Essence of Scientific Medical Journal (2025), Volume 23, Number 1:1-8
P-ISSN.1979-0147, E-ISSN. 2655-6472

LITERATURE REVIEW

ARTIFICIAL INTELLIGENCE IN CARDIOTOCOGRAPHY FOR ASSESSING FETAL WELL-BEING: A NARRATIVE REVIEW OF CURRENT EVIDENCE, CHALLENGES, AND NEXT STEPS

Leonardo Alpha Diaz,[1] Davin Beta Tadeo,[1] Rafael Kurnia Pratama Tokan, [1] Richard Christian Suteja,[1] Andrea Ivena[1]
[1]Faculty of Medicine, Udayana University

ABSTRAK

Pendahuluan: *Cardiotocography* (CTG) telah lama menjadi alat utama dalam pemantauan janin, tetapi interpretasinya sering terkendala pada variabilitas pengamat. *Artificial Intelligence* (AI) muncul sebagai solusi yang menjanjikan dalam meningkatkan akurasi dan konsistensi dalam interpretasi pola CTG.
Pembahasan: Tinjauan ini mengeksplorasi AI, khususnya pada *machine learning* dan *deep learning*, dalam mendukung klasifikasi CTG. Model seperti *random forest* dan *Convolutional Neural Networks* (CNN) telah menunjukkan kinerja yang baik dalam mendeteksi kondisi janin. Meskipun demikian, AI masih belum siap untuk menggantikan penilaian klinis karena beberapa keterbatasan dalam kualitas data, interpretabilitas, dan penggunaan etika informasi medis. Pendekatan kombinasi AI dan keahlian manusia dapat menjadi jalan yang paling efektif.
Simpulan: AI memiliki potensi kuat sebagai alat pendukung klinis dalam interpretasi CTG. Dengan pengembangan lebih lanjut, integrasi multimodal, dan *explainable* AI *approaches*, pengambilan keputusan selama persalinan dapat ditingkatkan dengan cara yang lebih akurat dan personal.

Kata kunci: *Artificial Intelligence*, *Cardiotocography*, *Deep Learning*, *Machine Learning*, Pemantauan Janin

ABSTRACT

Introduction: *Cardiotocography (CTG) has long been a key tool for fetal monitoring, yet its interpretation is often limited by interobserver variability. Artificial intelligence (AI) has emerged as a promising solution to enhance accuracy and consistency in the interpretation of CTG patterns.*
Discussion: *This review explores how AI, particularly machine learning and deep learning, supports CTG classification. Models like random forest and Convolutional Neural Networks (CNN) have shown strong performance in detecting fetal conditions. However, AI is not yet ready to replace clinical judgment due to some limitations in data quality, interpretability, and ethical use of medical information. Hybrid approaches that combine AI with human expertise appear to be the most effective path forward.*
Conclusion: *AI holds strong potential as a clinical support tool in CTG interpretation. With further development, multimodal integration, and explainable AI approaches, decision-making during labor may be improved in a more accurate and personalized way.*

Keywords: *Artificial Intelligence, Cardiotocography, Deep Learning, Fetal Monitoring, Machine Learning*

[1]Faculty of Medicine, Udayana University, Bali

INTRODUCTION

Initially, fetal heart rate (FHR) was assessed using basic methods such as inspection and palpation, later progressed with the introduction of instruments like the fetoscope and phonocardiography. A major breakthrough came in the late 1960s with cardiotocography (CTG), which enabled continuous and simultaneous monitoring of FHR and uterine activity, establishing CTG as a cornerstone of intrapartum fetal surveillance in modern obstetrics.[1] By the late 1970s, CTG was already used in approximately 175,000 deliveries or 50% of labors and its adoption expanded to almost 85% of births in United States by 2002.[2] Today, CTG has become a standard of care in obstetrics practice and is widely used across the globe.[2,3]

