RJNMSR
Review Journal of Neurological & Medical Sciences Review

Volume 4, Issue 3, 2026

EVALUATING THE ACCURACY AND CLINICAL UTILITY OF AI ALGORITHMS FOR GESTATIONAL AGE PREDICTION: A COMPREHENSIVE SYSTEMATIC REVIEW

[1]Aimen Ali, [2]Syeda Saliha Shah, [3]Aysha Ayub, [4]Rabia Kamal, [5]Hilal Ahmad Malik, [*6]Shamia Kamal

[1]Department of Radiological Sciences and Medical Imaging Technology, Ibadat International University Islamabad, Pakistan.
[2]Department of Radiological Sciences and Medical Imaging Technology, Ibadat International University Islamabad, Pakistan.
[3]Department of Radiological Sciences and Medical Imaging Technology, Ibadat International University Islamabad, Pakistan.
[4]Department of Radiological Sciences and Medical Imaging Technology, Ibadat International University Islamabad, Pakistan.
[5]Department of Management Sciences,Ibadat International university Islamabad, Pakistan
[*6]Department of Radiological Sciences and Medical Imaging Technology, Ibadat International University Islamabad, Pakistan.
[1]naynokhan123@gmail.com, [2]syedasaliha52@gmail.com, [3]ayeshaayub292@gmail.com, [4]Rabiakamal302@gmail.com, [5]hilal.ahmed@dms.iiui.edu.pk, [*6]Shamiakamal18015@gmail.com

DOI: https://doi.org/

Keywords:
Artificial intelligence, Machine learning, Deep learning, Gestational age prediction, Neural networks, convolutional neural networks (CNN) ,Support vector machines (SVM), Algorithm validation

Article History
Received on 21 Feb, 2026
Accepted on 29 March, 2026
Published on 31 March, 2026

Corresponding Author: *
Shamia Kamal

Abstract

Background: Accurate gestational age (GA) estimation is vital for optimal prenatal care. This systematic review evaluates the accuracy and clinical utility of artificial intelligence (AI) algorithms for GA prediction using ultrasound data. By synthesizing evidence from diverse populations and imaging approaches, this review highlights current performance, potential benefits, and limitations, guiding future research and clinical adoption. Methods: This systematic review was conducted in accordance with PRISMA guidelines. A comprehensive literature search was performed across major databases, including PubMed, Scopus, Web of Science, and Google Scholar, to identify relevant studies on AI-based gestational age prediction. Results: This systematic review identified 22,350 records, including 22,300 from database searches and 50 from other sources. After removing duplicates, 18,400 records were screened, and 17,900 were excluded based on titles and abstracts. Of the 500 full-text articles assessed, 492 were excluded for not meeting inclusion criteria. Finally, 8 studies were included in the qualitative synthesis, providing valuable evidence on the accuracy and clinical utility of AI algorithms for gestational age prediction. Conclusion: . This review demonstrates that AI algorithms have the potential to improve prenatal care, particularly in settings with limited resources, and that their accuracy for predicting gestational age is encouraging. However, in order to support widespread adoption and confirm their clinical utility, additional large-scale studies and external validations are required.

INTRODUCTION

Accurate determination of gestational age (GA) is a cornerstone of safe and effective prenatal care. It guides critical decisions such as timing of screening tests, monitoring fetal growth, and planning delivery. Optimal prenatal care relies on accurate gestational age (GA) estimation for appropriate care of mother and child during pregnancy and beyond. The precise dating of pregnancy is also necessary to assess viability in premature labor and post-date deliveries . The most common methods used to estimate GA during pregnancy are based on the last menstrual period (LMP) and ultrasonographic findings(1). Fetal ultrasonography is the cornerstone of prenatal imaging and provides crucial information to guide maternal-fetal care, such as estimated gestational age (GA) and evaluation for fetal growth disorders. Currently, the clinical standard for estimating GA and diagnosing fetal growth disorders is determined through manual acquisition of fetal biometric measurements, such as biparietal diameter, head circumference, abdominal circumference (AC), femur length, or crown-rump length(2). During the first trimester of pregnancy, the risk of miscarriage is higher than in later trimesters . It is a critical period for fetal development, and various factors can contribute to this increased risk. On a global scale, approximately 15 million infants are born preterm each year (3) . Early obstetric ultrasound determines the location of the pregnancy, detects fetal cardiac activity, estimates gestation, identifies multiple pregnancies and fetal anomalies, reduces the likelihood of inductions after term, and enhances the experience of pregnancy.

