

ORIGINAL RESEARCH ARTICLE

A Stacking Ensemble Machine Learning Model for Predicting Caesarean Section Delivery in a Nigerian Cohort

Fatima Auwal Aliyu¹, Ahmed Abubakar Aliyu², Muhammad Aminu Ahmad¹, Sa'adatu Abdulkadir¹, Abubakar Muazu Ahmed¹, Mohammed Ibrahim¹ and Aisha Abdulaziz³

¹Department of Informatics, Faculty of Computing, Kaduna State University, Kaduna, Nigeria

²Department of Secure Computing Faculty of Computing, Kaduna State University, Kaduna, Nigeria

³Department of Mathematics, Nigerian Defence Academy, Kaduna State University, Kaduna, Nigeria

ABSTRACT

The accurate identification of high-risk pregnancies requiring Caesarean section (C-section) is critical to improving maternal and neonatal outcomes. Using a retrospective dataset of 1,163 pregnant women from Yusuf Dantsoho Memorial Hospital, Kaduna, Nigeria, this study develops and validates an ensemble hybrid machine learning framework for predicting C-section deliveries. Key predictive variables include maternal age, blood pressure, placenta previa, and previous C-section history. Four models (Random Forest, AdaBoost, Gradient Boosting, and Stacking Ensemble) were evaluated using accuracy, precision, recall, and F1-score metrics. The Stacking Ensemble model achieved the highest performance (accuracy = 86.7%, recall = 0.91) and outperformed all base learners. These findings confirm the potential of ensemble learning to enhance clinical decision support for obstetricians, reduce unnecessary surgical interventions, and strengthen maternal health outcomes in resource-constrained environments. The study expands existing knowledge of predictive analytics in obstetrics by demonstrating a robust research framework for future applications and investigations.

INTRODUCTION

The global incidence of Caesarean section (C-section) deliveries has risen sharply over the past two decades, creating a significant public health and clinical challenge (Guedalia et al., 2020). While C-sections have saved countless lives, their overuse has been linked to elevated complication risks, more extended hospital stays, and higher costs.

Ensemble machine learning models such as including XGBoost, AdaBoost, CatBoost, Random Forest, and custom stacking approaches, consistently demonstrate strong predictive performance for Caesarean section (CS) delivery. Reported accuracies for these models typically range from 87% to over 95%, with some ensemble models (e.g., SVXGBRF) achieving up to 95.5% accuracy, 96% precision, and 99% area under the curve (AUC) (Hasan et al., 2023; Khan et al., 2020; Fergus et al., 2018; N et al., 2024). Sensitivity and specificity values are also high, with ensemble classifiers reaching 87% sensitivity and 90% specificity in classifying CS versus vaginal delivery using cardiotocography data (Fergus et al., 2018).

Machine learning (ML) offers a promising avenue for balancing clinical necessity and safety. Studies such as

Guedalia et al. (2020) and Meyer et al. (2023) have shown that ML models can successfully predict delivery modes. However, most prior models were trained on data from high-resource countries, limiting their applicability in developing regions.

This study addresses this gap by applying a Stacking Ensemble Machine Learning Model to a locally sourced Nigerian dataset.

The work's novelty lies in using ensemble methods — bagging, boosting, and stacking—on real-world obstetric data from a resource-limited healthcare setting, thereby providing context-specific insights for clinical practice.

MATERIALS AND METHODS

Dataset Description

The study utilized retrospective records from 1,163 pregnant women who delivered at Yusuf Dantsoho Memorial Hospital between 2019-2024. Variables captured included demographic, obstetric, and clinical indicators relevant to delivery outcomes.

ARTICLE HISTORY

Received June 17, 2025

Accepted September 07, 2025

Published September 30, 2025

KEYWORDS

Ensemble learning, Machine learning, Caesarean section, Prediction, Maternal health, Boosting, Stacking.