However, despite being in clinical use for over 50 years, the interpretation of CTGs remains highly subjective. According to the Each Baby Counts report in 2015, 62% of stillbirths, neonatal deaths, and brain injuries among term babies were linked to CTG misinterpretation and mismanagement.[4] The efficacy of CTG to identify hypoxic injury enables clinicians to do intervention early when needed. However, CTG faces a notable challenge, as one of the main issues is the high inter-observer variability in its interpretation, which results in substantial subjectivity between different observers or even within the same observer at different times. A study assessing the 2015 International Federation of Gynecology and Obstetrics (FIGO) guidelines in CTG interpretation reported an overall interobserver agreement (Pa) of 0.60, indicating a moderate level of consistency and raising concerns about the subjectivity of CTG classification.[5] Another study comparing four FHR classification systems (including a 2-tier system, NICHD 2008 (3-tier), CNGOF 2013 (5-tier), and FIGO 2015 (3-tier)) found only moderate interobserver agreement across all, highlighting the persistent subjectivity in CTG interpretation regardless of the guideline used.[6] Moderate agreement in CTG interpretation is concerning due to the influence of CTG on clinician intervention's decision, such as cesarean procedure. A recent meta-analysis further demonstrated that continuous CTG is consistently associated with increased rates of cesarean delivery and instrumental vaginal births, likely reflecting its high false-positive rate and the challenges of subjective interpretation leading to unnecessary interventions.[7]

In this advanced digital era, artificial intelligence (AI) could offer solutions and assistance to improve objectivity and accuracy of CTG interpretation. Machine learning, one of the subfields of AI, enables the development of predictive models that can analyze complex CTG patterns, reduce observer variability, and potentially help clinicians in interpreting CTG more consistently.[8,9] Despite all potentials described above, the use of AI faces critical challenges, including potential algorithmic bias and poor performance stemming from non-representative or poor-quality data, as well as lack of transparency in 'black-box' models.[10] With the increasing reliance on AI in clinical decision-making, a comprehensive review is essential to assess its reliability, accuracy, and readiness for its integration into obstetric practice. This paper aims to review current evidence of AI in assisting CTG interpretation and enhancing clinical decision-making regarding interventions based on CTG results.

DISCUSSION

Brief Overview of CTG

CTG is an instrument used to assess fetal well-being by simultaneously recording FHR and uterine contractions. It employs two transducers to capture fetal cardiac activity and uterine activity, which may be recorded internally or externally. The first transducer is used to monitor the FHR using a Doppler Ultrasound Transducer (external transducer) or fetal electrode (internal transducer).[11] The second transducer is used to monitor the uterine activity using tocodynamometer (external transducer) or intrauterine catheter (internal transducer).[3] Although internal transducers can provide more accurate recordings, their use is invasive and associated with certain risks; hence, external transducers are recommended for routine intrapartum monitoring.[3, 7]

The CTG output is typically displayed on a continuous strip, either on paper or digitally, allowing real-time interpretation of several key parameters: baseline FHR, variability, accelerations, decelerations, and uterine contractions. The baseline FHR is the mean heart rate calculated over a 10-minute period and expressed in beats per minute (bpm). Variability refers to the fluctuations in the FHR signal, assessed by calculating the range of the signal's average amplitude over a one-minute interval. Acceleration is the abrupt increase of FHR above the baseline for more than 15 bpm for 15 seconds but less than 10 minutes. Deceleration refers to the decrease of FHR below the baseline by more than 15 bpm for more than 15 seconds.[3]

Interpretation of CTG is conducted by systematically evaluating these parameters to classify the tracing into one of three categories: normal, suspicious, or pathological.[3] According to the FIGO 2015 guidelines, a suspicious CTG is defined by the presence of one inconclusive feature. Meanwhile, a pathological CTG is identified by one or more abnormal findings, such as baseline heart rate below 100 bpm, absent or markedly increased variability, sinusoidal patterns, or repetitive late or prolonged decelerations. These categorizations have direct clinical implications. Suspicious tracings require continuous observation and reevaluation, whereas pathological patterns necessitate prompt clinical intervention to prevent fetal compromise.[3]

