

Foundation Model-driven Key Anatomy Frame Selection for Blind-sweep Ultrasound Fetal Birth Weight Estimation

Le Ou^{1,2*}, Xiliang Zhu^{3*}, Huanwen Liang^{1,2}, Wenxiong Pan³,
Yuhao Huang^{1,4}, Yuxiang Deng³, Xuan Sheng^{3,5}, Hong Yin⁵, Juhua Xiao⁶,
Xin Zhou^{6(✉)}, and Dong Ni^{1,3,7(✉)}

¹ Medical Ultrasound Image Computing (MUSIC) Lab, Shenzhen University, Shenzhen, China
nidong@szu.edu.cn

² School of Biomedical Engineering, Shenzhen University, Shenzhen, China

³ School of Biomedical Engineering and Informatics, Nanjing Medical University, Nanjing, China

⁴ Centre for Artificial Intelligence and Robotics, Hong Kong Institute of Science & Innovation, Chinese Academy of Sciences, Hong Kong, China

⁵ Shandong Provincial Maternal and Child Health Care Hospital Affiliated to Qingdao University, Jinan, China

⁶ Jiangxi Maternal and Child Health Hospital, Nanchang, China
jxsfbcsk@163.com

⁷ School of Artificial Intelligence, Shenzhen University, Shenzhen, China

Abstract. Accurate fetal birth weight (FBW) estimation shortly before delivery is clinically valuable yet challenging due to its reliance on operator expertise, particularly in low-resource settings. To reduce this reliance, we study near-term birth-weight regression from blind-sweep ultrasound (US) videos acquired within 48 hours prior to delivery, with post-delivery weighing as ground truth. Accordingly, we propose a foundation model-driven key anatomy frame selection framework that enables accurate FBW regression despite the absence of plane constraints in blind sweeps. Our highlights are as follows: (1) We believe this is the first work to estimate FBW using blind-sweep US videos, enabling operator-independent assessment. (2) An Anatomy-Guided Frame Selection module equipped with a vision-language foundation model is proposed for keyframe collection in unconstrained sweeps. (3) A Redundancy-Aware Feature Compression module is designed to compress frame features while preserving task-relevant information, alleviating temporal redundancy. Extensively validated on prospectively collected data from 839 patients, our method achieves an MAE of 161.3 g, with 90.23% and 100% of cases falling within 10% and 15% absolute percentage error, outperforming typical Hadlock estimation and strong competitors. Codes are available at <https://github.com/ouleoule/BlindSweep-EBW>.

Keywords: Blind Sweep Ultrasound· Fetal Birth Weight· Foundation Model

* Le Ou and Xiliang Zhu contributed equally.

1 Introduction

Accurate estimation of fetal birth weight (FBW) prior to delivery is crucial for determining the mode of delivery and managing perinatal risks. Ultrasound (US) remains the primary modality for estimation of fetal weight in clinical practice [3]. However, in real-world clinical settings, accurate FBW estimation requires obtaining standard planes and performing precise biometric measurements [14]. This process is highly dependent on operator expertise. Such dependency is particularly problematic in low-resource setting, where the shortage of experienced sonographers and the high demand for obstetric services further exacerbate assessment inaccuracies and potential clinical risks.

Recent artificial intelligence (AI) has advanced rapidly in US analysis [21], demonstrating strong performance in plane localization, segmentation, and biometric measurement [20,4,17,5,12]. Plotka et al. [10,9] proposed a deep learning framework to extract key features from fetal abdominal images for real-time weight estimation. Wang et al. [19] utilized 3D ultrasound data to capture richer anatomical information, significantly improving accuracy under challenging fetal postures. These studies have demonstrated strong performance in fetal weight (FW) estimation. However, they still rely heavily on high-quality, standardized input data. When operators have limited experience, suboptimal image acquisition can lead to substantial degradation in prediction accuracy. Hence, current methods have not fundamentally addressed the issue of operator dependency.