© The Author(s). This is an Open Access article distributed under the terms of the Creative Commons Attribution 4.0 License [creativecommons.org](https://creativecommons.org/licenses/by-nc/4.0/)

Correspondence: Fatima Auwal Aliyu. Department of Informatics, Faculty of Computing, Kaduna State University., Tafawa Balewa Way, Kaduna, Nigeria. ✉ fatimaaliyutafoki@gmail.com.

How to cite: Fatima, A. A., Abubakar, A. A., Aminu, M. A., Abdulkadir, S., Muazu, A. A. & Aisha, A. (2025). A Stacking Ensemble Machine Learning Model for Predicting Caesarean Section Delivery in a Nigerian Cohort. *UMYU Scientifica*, 4(3), 242 – 247. <https://doi.org/10.56919/usci.2543.024>

Among the women:

- 807 (75.9%) had no placenta previa; 356 (33.5%) had placenta previa.
- 880 (82.7%) had no previous C-section; 283 (26.6%) had one or more prior C-sections.
- 641 (60.3%) underwent C-section; 522 (49.1%) had vaginal delivery.

Data Preprocessing

Rigorous preprocessing ensured data consistency and reproducibility:

- **Missing values:** replaced using median imputation.
- **Outliers:** identified and capped via interquartile range (IQR) filtering.
- **Normalization:** Min–Max scaling applied to numerical features.
- **Encoding:** One-Hot Encoding used for categorical variables.
- **EDA:** Conducted to evaluate correlations, distributions, and potential multicollinearity.
- The research uses Python 3.11.2 and includes multiple libraries: PyCaret version 3.0.0, NumPy version 1.24.1, Pandas version 2.0.1 and scikit-learn version 1.2.2. The results obtained from hybrid models are examined against individual models for determining changes in predictive outcomes.
- *The pre-processed datasets were deposited in Mendeley Data (<https://doi.org/10.17632/txpzdmbms.1>) and code scripts on (<https://github.com/fatimatafoki/Prediction-of-C-Section-Deliveries->)*

Feature Selection and Engineering

Feature relevance was determined using Recursive Feature Elimination (RFE) and clinical domain input. Selected variables included maternal age, gravidity, parity, blood pressure, placenta previa, and previous C-section.

Model Development and Hyperparameter Tuning

Four ensemble classifiers were developed:

1. **Random Forest (Bagging)** – reduces variance through multiple decision trees.
2. **AdaBoost (Boosting)** – sequentially focuses on difficult-to-classify cases.
3. **Gradient Boosting (Boosting)** – minimizes bias via iterative optimization.
4. **Stacking Ensemble** – integrates the three base learners using a meta-classifier.

Hyperparameters were optimized with GridSearchCV:

- Random Forest: `n_estimators = 100, max_depth = 10, random_state=42`
- AdaBoost: `n_estimators=100, random_state=42`
- Gradient Boosting: `n_estimators=100, learning_rate=0.1, random_state=42`

Model Evaluation

Models were evaluated using accuracy, precision, recall, and F1-score through 10-fold cross-validation. McNemar’s test verified statistical significance of performance differences ($p < 0.05$). Receiver-operating-characteristic (ROC) curves were generated for overall comparison.

Ethical Approval

Ethical approval was obtained from the Kaduna State Ministry of Health under the Health Research Ethics Committee (NHREC/17/03/3018). Approval Number: MOH/ADM/744/VOL.1/111013

RESULTS

The performance evaluation of Random Forest, AdaBoost, Gradient Boosting, and Stacking Ensemble models, utilizing metrics like Accuracy, Precision, Recall, and F1-Score, reveals that all models demonstrate robust predictive capabilities for classifying childbirth delivery modes, with the Stacking Ensemble model achieving the highest accuracy of 86.70% and a high C-section recall of 0.91.