Recent Evidences of AI in CTG

AI has advanced significantly in recent years. Some of the most popular uses of AI in daily applications are Artificial intelligence-large language models (AI-LLMs), such as ChatGPT-4o and Gemini Advanced. Interestingly, a blinded study written by Gumilar *et al.* (2025) on the interpretation of seven CTG images reported that ChatGPT-4o's performance was better than junior doctors. Furthermore, ChatGPT-4o's performance score was comparable to senior doctors without significant difference. However, the study also noted that junior doctors scored higher than both Gemini Advanced and Copilot.[12]

Machine learning, a subfield of AI, has emerged as a transformative tool in medical diagnostics, offering powerful capabilities in pattern recognition, classification, and prediction. There are two broad categories of machine learning algorithms used in studies for CTG interpretation: traditional (classic) machine learning methods and deep learning architectures. Traditional machine learning relies on handcrafted features derived from quantitative data extracted from raw CTG signals, based on parameters defined by clinical guidelines such as those issued by FIGO. These parameters include baseline fetal heart rate, variability, acceleration, deceleration, and uterine activity.[3,13] Some examples of dominant models from traditional machine learning are random forests (RF), support vector machines (SVM), and gradient boosting.[14]

Random forest is an ensemble learning method that uses data collected from multiple decision trees and combines their outputs to improve overall prediction accuracy and reduce the risk of overfitting, making it generally more robust than a single decision tree model. A study demonstrated that the RF method achieved 95.11% accuracy in predicting fetal conditions based on 22 cardiotocography features, outperforming Naïve Bayes and Decision Tree classifiers.[15] Similarly, another study reported combination of RF with synthetic minority oversampling technique (SMOTE) to address class imbalance in CTG data achieved high performance, with 94.40% accuracy, 94.45% precision, 94.40% recall, and a 94.38% F1-score, highlighting the model's robustness in classifying fetal health conditions.[16] Other traditional models like SVM and gradient boosting have also been reported to show strong performance in CTG classification.[14,17]

Unlike traditional methods, deep learning models like Convolutional Neural Networks (CNNs) and Recurrent Neural Networks (RNNs) can learn patterns directly from the raw CTG data, reducing dependency on manual feature engineering.[18] While traditional machine learning focused on testing whether machines could follow human-defined rules more effectively, deep learning explores whether raw CTG signals contain hidden patterns that are undetectable by humans and not captured by existing guidelines that can predict fetal conditions. The true potential of AI may lie not just in creating ‘super-interpreters’ of current rules, but also in uncovering entirely new physiological markers of fetal distress. This suggests that clinical guidelines followed for decades might be incomplete, which is why studies discovering invisible patterns are considered so significant and why the field is rapidly moving toward deep learning models.[19] A study demonstrated that CNN could achieve high accuracy (AUC = 0.896) in identifying late decelerations using only fetal heart rate without uterine contraction data.[20] Sriram *et al*. proposed a hybrid auto-encoder with RNN that achieved an AUC of 99.77% in classifying fetal distress from CTG signals, indicating near-perfect performance in distinguishing compromised from healthy cases.[21] These findings show how CNNs and RNNs, through deep abstraction, may uncover subtle signal variations that are often imperceptible to the human eye, resulting in more precise and personalized monitoring of fetal well-being.

The Accuracy of AI in Comparison to Human Judgment for CTG Interpretation

The rapid advancement of AI in medicine raises a new question: Could AI substitute human judgement entirely? However, AI still faces some challenges and obstacles. After exploring the role and performance of AI in CTG interpretation, the next crucial step is to evaluate how AI compares to the current gold standard: human expert judgment. In the previous chapter, we already discussed a study comparing AI-LLMs with human judgment, which reported that ChatGPT-4o significantly outperformed junior doctors, while showing no significant difference compared to senior doctors. Further analysis revealed that Chat-GPT4o scored highest in the “depth” category, suggesting AI could potentially provide more detailed and comprehensive interpretation.[12] In contrast, senior doctors outperformed in terms of relevance, coherence, and focus, demonstrating their ability to filter information and highlight the most clinically important aspects within the specific context of each patient. Interestingly, another study reported that human doctors outperformed AI achieving a statistically significant area under the curve (AUC) of 0.693, while the machine learning and deep learning models performed at a level indistinguishable from random chance. The study also reported that a hybrid human-AI approach did not surpass the overall diagnostic accuracy of human doctors alone. However, this hybrid system did improve specificity, showing its potential as a useful adjunct to support doctor's judgment in reducing false-positive interpretations.[19]