This tight coupling between performance and standardized operation limits the scalability of AI in primary care. To address this challenge, Stringer et al. [2,11,18,16] proposed replacing the conventional plane-based paradigm with blind-sweep US, which acquires data through continuous scanning along predefined trajectories without requiring real-time identification of standard anatomical structures. Tanya Akumu et al. [1] proposed a clustering-based strategy to adaptively select informative frames from blind-sweep videos, effectively reducing redundancy and enhancing the robustness of gestational age estimation. These studies have demonstrated that blind-sweep US can reduce dependence on operator expertise while maintaining competitive performance.

In this study, we propose a foundation model-driven key anatomy frame selection framework for blind-sweep US FBW estimation. Our contributions are threefold. (1) To the best of our knowledge, this is the first study to apply the blind-sweep paradigm to FBW estimation, effectively decoupling accuracy from the need for expert-level image acquisition. (2) We propose an Anatomy-Guided Frame Selection (AFS) module that leverages a foundation model and anatomical priors to implicitly localize key structures within unconstrained blind-sweep sequences. (3) We introduce a Redundancy-Aware Feature Compression (RAFC) module that compresses frame features while enhancing task-relevant information, mitigating the impact of temporal redundancy in blind-sweep videos. We prospectively collected a dataset of 870 exams from 839 patients. Experimental results demonstrate the effectiveness of our method, which reduces operator dependency without requiring high-quality input data, providing reliable FBW estimation in resource-constrained settings.

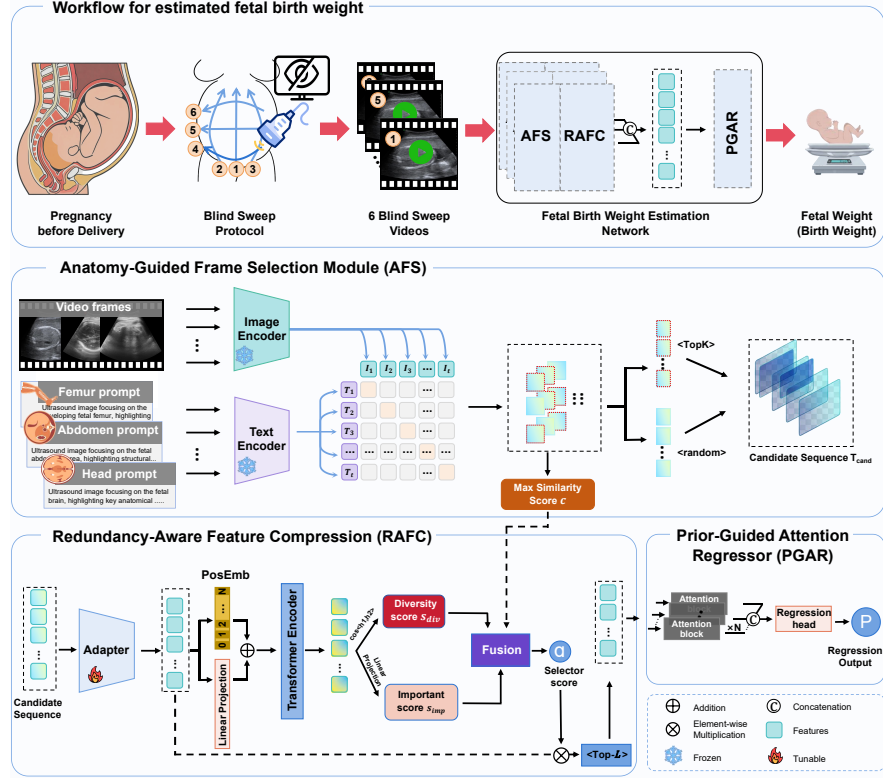


Fig. 1: The overview of our proposed framework.

2 Methodology

Blind-sweep US videos exhibit variable temporal lengths, rapid viewpoint transitions, sparse standard anatomical planes, and substantial acquisition noise. Directly regressing FBW from all frames introduces significant redundancy and may dilute discriminative anatomical cues. To address this challenge, we propose a hierarchical anatomy-guided frame selection framework for blind-sweep FBW estimation (Fig. 1). Without relying on explicit standard plane detection, our method progressively refines informative evidence through two selection stages, followed by prior-guided regression. Specifically, the framework consists of: (1) an AFS module that constructs a coarse informative frame subset using anatomical priors; (2) a RAFC module that suppresses temporally redundant frames and enhances task-relevant representations; and (3) a Prior-Guided Attention Regressor (PGAR) integrating refined frame features for FBW estimation.