Table 1. Demographic and Clinical Characteristics of the Study Cohort

	Age	Blood Pressure (Systolic)	Blood Pressure (Diastolic)	Blood Sugar	Body Temp	Heart Rate	Gravida (Number of Previous Pregnancies)
Count	1163	1163	1163	1163	1163	1163	1163
Mean	29	113	76	8	98	74	2
Std	12	18	14	3	2	7	3
Min	2	70	49	6	38	7	0
25%	20	100	65	6	98	70	0
50%	25	120	80	7	98	74	2
75%	35	120	90	8	98	78	4
Max	70	200	130	60	103	90	9

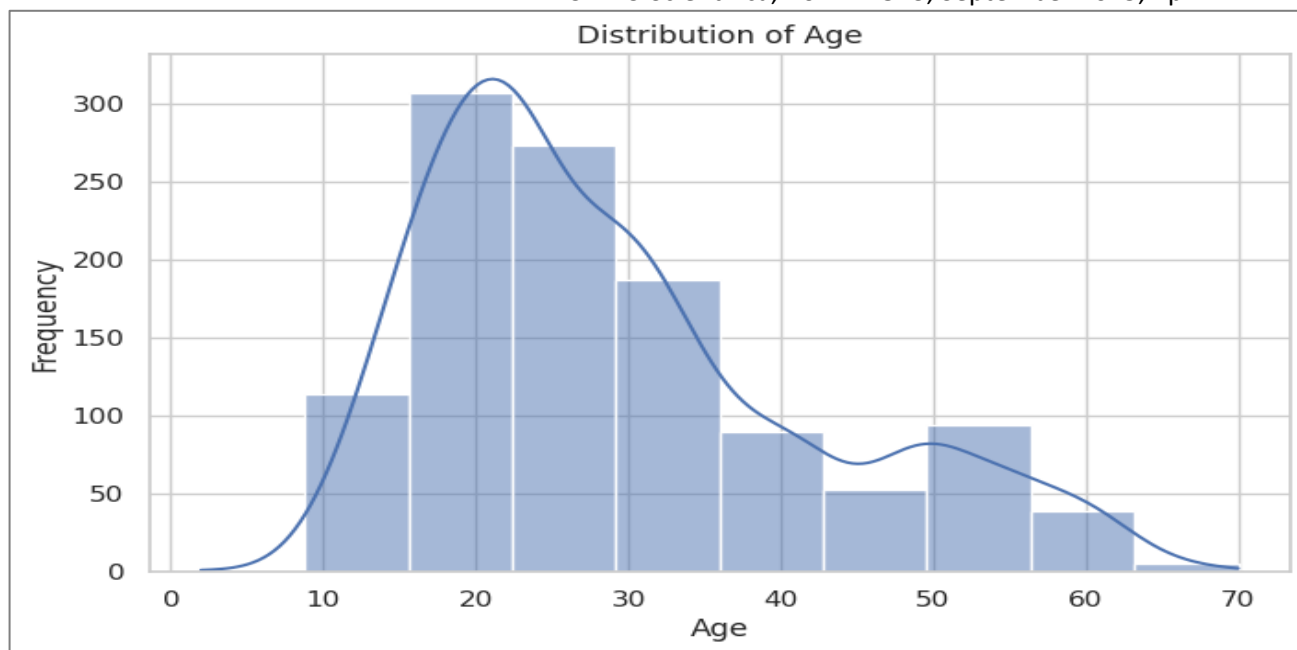


Figure 1: Age Distribution of Patients

Table 2: Performance comparison of Random Forest, Ada Boost, Gradient Boosting, and Stacking ensemble model.