In addition to direct comparisons between AI and clinicians, recent study has shown that integrating clinical knowledge into AI systems can enhance performance. Zhao *et al*. (2024) introduced a hybrid model that combined deep learning from CTG signals with expert-derived clinical features, reflecting how obstetricians traditionally interpret fetal heart rate. The results revealed that this knowledge-enhanced AI outperformed the signal-only models, supporting the idea that AI systems are most effective when grounded in human medical expertise.[22] These findings suggest that instead of replacing clinicians, AI systems that incorporate expert knowledge can enhance diagnostic precision and serve as reliable decision-support tools during labor.

Further supporting the potential of AI, a recent study by Melaet *et al*. (2024) evaluated an AI system designed to predict FHR responses during labor, based solely on CTG data from healthy pregnancies. The model's predictions were directly compared with minute-by-minute annotations made by three experienced obstetricians who had access to the entire CTG recordings and neonatal outcomes. Despite the AI operated without future outcome knowledge, it achieved high agreement with the expert annotations.[23] Notably, the model demonstrated an area under the receiver operating characteristic (ROC) curve (AUC) of 0.96 for distinguishing between normal and pathological events, an accuracy level comparable to expert consensus.[23] These findings suggest that AI may reach expert-level performance in certain tasks of CTG interpretation and has strong potential as a supportive tool in clinical decision-making.

In a large-scale, multicenter study conducted in South Korea, Park *et al*. (2025) developed a deep learning-based CTG interpretation model and validated it against expert annotations made through a two-step process involving board-certified obstetricians and final judgment by a senior obstetrician with over 15 years of experience. The model demonstrated strong diagnostic performance, achieving an AUC of 0.880 in internal validation and up to 0.895 in external datasets. These results suggest that AI models can closely match the interpretive accuracy of experienced human experts across a range of clinical conditions. Notably, the model maintained high performance even in high-risk subgroups, such as preterm births and emergency deliveries, further highlighting its potential to support clinicians in critical decision-making during labor.[24]

Challenges and Limitations

One of the main challenges of AI implementation in medicine is data quality and availability. Most AI models require large, diverse, and well-labelled datasets to perform reliably, but medical data is often fragmented, inconsistent, or biased toward certain populations.[25] This problem is especially apparent in the development of AI for CTG interpretation. Serious conditions like hypoxic-ischemic encephalopathy (HIE) are very rare, so datasets are often unbalanced, making it hard for AI to recognize such events and develop reliable predictive models. To deal with this, researchers use indirect labels like low pH or Apgar scores, but these are not always reliable. Many CTG datasets also lack detailed labels for specific time points, and signals can look different depending on the stage of labor. What appears abnormal in the first stage may be entirely normal during the second stage, and yet most studies overlook this temporal context. Even when expert annotations are available, they vary a lot, which raising concerns about accuracy and objectivity of human-labeled data. These problems show the need for better-quality data and more consistent labeling to make AI models more dependable in real clinical use.[26] Oversampling methods like SMOTE are often used to address class imbalance, but if the original minority class is poorly labeled or noisy, such techniques may end up replicating and amplifying those errors rather than correcting them.[27]

The challenge of fragmented and insufficient data for AI is further complicated by the problem of data silos. Data silos refer to isolated systems that prevent smooth integration of information across healthcare settings. Medical records are often stored in separate and incompatible formats, making it difficult to compile the comprehensive datasets needed for AI training. Standards like Fast Healthcare Interoperability Resources (FHIR) and open Electrical Health Record (openEHR) aim to improve data sharing and interoperability, but they also introduce concerns about privacy and security, highlighting the need for strong data governance and consent management frameworks. Beyond technical integration, ethical concerns around data usage must also be addressed. Patients are the rightful owners of their personal health information, while healthcare institutions act merely as custodians.[28] This distinction brings attention to critical issues such as informed consent, transparency, and the potential for unauthorized data sharing. As AI systems increasingly depend on sensitive clinical data, it becomes essential to ensure that data practices respect individual rights and maintain public trust.[29]