Vietnam's Ministry of Health recommends at least four antenatal care (ANC) visits. In urban and rural areas, nearly all women attend at least one. As of the early 21st century, ANC includes three routine ultrasounds and a third-trimester growth scan (3) . During the ultrasound examination, various important aspects of the fetus must be identified, sized, grown, oriented, and gestational age determined. The World Health Organization recommends that all pregnant people receive at least 1 ultrasonography examination prior to 24 weeks.10 Although this policy recommendation remains largely aspirational in many low- and middle-income countries (LMICs), recent advances in ultrasonography hardware11,12 and artificial intelligence (AI)–enabled medical image analysis13,14 could facilitate broader access to this critical diagnostic tool. In 2022, a deep learning algorithm developed in an international study of 4695 pregnant volunteers that could estimate GA from blindly obtained ultrasound sweeps of the gravid abdomen was examined(4).

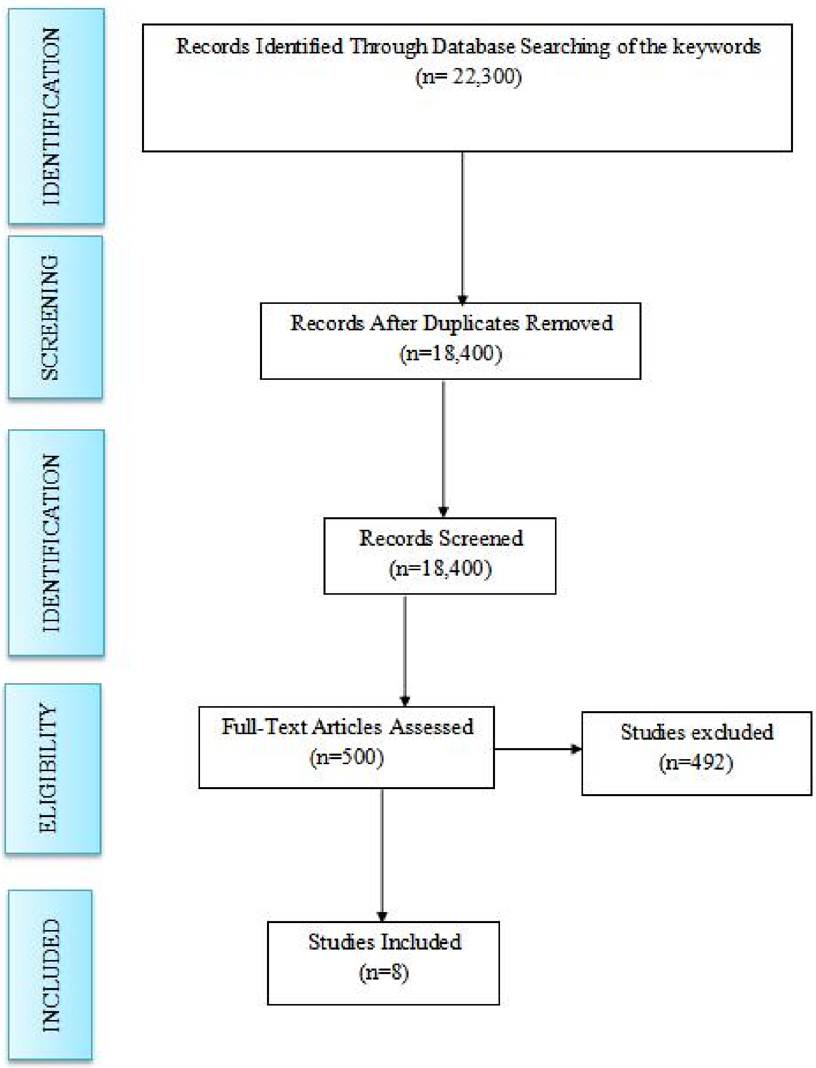
standard ultrasonography examination requires expensive equipment and skilled operators that are often not available in resource-limited setting (5) . AI has made significant strides in prenatal diagnosis, mainly through ultrasound image analysis and genetic screening. Traditional ultrasound diagnostics are highly dependent on the skills and experience of the sonographer, leading to variability in detection rates. AI algorithms, however, provide a significant advantage by analyzing ultrasound images with greater accuracy and consistency (6) . The exponential growth of artificial intelligence (AI) is defined as the use of neural networks, machine-learning (ML), or deep-learning methods; there has been a growing approach to its use in obstetric.The current uses of AI in obstetric ultrasound include first trimester pregnancy ultrasound, assessment of placenta, fetal biometry, fetal echocardiography, fetal neurosonography, assessment of fetal anatomy, and other uses including assessment of fetal lung maturity and screening for risk of adverse pregnancy outcomes. AI holds the potential to improve the ultrasound efficiency, pregnancy outcomes in low resource settings, detection of congenital malformations and prediction of adverse pregnancy outcomes (7) . This systematic review aimed to critically evaluate how well artificial intelligence (AI) tools can estimate gestational age by comparing their output—based on 2D ultrasound images and blind- sweep videos—to standard ultrasound biometry measurements. Additionally, the review aimed to compare different AI approaches (such as CNN and DNN) to determine which methods offered the most accurate results