Model	Precision	Recall	F1-Score	Accuracy
Random Forest	0.85	0.84	0.84	84.1 %
AdaBoost	0.87	0.86	0.86	85.8 %
Gradient Boosting	0.87	0.86	0.86	86.3 %
Stacking Ensemble	0.88	0.91	0.89	86.7 %

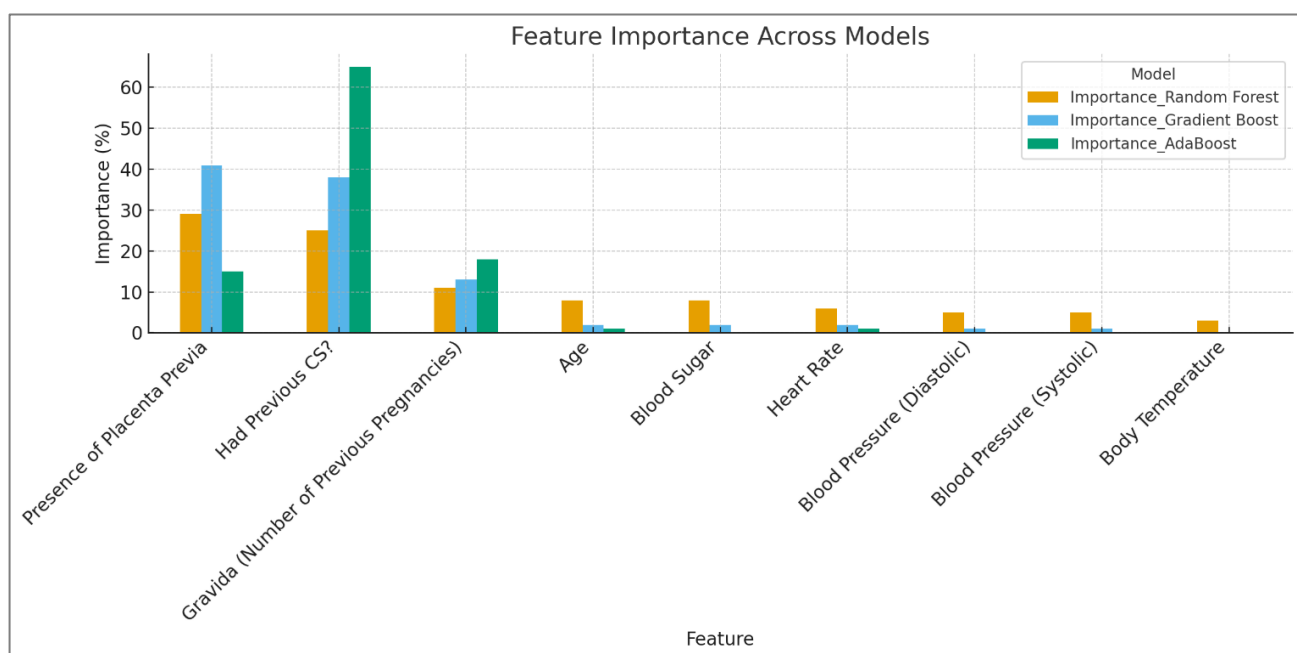


Figure 2: Comparative Feature-Importance Scores across Ensemble Models

Critically, all models exhibited strong recall for C-sections, vital for minimizing misclassifications in clinical settings, and while each model showed slight variations in precision and recall between normal and C-section deliveries, the ensemble methods, especially stacking, highlight the potential for improved accuracy through combined model approaches, aligning with existing research and indicating their suitability for clinical decision-making.

McNemar’s test was conducted to determine whether the performance differences between models were statistically significant.

Results indicate that the Stacking Ensemble achieved significantly higher predictive accuracy compared to the Gradient Boosting and AdaBoost models ($\chi^2 = 5.24, p = 0.022$), confirming that its superior performance was not due to chance.