A further scoping review highlighted key barriers in applying AI to CTG interpretation.[30] Most studies relied on a single public dataset without external validation, making their findings hard to generalize. Labels used to define fetal distress varied widely, such as different pH cutoffs, leading to inconsistent training targets. Many models also ignored important clinical features like uterine contractions and used black-box algorithms with little interpretability.[30] These issues raise doubts about the reliability, transparency, and clinical readiness as well as serious barriers to clinical trust and ethical acceptance of current AI tools in fetal monitoring.[30,31] These challenges should not be seen as reasons to abandon AI in medicine, but rather as essential considerations for its responsible development. By identifying these limitations early, future studies can address them more effectively and build systems that are both clinically useful and ethically accepted.

In line with these observations, several studies have explored ML approaches for CTG analysis, each showing methodological progress yet sharing similar limitations. The early work by Islam and Yulita (2024) demonstrated promising results using the Random Forest algorithm, achieving 95.11% accuracy and highlighting the strength of ensemble methods in fetal condition classification. However, their dataset was relatively small and imbalanced, which limited generalizability to diverse clinical populations. Later, Salini *et al. (*2024*)* expanded this approach by comparing multiple algorithms, including Random Forest, Logistic Regression, Support Vector Classifier, K-Nearest Neighbor, and Gradient Boosting. Their study improved methodological rigor through stratified sampling and careful data preprocessing but still relied on a single public dataset without external validation. More recently, Nazli *et al. (*2025) further advanced this field by integrating deep learning models and applying the Synthetic Minority Oversampling Technique (SMOTE) to mitigate data imbalance, emphasizing balanced accuracy as a fairer evaluation metric. Despite these improvements, all three studies faced recurring challenges, such as limited dataset diversity, lack of multi-center or external validation, and minimal attention to model interpretability and real-time clinical integration. Collectively, these findings reflect continuous progress in AI-based CTG analysis while underscoring the persistent barriers that must be addressed for reliable clinical adoption.

Future Directions and Research Opportunities

While the challenges and limitations are considerable, the field of AI in medicine, specifically in CTG interpretation as discussed in this review has so many opportunities for future research and development. This section outlines opportunities that could redefine the role of AI in enhancing maternal and fetal safety. Exploring these avenues is essential to bridge the gap between technological innovation and real-world clinical application.

Data availability and quality remain major barriers to building robust AI models. Developing clinically reliable systems may require millions of CTG recordings far beyond what any single institution currently holds. A significant step forward is the creation of the Oxford Maternity (OxMat) dataset, the largest curated CTG collection to date. It includes over 177,000 digital CTG recordings from more than 51,000 pregnancies, along with more than 200 clinical variables spanning antepartum, intrapartum, and postpartum periods.[35] Unlike many older datasets, OxMat provides raw, time-series data, making it ideal for machine learning. Its comprehensive structure, especially the rich antepartum data, offers a valuable opportunity to train more accurate, generalizable, and clinically useful AI models.[35]

Future AI systems will likely move from analyzing CTG data alone to multi-modal models that combine CTG with other clinical data of the patients. Early studies have shown that adding factors like gestational age, parity, maternal, hypertension, and labor duration can improve accuracy.[26] A recent study by Bai *et al*. demonstrated that integrating computerized CTG (cCTG), ultrasound data, and electronic health records into a multimodal neural network significantly enhanced the prediction of delivery mode, achieving an AUC of 97.1% and accuracy of 93.3%.[36] These findings highlight the potential of combining real-time and historical data sources to build more robust and personalized AI models for labor management.