2.1 Anatomy-Guided Frame Selection

Blind-sweep US videos inherently contain noise-dominant or low-quality frames, which may obscure discriminative anatomical structures and degrade regression performance. To preserve frames with vital anatomical content, we propose an AFS module that leverages anatomical priors to identify informative frames.

We adopt FetalCLIP [7] as the backbone encoder due to its strong cross-modal representation capability learned from large-scale fetal image-text alignment. It consists of an image encoder $\mathcal{E}_{img}$ and a text encoder $\mathcal{E}_{txt}$ that project visual frames and anatomical concept descriptions into a shared embedding space.

Given a blind-sweep video $V = \{v_1, \dots, v_N\}$ and a anatomical concepts set $T = \{t_1, \dots, t_M\}$, where each concept t_M is described using the predefined and fixed text prompts provided in the public FetalCLIP repository. we extract embeddings for the videos and concepts set as:

$$X = \mathcal{E}_{img}(V), \quad (1)$$

$$P = \mathcal{E}_{txt}(T), \quad (2)$$

where N represents the number of video frames and M represents the number of concepts. The anatomical relevance of each frame is quantified by the cosine similarity between frame and concept embeddings $\mathbf{A} = \cos(X, P)$. We define the anatomical confidence of each frame as the maximum similarity over all concepts:

$$c = \max_j \mathbf{A}_{:,j} \in \mathbb{R}^N. \quad (3)$$

Frames with higher confidence are more likely to contain informative anatomical structures. The top- K frames are selected according to c to form the coarse informative frame subset:

$$\mathcal{I} = \text{TopK}(c), \quad X_c = X[\mathcal{I}, :]. \quad (4)$$

This selection effectively filters out noise-dominant frames while preserving structurally meaningful evidence for subsequent redundancy-aware refinement and regression.

2.2 Redundancy-Aware Feature Compression

Blind-sweep videos often contain many temporally adjacent frames with high visual similarity, which can dilute the model’s attention and reduce the discriminative power of frame representations. To address this, we propose a Redundancy-Aware Feature Compression (RAFC) module that compresses frame features while enhancing task-relevant information.

Given X_c obtained from the AFS module, we first adapt the features to the fetal weight estimation task using a lightweight adapter:

$$\tilde{X} = \text{Adapter}(X_c). \quad (5)$$

Next, a transformer layer is applied to model inter-frame relationships and learn frame-level importance: $H = \text{Transformer}(\tilde{X})$, H is the $K \times D$ feature matrix. To explicitly estimate the importance of each frame, we compute a task-specific importance score via a linear mapping followed by a sigmoid activation σ :

$$s_{\text{imp}} = \sigma(W_{\text{imp}}H), s_{\text{imp}} \in \mathbb{R}^K \quad (6)$$

To account for temporal redundancy after AFS candidate selection, we compute a local diversity score for each candidate frame within a sliding window of length w :

$$d_i = 1 - \max_{j \in \mathcal{W}_i} \cos(h_i, h_j), \quad i = 1, \dots, K, \quad (7)$$

where $\mathcal{W}_i$ denotes the set of neighboring frames within the window centered at frame i . Frames that are highly similar to their neighbors have lower d_i , indicating higher redundancy. For the entire video, we denote local diversity scores as $s_{\text{div}} = [d_1, d_2, \dots, d_K] \in \mathbb{R}^K$.

Finally, we adaptively fuse c from AFS, frame importance s_{imp} , and redundancy s_{div} to obtain a selector score for each frame:

$$s_i = \sigma\left(W_g[s_{\text{imp},i}, s_{\text{div},i}, c_i]\right), \quad i = 1, \dots, K, \quad (8)$$

The top- L frames with the highest selector scores are then retained as the fine-grained frame subset H_f for subsequent regression.