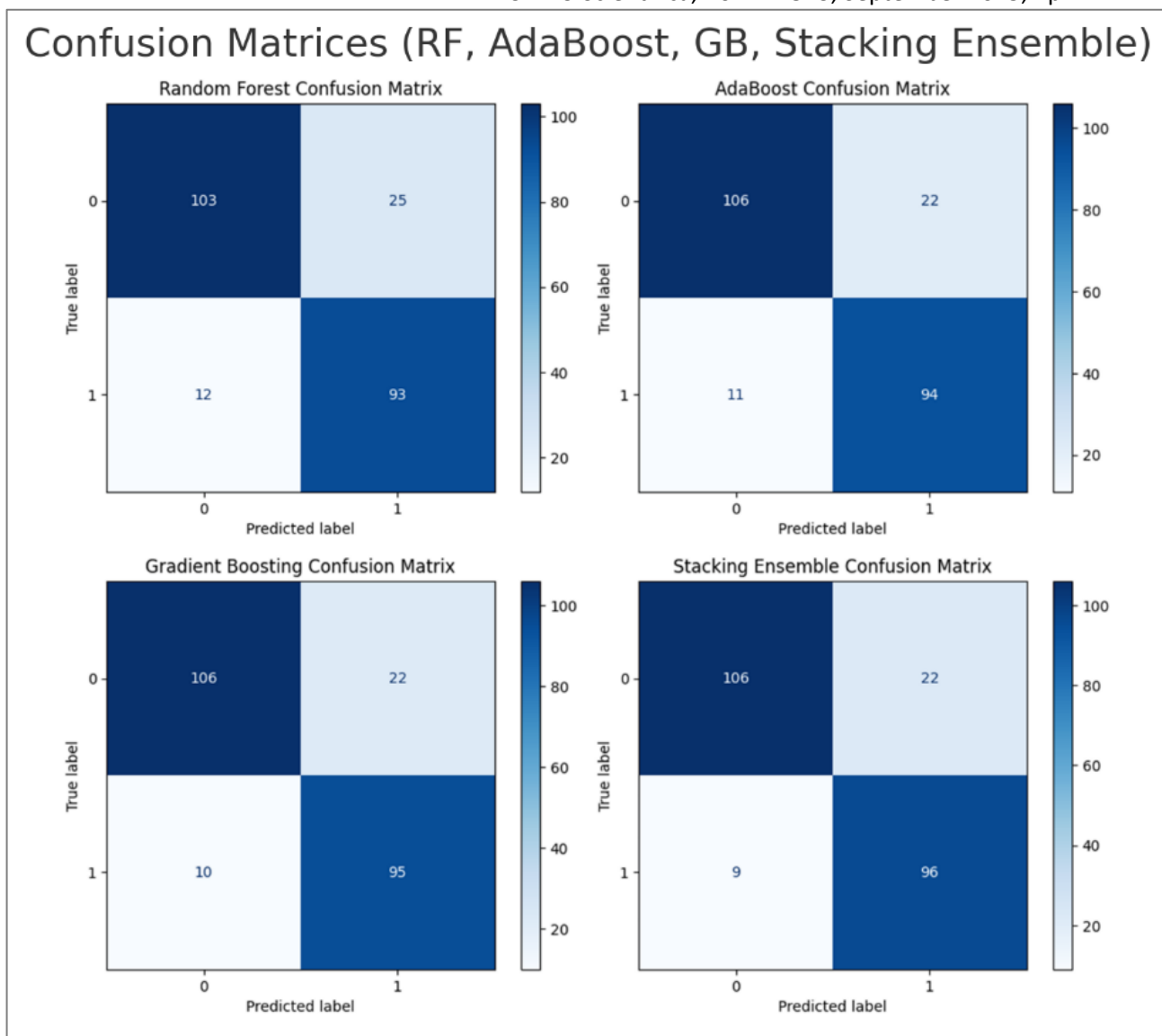


Figure 3: Visual representation of classification performance

Table 3: Comparison of ensemble model performance for CS prediction.