METHOD AND MATERIAL:

The PRISMA guidelines are followed in this systematic review. A comprehensive literature search was conducted between time period of (year 2019 – 2025) including are PubMed , Scopus, Web of Science and Google Scholar etc.Terms included combinations of keywords and Medical Subject Headings (MeSH) such as

Artificial Intelligence
Machine Learning
Deep Learning
Prediction of gestational age
Neural Networks
Convolutional Neural Networks (CNN)
SVM, or Support Vector Machines
Unsupervised Learning
Algorithm Validation

This systematic review aims to critically evaluate and synthesize the existing evidence regarding the accuracy, performance, and clinical utility of artificial intelligence (AI) algorithms for predicting gestational age using a variety of modalities, with an emphasis on ultrasound imaging and related diagnostic tools. This review will include English-language original studies with peer review that evaluate the accuracy or performance of artificial intelligence, machine learning, or deep learning algorithms for predicting gestational age in human pregnancies based on clinical imaging (such as ultrasound) or other biomedical data. Non-human or in vitro studies, commentaries, letters, conference abstracts without full text, studies that do not specifically evaluate an AI method for gestational age prediction, and studies that do not report sufficient diagnostic performance data will all be excluded.



Results

In conducting this systematic review on the accuracy and clinical utility of AI algorithms for gestational age prediction, a comprehensive search across multiple databases yielded a total of 22,300 records. An additional 50 articles were identified through other sources such as reference lists, bringing the total to 22,350. After removing duplicates, 18,400 unique records remained for screening. During the initial screening phase, 17,900 records were excluded due to irrelevance based on title and abstract. The majority of these were studies that did not focus on AI, did not target the estimation of gestational age, or did not involve human subjects. This left 500 articles for full-text review. After thorough evaluation, 492 articles were excluded for reasons such as inadequate methodology, lack of performance data, or insufficient focus on clinical utility. In the end, only eight studies were included in the final qualitative synthesis because they met all of the inclusion requirements. The current state of AI-based gestational age prediction tools, their methodological sturdiness, and their potential clinical applicability are examined in these studies.