While these advances demonstrate impressive technical performance, their impact depends on whether clinicians are willing to trust and adopt such systems. Clinicians must be able to trust AI recommendations for AI to be adopted in high-stakes fields like medicine. This has raised a critical research area known as Explainable AI (XAI), which turns ‘black box’ problems into transparent and more trust-worthy models. Explainable AI provides a clear rationale for making the models’ conclusions easier to understand and evaluate. Tools like heatmaps can help show which CTG features influence predictions. The goal is to create models that are both accurate and transparent, so clinicians can trust and safely use them in practice.[37]

Since data sharing across multiple institutions is often limited, Federated Learning (FL) provides a promising approach for training robust AI models. This decentralized approach enables model training on diverse datasets while keeping sensitive health information within institutional boundaries, addressing both privacy concerns and the problem of data silos. Moreover, FL can help reduce algorithmic bias by allowing underrepresented groups' data to be included from geographically dispersed sources. However, several challenges remain. Model updates in FL may still leak sensitive data unless protected by techniques such as global differential privacy, secure multi-party computation, or encryption protocols.[38] Even with these protections, overfitting may increase the risk of membership inference attacks, which could expose patient identities.[38]

The current available studies mostly compare the accuracy of a CTG interpretation using AI with expert opinions and studies that analyze CTG result from different machine learning algorithms. The studies that discuss the usage of AI in interpretating CTG on a real time clinical setting are still limited, thus the effectiveness of AI for assisting a diagnosis is yet to be determine. Studies that cover the limitations and negative impact of AI in interpreting a CTG are also rare to be found, as the new technology is still developing. AI promises great potential in assisting clinicians at assessing fetal condition through CTG interpretation, therefore future research on this topic is greatly encouraged.

Future studies should prioritize the use of multicenter and prospectively collected datasets to improve generalizability and reduce institutional bias. External validation using diverse populations is essential to assess the robustness of AI models under different clinical conditions. Researchers are also encouraged to incorporate temporal and contextual variables such as stage of labor, maternal comorbidities, and medication use into training datasets to make models more clinically relevant. In addition, future work should emphasize developing interpretable and user-friendly AI systems that can be seamlessly integrated into real-time obstetric monitoring workflows. Collaborative research involving clinicians, data scientists, and ethicists will be crucial to ensuring that AI adoption in obstetrics is both effective and ethically responsible.

CONCLUSION

Artificial intelligence holds strong potential to improve the accuracy and consistency of CTG interpretation. Though not without challenges, current models have shown performance comparable to expert obstetricians. Rather than replacing clinicians, AI should serve as a supportive tool to enhance clinical decision-making during labor.

REFERENCES

1. Schmidt J V, McCartney PR. History and development of fetal heart assessment: a composite. J Obstet Gynecol neonatal Nurs.2000;29(3):295-305.
2. Heelan L. Fetal monitoring: creating a culture of safety with informed choice. *J Perinat Educ*. 2013;22(3):156-165.
3. Ayres-de-Campos D, Spong CY, Chandraharan E, Panel FIFMEC. FIGO consensus guidelines on intrapartum fetal monitoring: Cardiotocography. Int J Gynecol Obstet [Internet]. 2015;131:13–24. Available from: https://doi.org/10.1016/j.ijgo.2015.06.020
4. Nzelu O, Pereira S, Chandraharan E. Human factors. The dirty Dozen in CTG misinterpretation. Global Journal of Reproductive Medicine. 2018 Oct 11;6(2)
5. Rei M, Tavares S, Pinto P, Machado AP, Monteiro S, Costa A, *et al*. Interobserver agreement in CTG interpretation using the 2015 FIGO guidelines for intrapartum fetal monitoring. Eur J Obstet Gynecol Reprod Biol. 2016;205:27–31.
6. Garabedian C, Butruille L, Drumez E, Servan Schreiber E, Bartolo S, Bleu G, *et al*. Inter-observer reliability of 4 fetal heart rate classifications. J Gynecol Obstet Hum Reprod. 2017;46:131–5.
7. Alfirevic Z, Devane D, Gyte GML. Continuous cardiotocography (CTG) as a form of electronic fetal monitoring (EFM) for fetal assessment during labour. Cochrane database Syst Rev. 2006;(3):CD006066.
8. Nazli I, Korbeko E, Dogru S, Kugu E, Sahingoz OK. Early Detection of Fetal Health Conditions Using Machine Learning for Classifying Imbalanced Cardiotocographic Data. Diagnostics (Basel, Switzerland). 2025;15.
9. McCoy JA, Levine LD, Wan G, Chivers C, Teel J, La Cava WG. Intrapartum electronic fetal heart rate monitoring to predict acidemia at birth with the use of deep learning. Am J Obstet Gynecol [Internet]. 2025;232:116.e1-116.e9. Available from: https://doi.org/10.1016/j.ajog.2024.04.022
10. Athavale R, Gutiérrez VB, Jha S. AI in medicine: an introduction to the potential benefits and challenges, and why doctors need to be involved. Obstet Gynaecol [Internet]. 2024;26:177–82. Available from: https://doi.org/10.1111/tog.12950
11. Grivell RM, Alfirevic Z, Gyte GML, Devane D. Antenatal cardiotocography