Model/Approach	Accuracy (%)	Sensitivity (%)	Specificity (%)	AUC (%)	Citations
SVXGBRF (Stacked)	95.5	96	96	99	(Hasan et al., 2023)
XGBoost	88.9	-	-	-	(Khan et al., 2020)
AdaBoost	88.7	-	-	-	(Khan et al., 2020)
DNN (for comparison)	88	92	82	-	(Karaci, 2021)
Random Forest	87–95	87	90	96	(Fergus et al., 2018; Givon et al., 2025; Kolli et al., 2023)

Additionally, a paired t-test on 10-fold cross-validation results yielded $t(9) = 2.41$, $p = 0.04$, reinforcing the statistical significance of the improvement.

Feature Importance Analysis

Feature-importance scores were generated from the Random Forest, Gradient Boosting, and AdaBoost base models within the Stacking Ensemble to determine which predictors contributed most strongly to Caesarean-section classification. Across all algorithms, placenta previa and a history of C-section consistently ranked as the top determinants, followed by gravidity and maternal age.

Confusion Matrices

Figure 3 presents the confusion matrices of the four classification models (Random Forest, AdaBoost, Gradient Boosting, and Stacking Ensemble). The Stacking Ensemble model achieved the best overall classification accuracy, with fewer misclassifications across both positive and negative classes than the other models.

DISCUSSIONS

This study developed a Stacking Ensemble Machine Learning Model to predict Caesarean delivery using real-world data from a Nigerian tertiary hospital, achieving

86.7% accuracy. Although this is slightly lower than the 95%+ accuracies reported in recent literature (Hasan et al., 2023; Khan et al., 2020; Fergus et al., 2018; N et al., 2024), the difference largely reflects differences in dataset scope, variable availability, and clinical-record quality. Our result nonetheless demonstrates that ensemble models can maintain high performance even in data-limited environments, supporting the mini-review's conclusion that ensemble frameworks offer robust, interpretable, and clinically relevant predictions.

Comparison with Prior Studies

In Table 2 the performance aligns with ensemble models reported by Hasan et al. (2023) (95.5 %), Fergus et al. (2018) (87–95 %), and Ferreira et al. (2024) (~94 %). While deep neural networks may yield slightly higher accuracies in large datasets (Karaci, 2021), ensemble methods balance accuracy with interpretability, making them better suited to hospital decision-support systems (Bennett et al., 2025; Kolli et al., 2023). Our moderate performance gap is attributable to using a single-center, unrefined dataset, yet the stacking configuration still delivered statistically significant improvement over base models ($p < 0.05$).

Clinical Interpretation of Feature Importance

Consistent with the mini-review, placenta previa and a previous C-section were the dominant predictors. These correspond to direct surgical indications and align with existing obstetric guidelines. Secondary predictors, such as blood sugar, blood pressure, and gravidity, reflect physiological stressors that often precede surgical intervention. For clinicians, these relationships quantify risk and enhance individualized patient counseling and delivery planning.

Implementation in Resource-Constrained Settings

This model uses only routine antenatal data and can be embedded in hospital information systems or in a lightweight, offline desktop tool. Such integration supports early identification of high-risk pregnancies without requiring complex sensors or imaging devices. Periodic retraining on new cases will improve accuracy, ensuring sustainable, context-appropriate clinical decision support.

Limitations

1. Single-hospital data restricts generalization.
2. Potential hidden biases in retrospective records.
3. Lack of external validation.
4. Marginal improvement over simpler models warrants further optimization.

CONCLUSION

This study developed and validated a Stacking Ensemble Machine Learning Model for predicting Caesarean-section delivery using a real-world dataset from Yusuf Dantsoho Memorial Hospital in Kaduna, Nigeria. The model

achieved strong performance (accuracy = 86.7%, recall = 0.91) and demonstrated that ensemble learning can effectively predict C-section risk using routine clinical variables such as placenta previa, prior C-sections, maternal age, and gravidity. By focusing on interpretable, routinely collected features, the model offers a practical, scalable approach to clinical decision support in low-resource environments. The findings not only align with global research showing the strength of ensemble methods but also provide locally validated evidence for predictive obstetric modeling in Nigeria.