2.3 Prior-Guided Attention Regressor

Given the fine-grained frame features H_f retained after redundancy-aware compression, along with their corresponding frame scores a , we perform prior-guided feature aggregation for final fetal weight estimation.

The frame scores are first converted into a normalized prior distribution:

$$p_i = \text{softmax}\left(\frac{a_i}{\tau}\right), \quad (9)$$

where τ is a temperature parameter controlling distribution sharpness. Instead of directly applying attention on the original features, we first re-weight the frame representations using the prior distribution: $\tilde{H}_f = p \times H_f$, where $\times$ denotes row-wise scaling of frame features.

The re-weighted features are then fed into an attention module to model inter-frame dependencies:

$$H_{\text{att}} = \text{Attention}(\tilde{H}_f). \quad (10)$$

Finally, the global average pooling will be applied to obtain a compact representation z , which is then passed through a regression head to predict the FBW, with the formula as: $\hat{y} = \mathcal{R}(z)$. This prior-guided re-weighting strategy can ensure that anatomically relevant and non-redundant frames contribute more prominently during relational modeling and regression.

Table 1: Dataset statistics and split protocol.

Patients	Exams	Videos	Frames/video	Resolution	Train/Val/Test (exams)
839	870	5220	137–588	224×224	609 / 87 / 174

3 Experiments and Results

Datasets and Implementations. With the Institutional Review Board (IRB) approval, we prospectively collected 870 ultrasound exams from 839 pregnancies, including cases with up to two eligible scans per pregnancy. All scans were acquired by 10 trained sonographers and performed within 48 hours prior to delivery using a GE LOGIQ P9 system. A standardized blind-sweep protocol was adopted, consisting of a fixed number of vertical and horizontal freehand sweeps across the maternal abdomen, yielding six blind-sweep loop video per exam. Each loop video was set to 10 seconds. Prior to delivery, the estimated fetal weight (EFW) was determined by a senior doctor using the Hadlock formula based on manual biometric measurements. Post delivery, the actual birth weight (FBW) was recorded as the gold standard for this study. The mean FBW in this study was 3282.47 ± 400.27 g, ranging from 1410 g to 4360 g. The dataset was split into training, validation, and testing sets at the pregnancy level with a ratio of 7:1:2 to avoid data leakage. Detailed statistics of the dataset are provided in Table 1.

Prior to being fed into the network, all ultrasound frames were resized to 224×224 pixels. In the hierarchical selection process, the number of coarse candidate frames K in the AFS module was set to 100. Subsequently, the PAFC module selected $L = 50$ fine-grained frames from this candidate pool, with the sliding-window length set to $w = 10$. For each exam, the six blind-sweep loops were processed independently, and the L frame features selected from each loop were concatenated into $6L$ features for a single FBW prediction. These selection hyperparameters were determined through validation-set sensitivity analyses. To enhance the model’s generalizability and robustness, online data augmentation was applied during training, including random rotations, horizontal/vertical flipping, and adjustments to contrast and brightness. We trained the model for 100 epochs with a batch size of 4 using the Adam optimizer with an initial learning rate of $5e^{-4}$, which followed a cosine annealing decay schedule. To prevent overfitting, an early stopping strategy was incorporated with a patience of 5 epochs based on the validation loss. All experiments were conducted using PyTorch, and the models were trained on an NVIDIA H20 GPU with 141 GB memory.

For evaluation metrics, we used Mean Absolute Error (MAE) and Mean Absolute Percentage Error (MAPE) to evaluate the performance. In addition, we calculated the percentage of cases with absolute percentage error within the clinically relevant thresholds of 10% and 15% [14] (denoted as PC10 and PC15).

Quantitative Analysis. We comprehensively compared our approach with several competitor methods, including the clinical Hadlock-4 [3] (based on senior sonographers’ manual biometry), INTERGROWTH-21st [15], the blind-sweep-

Table 2: Test-set performance for near-term blind-sweep birth weight estimation.