Author and Year	Country	Income Region	Study Design	Sample Size	Input Measures (2D images vs blind)	Input Measures (video vs still)	Size of training and test set	GA range by Gold Standard	AI mode used	Performance Metrics (AI Algorithm)	Validation Method	External validation
Chace Lee et al., 2023	USA and Zambia	High-income (USA) and Low-income (Zambia)	Retrospective cohort study	Total: 3,842 participants	Standard plane 2D still images sweep videos	Both – Still images and video sweeps	Training: 3,438 cases Test: 404 cases	typically 9–40 weeks Standard fetal biometry	Three models image-only, video-only, and ensemble (multimodal)	Ensemble model: Mean Absolute Error (MAE) −1.51 days (SD 3.96) vs. standard fetal biometry	Internal validation with Fetal Age Machine Learning Initiative (FAMLI) cohort	No
Fatima Rauf et al., 2024	Pakistan, Saudi Arabia, South Korea	Lower-middle-income (Pakistan) and high-income	Experimental study evaluating deep learning models for scan	9,938 still 2D ultrasound images	Still 2D images ; not blind sweeps	Still images only	60% training 20% validation 20% testing	8_ 37 Weeks Standard fetal biom	Two CNN architectures one with 3 residual blocks, one with 4	Classification accuracy: 98.5% (3 blocks) and 88.6% (4 blocks)	Internal experimental validation using public datasets	Not performed

		(Saudi Arabia, South Korea)	plane classification					etry	residual blocks			
Stringer et al., 2024	Zambia and USA	Low-income (Zambia) and high-income (USA)	Prospective diagnostic accuracy study	407 cases (USA), 114 cases (Zambia), 613 cases (malpresentation), training on 4,695 prior cases	Blind ultrasound video sweeps	Video sweeps only	Training: 4,695 Test: 407 (USA), 114 (Zambia), 613 (malpresentation	9–14 weeks Standard fetal biometry	Deep learning model for direct GA prediction & fetal malpresentation detection	GA prediction: MAE 3.2 days vs. 3.0 days (standard biometry); 90.7% within ±7 days; Malpresentation: AUC ~0.977	Prospective validation using blinded CRL as gold standard	Yes — separate novice user cohort in Zambia & across different devices
Naz et al., 2025	Multiple countries across high-, upper-middle-, and low/mi	High-, upper-middle-, and low/middle-income countries	Comprehensive systematic review and meta-analysis (17 studies; 10 in meta-	Typical: 400 training, 100 test cases per study	Both 2D still ultrasound images and blind sweep videos	Both — still images and video sweeps	400 training 100 test	8–40 weeks Standard fetal biometry	Various CNN and DNN models	Mean error: 4.32 days (2D images), 2.55 days (blind sweep videos); second trimester	Internal or cross-validation in most studies	Only a few studies performed robust external validation

	ddle-income regions		analysis)							mean error: 2.35 days		
Szymon Płotka et al., 2022	Multiple countries	Multiple income regions	Prospective cohort study	Total: 750 cases	Blind ultrasound sweeps focusing on biometric parameters	Video sweeps	80% training 20 percent testing	8–40 weeks Standard fetal biometry	CNN-based model for GA estimation using blind sweeps	Mean Absolute Error: 0.05 ± 0.01 weeks	Internal validation only	Not performed
Thomas LA van den Heuvel et al., 2019	Ethiopia and the Netherlands	Low-income (Ethiopia) and high-income (Netherlands)	Prospective study	183 cases	2D free-hand ultrasound sweeps	Video sweeps	60% training 20% testing	20–40 weeks Standard fetal biometry	U-Net architecture for GA estimation	Mean interval difference: 3.6 ± 9.8 days compared to standard measures	Internal validation only	No
Teeranan Pokaprakarn et al., 2022	North Carolina, USA and Zambia	High-income (USA) and low-income (Zambia)	Prospective study	4,695 pregnant participants	Blind ultrasound cineloop video	Video sweeps	80% training 10% tuning 10% testing	9–37 weeks Standard fetal biom	End-to-end deep learning CNN model	MAE: 2.1 ± 0.19 days (1st trimester), 3.1 ± 0.16 days (2nd),	Internal validation	No