for fetal assessment. Cochrane database Syst Rev. 2015;2015:CD007863.

12. Gumilar KE, Wardhana MP, Akbar MIA, Putra AS, Banjarnahor DPP, Mulyana RS, *et al*. Artificial intelligence-large language models (AI-LLMs) for reliable and accurate cardiotocography (CTG) interpretation in obstetric practice. Comput Struct Biotechnol J. 2025;27:1140–7.
13. Ben M'Barek I, Jauvion G, Ceccaldi PF. Computerized cardiotocography analysis during labor - A state-of-the-art review. Acta Obstet Gynecol Scand. 2023;102:130–7.
14. Salini Y, Mohanty SN, Ramesh JVN, Yang M, Chalapathi MM V. Cardiotocography Data Analysis for Fetal Health Classification Using Machine Learning Models. IEEE Access. 2024;12:26005–22.
15. Islam SFN, Yulita IN. Predicting Fetal Condition from Cardiotocography Results Using the Random Forest Method. In EAI; 2020.
16. Latif A, Siti Khotimatul Wildah. Optimization of Random Forest Model with SMOTE for Fetal Health Classification Based on Cardiotocography . J Teknol DAN OPEN SOURCE [Internet]. 2025;8:390–7. Available from: https://ejournal.uniks.ac.id/index.php/JTOS/article/view/4360
17. Ricciardi C, Amato F, Tedesco A, Dragone D, Cosentino C, Ponsiglione AM, *et al.* Detection of Suspicious Cardiotocographic Recordings by Means of a Machine Learning Classifier. Vol. 10, Bioengineering. 2023.
18. Xie W, Cai P, Hu Y, Lu Y, Chen C, Cai Z, *et al*. AI-driven paradigm shift in computerized cardiotocography analysis: A systematic review and promising directions. Neurocomputing [Internet]. 2024;607:128446. Available from: https://www.sciencedirect.com/science/article/pii/S0925231224012177
19. Miyata K, Shibata C, Fukunishi H, Hemmi K, Kinoshita H, Hirakawa T, *et al.* Cardiotocography-Based Experimental Comparison of Artificial Intelligence and Human Judgment in Assessing Fetal Asphyxia During Delivery. Cureus. 2025;17:e78282.
20. Sato I, Hirono Y, Shima E, Yamamoto H, Yoshihara K, Kai C, *et al*. Comparison and verification of detection accuracy for late deceleration with and without uterine contractions signals using convolutional neural networks. Front Physiol. 2025;16:1525266.
21. Sriram A, Kumar P, Lanka A. Autonomous Fetal Distress Detection with Cardiotocogram Classification using Auto-Encoder Model Combined with Recurrent Neural Networks. Nanotechnol Perceptions. 2024;20:222–33.
22. Zhao Z, Zhu J, Jiao P, Wang J, Zhang X, Lu X, *et al*. Hybrid-FHR: a multi-modal AI approach for automated fetal acidosis diagnosis. BMC Med Inform Decis Mak. 2024;24:19.
23. Melaet R, de Vries IR, Kok RD, Guid Oei S, Huijben IAM, van Sloun RJG, *et al*. Artificial intelligence based cardiotocogram assessment during labor. Eur J Obstet Gynecol Reprod Biol. 2024;295:75–85.
24. Park CE, Choi B, Park RW, Kwak DW, Ko HS, Seong WJ, *et al*. Automated interpretation of cardiotocography using deep learning in a nationwide multicenter study. Sci Rep [Internet]. 2025;15:19617. Available from: https://doi.org/10.1038/s41598-025-02849-4
25. Dhanda SS, Panwar D, Lin CC, Sharma TK, Rastogi D, Bindewari S, *et al*. Advancement in public health through machine learning: a narrative review of opportunities and ethical considerations. J Big Data [Internet]. 2025;12:154. Available from: https://doi.org/10.1186/s40537-025-01201-x
26. O'Sullivan ME, Considine EC, O'Riordan M, Marnane WP, Rennie JM, Boylan GB. Challenges of Developing Robust AI for Intrapartum Fetal Heart Rate Monitoring. Front Artif Intell. 2021;4:765210.
27. Wang AX, Le VT, Trung HN, Nguyen BP. Addressing imbalance in health data: Synthetic minority oversampling using deep learning. Comput Biol Med [Internet]. 2025;188:109830. Available from: https://www.sciencedirect.com/science/article/pii/S0010482525001805
28. Huang T, Xu H, Wang H, Huang H, Xu YJ, Li B, *et al*. Artificial intelligence for medicine: Progress, challenges, and perspectives. Innov Med. 2023;1:100030.
29. Jiang L, Wu Z, Xu X, Zhan Y, Jin X, Wang L, *et al*. Opportunities and challenges of artificial intelligence in the medical field: current application, emerging problems, and problem-solving strategies. J Int Med Res. 2021;49:3000605211000157.
30. Francis F, Luz S, Wu H, Stock SJ, Townsend R. Machine learning on cardiotocography data to classify fetal