Method	MAE (g)	MAPE (%)	$\leq 10\%$ (%)	$\leq 15\%$ (%)	Std. (%)
ResNet50+WAA[11]	309.50	9.15	58.15	79.11	7.09
SelectGA [1]	287.03	8.49	65.69	84.52	6.87
USFM [6]	247.26	6.67	79.60	91.54	6.25
EchoNet [8]	295.72	8.90	69.75	82.01	6.79
ViFi-CLIP [13]	284.94	8.62	63.79	84.48	6.57
INTERGROWTH-21st [15]	216.63	6.48	77.59	94.83	4.34
Hadlock-4 [3]	183.57	5.60	82.18	97.13	4.13
Ours	161.34	4.89	90.23	100.0	3.47

Table 3: Ablation results of the proposed components. $\downarrow$ ($\uparrow$) indicates lower (higher) values are preferred.

ASF	RAFC	MAE (g) $\downarrow$	Std. (%) $\downarrow$	$\leq 10\%$ (%) $\uparrow$	$\leq 15\%$ (%) $\uparrow$	MAPE (%) $\downarrow$
$\times$	$\times$	253.04	5.91	72.41	89.08	6.82
$\checkmark$	$\times$	205.49	4.66	80.02	94.17	6.29
$\times$	$\checkmark$	183.54	4.96	83.62	93.79	5.73
$\checkmark$	$\checkmark$	161.34	3.47	90.23	100.00	4.89

specific SelectGA [1] (re-purposed for FBW estimation), and several deep learning architectures, including USFM [6], EchoNet [8], and ViFi-CLIP [13]. For a fair comparison, all deep learning baselines used their original backbones with their final layers adapted to our regression head for task-specific optimization.

As shown in Table 2, our method achieved the best performance with an MAE of 161.34 g, a MAPE of 4.89%, and PC10/PC15 values of 90.23% and 100%, respectively. Notably, it slightly outperformed the clinical Hadlock-4 benchmark, demonstrating its competitive advantage. Although USFM is pre-trained on a large number of standard ultrasound planes and shows advantages in related tasks, it fails to reach the Hadlock baseline in our blind-sweep scenario, where standard planes are scarce. Given that a PC10 of 80% and PC15 of 95% are generally considered thresholds for clinical acceptance, our method is the only one to meet these practical requirements, highlighting its potential for real-world deployment. Ablation experiments were conducted to quantify the contribution of each module, as shown in Table 3. ASF contributed to a 18.8% improvement in MAE and an 10.5% increase in PC10. RAFC contributed to a 27% improvement in MAE and an 11.2% increase in PC10. Combining RAFC with ASF further boosted performance by an additional 8.7% in MAE and 6.6% in PC10. These results demonstrate the effectiveness of our proposed modules.

Qualitative Analysis. We visualize several representative cases in Fig. 2. The results reveal that our framework effectively selects structure-rich frames. Frames selected by the first-level AFS module contain key fetal anatomical structures, including the head and abdomen, while non-informative or noisy frames are filtered out. After the second-level RAFC module performs frame compression, the remaining frames preserve complete anatomical information, and low-