								etry		4.7 ± 0.18 days (3rd)		
Burgos- Artizzu et al., 2021	Spain (BCNatal, Barcelona)	High-income	Retrospective Study	Training: 1,394 patients; Evaluation: 3,065 scans from 1,992 patients	Standard 2D axial fetal brain ultrasound images	Still images	Training: 1,394; Test: 3,065 scans	22 – 40 weeks Standard fetal biometry	CNN-based automated analysis	95% CI error: AI alone 14.2 days; AI + biometry 11 days; biometric alone 12.9 days; trimester breakdown: 2nd T— 6.7 vs 7 days; 3rd T— 14.3 vs 17 days	Internal validation	No

DISCUSSION

Using ultrasonography data from both high- and low-income settings, specifically Zambia and the United States, the study by Chace Lee et al. (2023) demonstrates the potential of artificial intelligence to improve the estimation of gestational age. Standard plane still images and "fly-to" ultrasonography videos were used to analyze a cohort of 3,842 participants and a separate test set of 404 cases. Three AI models were developed: an image-only model, a video-only model, and an ensemble model combining both data types. The ensemble model showed the best performance, with a mean absolute error of −1.51 days (SD 3.96) compared to standard fetal biometry, indicating improved accuracy. The Fetal Age Machine Learning Initiative (FAMLI) cohort was used internally for validation; however, no external validation was performed, indicating the need for additional research to verify generalizability across other populations and settings. Overall, the findings support the feasibility of using multimodal AI tools to enhance prenatal care, especially in low-resource contexts(8).

Fatima Rauf et al.'s (2024) investigation looked at deep learning models for classifying still 2D fetal ultrasound scan planes from Pakistan, Saudi Arabia, and South Korea, encompassing both low-middle-income and high-income settings. Using a dataset of 9,938 images spanning multiple anatomical regions, the study developed two novel CNN architectures with three and four residual blocks. The models achieved high classification accuracies of 98.5% and 88.6% with a data split of 60% training, 20% validation, and 20% testing. This research specifically targeted accurate scan plane recognition to support automated ultrasound workflows, in contrast to other studies that focused on estimating gestational age. However, the findings were based on experimental validation with public datasets, and no external clinical validation was performed, indicating a need for further testing in real-world clinical settings(9).

The study by Stringer et al. (2024) assessed a deep learning model for direct gestational age prediction and fetal malpresentation detection using blind ultrasound video sweeps in Zambia and the USA. Conducted as a prospective diagnostic accuracy study, it included 407 cases in the USA, 114 in Zambia, and 613 cases for malpresentation assessment, with training based on 4,695 prior cases. The model, optimized for mobile devices, achieved a mean absolute error of 3.2 days compared to 3.0 days for standard fetal biometry, with 90.7% of estimates within ±7 days. With an AUC of approximately 0.977, the malpresentation detection proved to be highly accurate. Importantly, the study validated the model prospectively using blinded crown–rump length as the gold standard and included external validation in a separate novice user cohort in Zambia and across different devices, demonstrating strong potential for real-world application in low-resource settings(10).

A comprehensive review was carried out by Naz et al. (2025), combining the findings of 17 prospective and retrospective studies that investigated the use of AI for estimating gestational age in a variety of settings, including countries with high, upper-middle, and low/middle incomes. A meta-analysis of 10 of these studies included blind sweep videos and 2D still ultrasound images. Each study typically divided the dataset into 400 training and 100 test cases, with gestational ages ranging from 8 to 40 weeks. The performance of various CNN and DNN models varied by trimester, with a mean error of 4.32 days for 2D images and 2.55 days for blind sweep videos. The subgroup results showed that accuracy was higher (mean error 2.35 days) in the second trimester than in the first or third. Most studies used internal or cross-validation, with only a few performing robust external validation. Overall, the findings highlight that AI, especially using blind sweep video data, holds promise for improving gestational age estimation accuracy but underline the need for broader validation across populations and scanning environments(1).

