEEE Global Humanitarian Technology Conference. pp. 290–295 (2011). https://doi.org/10.1109/GHTC.2011.39

7. Drumm, J., Clinch, J., MacKenzie, G.: The ultrasonic measurement of fetal crown-rump length as a method of assessing gestational age. BJOG: An International Journal of Obstetrics & Gynaecology **83**(6), 417–421 (1976)
8. Gomes, R.G., Vwalika, B., Lee, C., Willis, A., Sieniek, M., Price, J.T., Chen, C., Kasaro, M.P., Taylor, J.A., Stringer, E.M., et al.: Ai system for fetal ultrasound in low-resource settings. arXiv preprint arXiv:2203.10139 (2022)
9. Hadlock, F.P., Deter, R.L., Harrist, R.B., Park, S.: Estimating fetal age: computer-assisted analysis of multiple fetal growth parameters. Radiology **152**(2), 497–501 (1984)
10. He, K., Zhang, X., Ren, S., Sun, J.: Deep residual learning for image recognition. In: Proceedings of the IEEE conference on computer vision and pattern recognition. pp. 770–778 (2016)
11. van den Heuvel, T.L., Petros, H., Santini, S., de Korte, C.L., van Ginneken, B.: Automated fetal head detection and circumference estimation from free-hand ultrasound sweeps using deep learning in resource-limited countries. Ultrasound in medicine & biology **45**(3), 773–785 (2019)
12. Imambi, S., Prakash, K.B., Kanagachidambaresan, G.: Pytorch. Programming with TensorFlow: solution for edge computing applications pp. 87–104 (2021)
13. Jiao, J., Zhou, J., Li, X., Xia, M., Huang, Y., Huang, L., Wang, N., Zhang, X., Zhou, S., Wang, Y., et al.: Usfm: A universal ultrasound foundation model generalized to tasks and organs towards label efficient image analysis. Medical Image Analysis **96**, 103202 (2024)
14. Kingma, D.P., Ba, J.: Adam: A method for stochastic optimization. arXiv preprint arXiv:1412.6980 (2014)
15. Lee, C., Willis, A., Chen, C., Sieniek, M., Watters, A., Stetson, B., Uddin, A., Wong, J., Pilgrim, R., Chou, K., et al.: Development of a machine learning model for sonographic assessment of gestational age. JAMA network open **6**(1), e2248685–e2248685 (2023)
16. Likas, A., Vlassis, N., Verbeek, J.J.: The global k-means clustering algorithm. Pattern recognition **36**(2), 451–461 (2003)
17. Ouyang, D., He, B., Ghorbani, A., Lungren, M.P., Ashley, E.A., Liang, D.H., Zou, J.Y.: Echonet-dynamic: a large new cardiac motion video data resource for medical machine learning. In: NeurIPS ML4H Workshop: Vancouver, BC, Canada. vol. 5 (2019)
18. Papageorghiou, A.T., Kemp, B., Stones, W., Ohuma, E.O., Kennedy, S.H., Purwar, M., Salomon, L.J., Altman, D.G., Noble, J.A., Bertino, E., et al.: Ultrasound-based gestational-age estimation in late pregnancy. Ultrasound in Obstetrics & Gynecology **48**(6), 719–726 (2016)
19. Papageorghiou, A.T., Ohuma, E.O., Gravett, M.G., Hirst, J., Da Silveira, M.F., Lambert, A., Carvalho, M., Jaffer, Y.A., Altman, D.G., Noble, J.A., et al.: International standards for symphysis-fundal height based on serial measurements from the fetal growth longitudinal study of the intergrowth-21st project: prospective cohort study in eight countries. bmj **355** (2016)
20. Pokaprakarn, T., Prieto, J.C., Price, J.T., Kasaro, M.P., Sindano, N., Shah, H.R., Peterson, M., Akapelwa, M.M., Kapilya, F.M., Sebastião, Y.V., et al.: Ai estimation of gestational age from blind ultrasound sweeps in low-resource settings. NEJM evidence **1**(5), EVIDoa2100058 (2022)
21. Rasheed, H., Khattak, M.U., Maaz, M., Khan, S., Khan, F.S.: Fine-tuned clip models are efficient video learners. In: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. pp. 6545–6554 (2023)

22. Ren, S., He, K., Girshick, R., Sun, J.: Faster r-cnn: Towards real-time object detection with region proposal networks. IEEE transactions on pattern analysis and machine intelligence **39**(6), 1137–1149 (2016)
23. Salomon, L., Alfirevic, Z., Da Silva Costa, F., Deter, R., Figueras, F., Ghi, T.a., Glanc, P., Khalil, A., Lee, W., Napolitano, R., et al.: Isuog practice guidelines: ultrasound assessment of fetal biometry and growth. Ultrasound in obstetrics & gynecology **53**(6), 715–723 (2019)
24. Sendra-Balcells, C., Campello, V.M., Torrents-Barrena, J., Ahmed, Y.A., Elattar, M., Ohene-Botwe, B., Nyangulu, P., Stones, W., Ammar, M., Benamer, L.N., et al.: Generalisability of fetal ultrasound deep learning models to low-resource imaging settings in five african countries. Scientific reports **13**(1), 2728 (2023)