outcomes: A scoping review. Comput Biol Med [Internet]. 2024;172:108220. Available from: https://www.sciencedirect.com/science/article/pii/S0010482524003044

31. Khaliq A, Ahmad F, Rehman HU, Alanazi SA, Haleem H, Junaid K, *et al*. Revolutionizing medical imaging: A cutting-edge AI framework with vision transformers and perceiver IO for multi-disease diagnosis. Comput Biol Chem [Internet]. 2025;119:108586. Available from: https://www.sciencedirect.com/science/article/pii/S1476927125002476
32. Islam S, Yulita I. Predicting fetal condition from cardiotocography results using the Random Forest Method. Proceedings of the Proceedings of the 7th Mathematics, Science, and Computer Science Education International Seminar, MSCEIS 2019. 2020.
33. Salini Y, Mohanty SN, Ramesh JV, Yang M, Chalapathi MM. Cardiotocography data analysis for fetal health classification using Machine Learning Models. IEEE Access. 2024;12:26005–22.
34. Nazli I, Korbeko E, Dogru S, Kugu E, Sahingoz OK. Early detection of fetal health conditions using machine learning for classifying imbalanced Cardiotocographic Data. Diagnostics. 2025 May 15;15(10):1250.
35. Khan J, Duta I, Albert B, Cooke W, Vatish M, Jones G. The OxMat dataset: a multimodal resource for the development of AI-driven technologies in maternal and newborn child health. 2024.
36. Bai J, Kang X, Wang W, Yang Z, Ou W, Huang Y, *et al*. A multimodal model in the prediction of the delivery mode using data from a digital twin-empowered labor monitoring system. Digit Heal. 2024;10:20552076241304936.
37. Noor AA, Manzoor A, Mazhar Qureshi MD, Qureshi MA, Rashwan W. Unveiling Explainable AI in Healthcare: Current Trends, Challenges, and Future Directions. WIREs Data Min Knowl Discov [Internet]. 2025;15:e70018. Available from: https://doi.org/10.1002/widm.70018
38. Topaloglu MY, Morrell EM, Rajendran S, Topaloglu U. In the Pursuit of Privacy: The Promises and Predicaments of Federated Learning in Healthcare. Front Artif Intell [Internet]. 2021;4:746497. Available from: http://www.frontiersin.org