Szymon Płotka et al. (2022) investigated a CNN-based approach for estimating gestational age using blind ultrasound sweeps in a prospective cohort spanning multiple countries and income regions. The study included two datasets: 700 cases for training and testing and an additional 50 freehand fetal ultrasound video scans for evaluation. The video-captured standard biometric parameters (HC, BPD, AC, and FL) were the focus of the data. With 80% of data used for training and 20% for testing, the model achieved a mean absolute error of just 0.05 ± 0.01 weeks, demonstrating strong accuracy. However, no external validation was carried out, indicating that additional research is required to verify its generalizability to a variety of clinical settings(11).

In their 2019 study, van den Heuvel et al. developed a deep learning model to automatically detect the fetal head and estimate head circumference from free-hand ultrasound sweeps. Tested on a dataset, the model achieved a mean absolute error (MAE) of about 1.99 mm for head circumference measurements compared to manual expert annotations, showing strong agreement. This level of accuracy highlights the potential of AI to deliver reliable fetal biometry in resource-limited countries, where access to skilled sonographers is often limited. However, to fully realize clinical benefits, further large-scale validation and real-world deployment are required to confirm consistency across different devices, operators, and populations(12).

Teeranan Pokaprakarn et al. (2022) conducted a prospective study in North Carolina, USA, and Zambia to assess an end-to-end deep learning model using blind ultrasound cineloop videos for gestational age estimation. 4,695 pregnant women participated in the study, and the data were divided into 80% for training, 10% for tuning, and 10% for testing. The CNN model produced promising mean absolute errors for gestational ages ranging from 9 to 37 weeks, with 2.1 0.19 days in the first trimester, 3.1 0.16 days in the second, and 4.7 0.18 days in the third. Despite excellent performance, no external validation was carried out, highlighting the requirement for additional testing to demonstrate its adaptability to a wider range of settings(13).

Burgos Artizzu et al. used standard 2D axial fetal brain ultrasound images to train and test a CNN-based automated system for estimating gestational age in their 2021 study in Barcelona, Spain. The dataset included 1,394 patients for training and 3,065 scans from 1,992 patients for evaluation, covering pregnancies from 22 to 40 weeks. Comparable to biometry alone, the AI model performed well, with a 95% CI error of 14.2 days for AI alone and an improvement to 11 days when combined with standard biometry. Notably, accuracy was higher in the second trimester (6.7 days vs. 7 days) and decreased in the third trimester (14.3 days vs. 17 days). Although the study relied solely on internal validation, it highlights the need for external testing to confirm generalizability and demonstrates the potential of AI to complement conventional biometry(14).

CONCLUSION

Systematic review demonstrates that artificial intelligence algorithms hold significant promise for accurately estimating gestational age, often matching or surpassing the performance of standard biometry, especially in low-resource and blind ultrasound settings. Across diverse study designs, populations, and imaging modalities, AI models showed consistently low prediction errors and high agreement with established clinical standards.

However, the need for more robust, large-scale prospective trials and external validations across various populations and clinical contexts is highlighted by variations in study quality, sample sizes, and validation methods.

LIMITATIONS AND FUTURE RESEARCH

This review has some limitations, like the fact that AI models, input data types, and validation methods vary a lot between studies, making it hard to compare them all. The generalizability of their findings was limited by the fact that many studies relied on small or single-center datasets and lacked robust external validation. Additionally, the need for more rigorous and standardized research in this area is highlighted by issues like potential publication bias, inconsistent reporting of model performance, and limited evidence on practical implementation and user acceptance. Future research should prioritize well-designed, large-scale, multicenter prospective trials that include diverse populations from both high-income and low-resource settings. Such studies should emphasize rigorous external validation to assess model generalizability and performance under real-world clinical conditions.