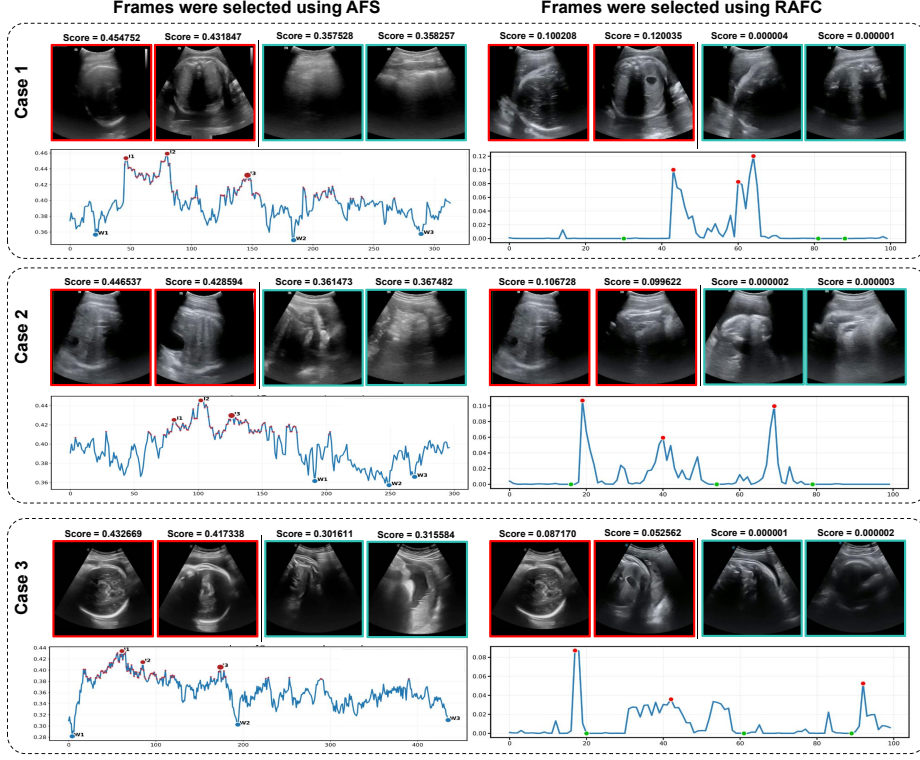


Fig. 2: **Visualization of frame selection for 3 representative cases using the AFS and RAFC modules.** For each case, the line plot shows the selection score over the frame sequence (x-axis: frame index, y-axis: selection score). The top images display the video frames corresponding to the highest (red boxes) and lowest (green boxes) scores selected by each method.

quality images are discarded. These results demonstrate the effectiveness of our proposed hierarchical anatomy-guided frame selection strategy.

4 Conclusion

In this study, we propose a foundation model-driven key anatomical frame selection framework to estimate near-term FBW from blind-sweep US videos. Our method employs a hierarchical anatomy-guided frame selection strategy to obtain more accurate FBW predictions. (1) We introduce an AFS module that leverages anatomical priors to implicitly select informative frames and construct a coarse-level subset, filtering out non-informative frames. (2) We propose a RAFC module that compresses temporally redundant frames to form a fine-grained subset, enhancing task-relevant representations. (3) We propose a PGAR to integrate frame features for accurate FBW estimation. Extensive experiments

show that our approach effectively reduces operator dependency and provides accurate FBW estimation in resource-constrained settings, with the potential to markedly improve prenatal care accessibility and quality in low-resource areas. In future work, we will perform comprehensive multi-center clinical validation and prospective evaluation.

Acknowledgments. This work was supported by the Frontier Technology Development Program of Jiangsu Province (No. BF2024078), National Natural Science Foundation of China (No. 12326619, 82260307), Science and Technology Planning Project of Guangdong Province (No. 2023A0505020002), Jiangxi Provincial Health Technology Project (No. 202610694, 202510072), Shandong Medical and Health Science and Technology Project (No. 202309020945, 202309020926), China Maternal and Child Health Association Research Project (No. 2023CAMCHS003A12, 2023CAMCHS003A05).

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

References

1. Akumu, T., Elbatel, M., Campello, V.M., Osuala, R., Martin-Isla, C., Valenzuela, I., Li, X., Khanal, B., Lekadir, K.: Adaptive frame selection for gestational age estimation from blind sweep fetal ultrasound videos. In: International Conference on Medical Image Computing and Computer-Assisted Intervention. pp. 3–12. Springer (2025)
2. 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.: A mobile-optimized artificial intelligence system for gestational age and fetal malpresentation assessment. *Communications medicine* **2**(1), 128 (2022)
3. Hadlock, F.P., Harrist, R., Sharman, R.S., Deter, R.L., Park, S.K.: Estimation of fetal weight with the use of head, body, and femur measurements—a prospective study. *American journal of obstetrics and gynecology* **151**(3), 333–337 (1985)
4. Huang, Y., Xu, Y., Dou, H., Deng, J., Yang, X., Zheng, H., Ni, D.: Uncertainty-aware diffusion and reinforcement learning for joint plane localization and anomaly diagnosis in 3d ultrasound. In: International Conference on Medical Image Computing and Computer-Assisted Intervention. pp. 650–660. Springer (2025)
5. Huang, Y., Yang, X., Huang, X., Liang, J., Zhou, X., Chen, C., Dou, H., Hu, X., Cao, Y., Ni, D.: Online reflective learning for robust medical image segmentation. In: International Conference on Medical Image Computing and Computer-Assisted Intervention. pp. 652–662. Springer (2022)
6. 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)
7. Maani, F., Saeed, N., Saleem, T.J., Farooq, Z., Alasmawi, H., Diehl, W., Mohammad, A., Waring, G., Valappil, S., Bricker, L., et al.: Fetalclip: A visual-language foundation model for fetal ultrasound image analysis. *npj Digital Medicine* (2026)
8. 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, p. 2 (2019)