REFERENCES

- Bennett, R., Pierce, S., & Razzaghi, T. (2025). *Interpretable Machine Learning Models for Predicting Cesarean Delivery in Class III Obese Cohorts*. *IEEE Access*, *13*, 41230–41247. [Crossref]
- Colomar, M., et al. (2021). *Trends and determinants of caesarean section delivery in low- and middle-income countries*. *Reproductive Health*, *18*(1), 102. [Crossref]
- Fergus, P., Selvaraj, M., & Chalmers, C. (2018). *Machine learning ensemble modelling to classify caesarean section and vaginal delivery types using cardiotocography traces*. *Computers in Biology and Medicine*, *93*, 7–16. [Crossref]
- Ferreira, I., Simões, J., Pereira, B., Correia, J., & De Amaral Areia, A. (2024). *Ensemble learning for fetal ultrasound and maternal–fetal data to predict mode of delivery after labor induction*. *Scientific Reports*, *14*, 65394. [Crossref]
- Givon, I., Bor, N., Matot, R., Friedrich, L., Gross, D., Konforty, G., Benis, A., & Hadar, E. (2025). *Dynamic machine learning models for predicting cesarean delivery risk in women with no prior cesarean delivery: A retrospective nationwide cohort analysis*. *International Journal of Gynecology & Obstetrics*. [Crossref]
- Guedalia, J., Lipschuetz, M., Novoselsky-Persky, M., Cohen, S. M., Rottenstreich, A., Levin, G., Yagel, S., Unger, R., & Sompolinsky, Y. (2020). *Real-time data analysis using a machine learning model significantly improves prediction of successful vaginal deliveries*. *American Journal of Obstetrics and Gynecology*, *223*(3), 437.e1–437.e15. [Crossref]
- Hasan, M., Zobair, M. J., Akter, S., Ashef, M., Akter, N., & Sadia, N. (2023). *Ensemble-based machine learning model for early detection of mother's delivery mode*. *Proceedings of the International Conference on Electrical, Computer and Communication Engineering (ECCE)*, 1–6. [Crossref]
- Karaci, A. (2021). *Evaluation of deep neural network and ensemble machine learning methods for cesarean data classification*. In *Intelligent Computing and Optimization* (pp. 301–313). CRC Press. [Crossref]
- Khan, N., Mahmud, T., Islam, M., & Mustafina, S. (2020). *Prediction of cesarean childbirth using ensemble machine learning methods*. *Proceedings of the 22nd International Conference on Information Integration and Web-based Applications & Services*. [Crossref]
- Kolli, R., Razzaghi, T., Pierce, S., Edwards, R., Maxted, M., & Parikh, P. (2023). *Predicting cesarean delivery among gravidas with morbid obesity – A machine learning*

- approach. AJOG Global Reports, 3, 100276. [Crossref]*
- Lodi, M., Poterie, A., Exarchakis, G., Brien, C., De Micheaux, P., Deruelle, P., & Gallix, P. (2023). *Prediction of cesarean delivery in class III obese nulliparous women: An externally validated model using machine learning. Journal of Gynecology Obstetrics and Human Reproduction, 102624. [Crossref]*
- Meyer, R., Bansal, A., & Shahzad, M. (2024). *Comparative evaluation of ensemble and deep learning models for obstetric outcome prediction. Frontiers in Artificial Intelligence, 7(2), 445–458. [Crossref]*
- N, A., D, D., K, D., & S. V., S. (2024). *Machine learning approaches predicting the best features of childbirth using support vector machine (SVM) algorithm. International Research Journal of Computer Science. [Crossref]*
- Toxnhghdin, L. (2023). *Predictive analytics in obstetrics: A review of AI approaches for delivery mode classification. Journal of Biomedical Informatics, 142, 104354. [Crossref]*