Moreover, future work should explore the integration of AI-based gestational age estimation into existing clinical workflows, assessing not only diagnostic accuracy but also practical implementation aspects, user training, cost-effectiveness, and patient acceptability

REFERENCES

1. Naz S, Noorani S, Jaffar Zaidi SA, Rahman AR, Sattar S, Das JK, et al. Use of artificial intelligence for gestational age estimation: a systematic review and meta-analysis. Vol. 6, Frontiers in Global Women's Health. Frontiers Media SA; 2025.
2. Lee C, Willis A, Chen C, Sieniek M, Watters A, Stetson B, et al. Development of a Machine Learning Model for Sonographic Assessment of Gestational Age. JAMA Netw Open. 2023 Jan 4;6(1):E2248685.
3. Rauf F, Attique Khan M, Albarakati HM, Jabeen K, Alsenan S, Hamza A, et al. Artificial intelligence assisted common maternal fetal planes prediction from ultrasound images based on information fusion of customized convolutional neural networks. Front Med (Lausanne). 2024;11.
4. Gimovsky AC, Eke AC, Tuuli MG. Enhancing Obstetric Ultrasonography with Artificial Intelligence in Resource-Limited Settings. Vol. 332, JAMA. American Medical Association; 2024. p. 626–8.
5. Pierucci UM, Tonni G, Pelizzo G, Paraboschi I, Werner H, Ruano R. Artificial Intelligence in Fetal Growth Restriction Management: A Narrative Review. Vol. 53, Journal of Clinical Ultrasound. John Wiley and Sons Inc; 2025. p. 825–31.
6. Benson M, Walton S, Hartley T, Meagher S, Seshadri S, Sleep N, et al. Fetal Gestational Age Estimation Using AI on Simple Ultrasound Images and Video [Internet]. 2025. Available from: https://www.researchsquare.com/article/rs-5907990/v1
7. Horgan R, Nehme L, Abuhamad A. Artificial intelligence in obstetric ultrasound: A scoping review. Vol. 43, Prenatal Diagnosis. John Wiley and Sons Ltd; 2023. p. 1176–219.
8. Lee C, Willis A, Chen C, Sieniek M, Watters A, Stetson B, et al. Development of a Machine Learning Model for Sonographic Assessment

of Gestational Age. JAMA Netw Open. 2023 Jan 4;6(1):E2248685.

9. Rauf F, Attique Khan M, Albarakati HM, Jabeen K, Alsenan S, Hamza A, et al. Artificial intelligence assisted common maternal fetal planes prediction from ultrasound images based on information fusion of customized convolutional neural networks. Front Med (Lausanne). 2024;11.
10. Stringer JSA, Vwalika B. Diagnostic accuracy of a novel machine learning algorithm to estimate gestational age. 2023.
11. Płotka S, Klasa A, Lisowska A, Seliga-Siwecka J, Lipa M, Trzciński T, et al. Deep learning fetal ultrasound video model match human observers in biometric measurements. Phys Med Biol. 2022 Feb 21;67(4).
12. van den Heuvel TLA, Petros H, Santini S, de Korte CL, van Ginneken B. Automated Fetal Head Detection and Circumference Estimation from Free-Hand Ultrasound Sweeps Using Deep Learning in Resource-Limited Countries. Ultrasound Med Biol. 2019 Mar 1;45(3):773–85.
13. Pokaprakarn T, Prieto JC, Price JT, Kasaro MP, Sindano N, Shah HR, et al. AI Estimation of Gestational Age from Blind Ultrasound Sweeps in Low-Resource Settings. NEJM Evidence. 2022 Apr 26;1(5).
14. Burgos-Artizzu XP, Coronado-Gutiérrez D, Valenzuela-Alcaraz B, Vellvé K, Eixarch E, Crispi F, et al. Analysis of maturation features in fetal brain ultrasound via artificial intelligence for the estimation of gestational age. Am J Obstet Gynecol MFM. 2021 Nov 1;3(6).