9. Płotka, S., Pustelnik, K., Szenejko, P., Żebrowska, K., Rzucidło-Szymańska, I., Szymecka-Samaha, N., Łęgowik, T., Kosińska-Kaczyńska, K., Korzeniowski, P., Biliński, P., et al.: Direct estimation of fetal biometry measurements from ultrasound video scans through deep learning. *American Journal of Obstetrics & Gynecology MFM* **7**(4) (2025)
10. Płotka, S.S., Grzeszczyk, M.K., Szenejko, P.I., Żebrowska, K., Szymecka-Samaha, N.A., Łęgowik, T., Lipa, M.A., Kosińska-Kaczyńska, K., Brawura-Biskupski-Samaha, R., Işgum, I., et al.: Deep learning for estimation of fetal weight throughout the pregnancy from fetal abdominal ultrasound. *American journal of obstetrics & gynecology MFM* **5**(12), 101182 (2023)
11. 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)
12. Ramirez Zegarra, R., Ghi, T.: Use of artificial intelligence and deep learning in fetal ultrasound imaging. *Ultrasound in Obstetrics & Gynecology* **62**(2), 185–194 (2023)
13. 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)
14. 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)
15. Stirnemann, J., Villar, J., Salomon, L., Ohuma, E., Ruyan, P., Altman, D., Nosten, F., Craik, R., Munim, S., Cheikh Ismail, L., et al.: International estimated fetal weight standards of the intergrowth-21st project. *Ultrasound in Obstetrics & Gynecology* **49**(4), 478–486 (2017)
16. Stringer, J.S., Pokaprakarn, T., Prieto, J.C., Vwalika, B., Chari, S.V., Sindano, N., Freeman, B.L., Sikapande, B., Davis, N.M., Sebastião, Y.V., et al.: Diagnostic accuracy of an integrated ai tool to estimate gestational age from blind ultrasound sweeps. *Jama* **332**(8), 649–657 (2024)
17. Venturini, L., Budd, S., Farruggia, A., Wright, R., Matthew, J., Day, T.G., Kainz, B., Razavi, R., Hajnal, J.V.: Whole examination ai estimation of fetal biometrics from 20-week ultrasound scans. *NPJ Digital Medicine* **8**(1), 22 (2025)
18. Viswanathan, A.V., Pokaprakarn, T., Kasaro, M.P., Shah, H.R., Prieto, J.C., Benabdelkader, C., Sebastião, Y.V., Sindano, N., Stringer, E., Stringer, J.S.: Deep learning to estimate gestational age from fly-to cineloop videos: A novel approach to ultrasound quality control. *International Journal of Gynecology & Obstetrics* **165**(3), 1013–1021 (2024)
19. Wang, J., Ni, Q., Yu, H., Yao, R., Ying, J., Zhang, B., Yang, X., Peng, J., Chen, J., Yu, J., et al.: Accurate and efficient fetal birth weight estimation from 3d ultrasound (2025)
20. Yang, X., Huang, Y., Huang, R., Dou, H., Li, R., Qian, J., Huang, X., Shi, W., Chen, C., Zhang, Y., et al.: Searching collaborative agents for multi-plane localization in 3d ultrasound. *Medical Image Analysis* **72**, 102119 (2021)
21. Zhang, Y., Huang, Y., Dou, H., Zhu, X., Ling, C., Yang, Z., Liang, L., Li, J., Liang, S., Li, R., et al.: Artificial intelligence for detecting fetal orofacial clefts and advancing medical education. *Nature Communications* (2026)