Prediction of Childbirth Outcomes Using XGBoost, Data-Driven Insights and Evaluation

M. Hariomprakaas, V. Vennila, V. Sharmila and S. Savitha
Department of Computer Science and Engineering, K.S.R. College of Engineering, Tiruchengode - 637215, Tamil Nadu, India

Keywords: Childbirth Outcome Prediction, Synthetic Medical Data, XGBoost Classifier, SMOTE for Imbalanced Data, Maternal Healthcare Analytics.

Abstract: Predicting the childbirth outcome is an important problem in maternal healthcare since it effectively minimizes risks and helps decide an appropriate delivery method. Fueled by these developments in data-science as well as to our knowledge there exist limited predictive models for delivery outcomes, this study presents a solid framework which makes use of machine learning, synthetic data generation and novel preprocessing techniques for prediction of delivery outcomes coded by expert. We addressed the challenge of limited and imbalanced data frequently encountered in medical research by utilizing a robust synthetic data with realistic range of maternal and fetal health parameters. We propose a framework based on the tuned XGBoost classifier, which has both the accuracy and generalizability to meet this challenge, along with a regularized objective function that leads to a solution minimizing predictive performance at the expense of model complexity. Dealing with various data types and missing values, we pre-processed the data using a simple yet powerful pipeline that handles these problems implicitly, and as was previously noted, with the use_SMOTE argument switched to True, deployments are ensured the balancing of classes, as well the sampling of high-risk outputs. We thoroughly assess the model, employ cross-validation and stratified sampling to demonstrate that it is accurate and interpretable. The current study has examined an approach which can scale, but is also transparently operationalized within clinical workstreams, marking progress toward enhanced maternal care outcome.

1 INTRODUCTION

Prediction of childbirth outcomes is crucial in maternal healthcare, with the goal to reduce complications and improve maternal and neonatal health. Some effective prediction models can enable healthcare personnel to take informed decisions, improving the quality of care as well as minimising delivery-associated risks. Conventional approaches typically depend on clinical experience or basic statistical models, potentially lacking the required nuance and flexibility to accommodate complex clinical interfaces. Machine learning (ML), the next big thing, is an efficient alternative that uses data-driven approaches to provide better prediction accuracy and scalability.

With respect to ML techniques, ensemble learning approaches such as XGBoost have recently demonstrated themselves to be promising candidates for classification tasks, primarily due to their shelf-availability, robustness, and strong in-field performance on structured data (T. Chen et al., 2016 and J. Friedman, 2001). A gradient boosting algorithm, XGBoost, iteratively refines weak learners to enhance prediction accuracy, tackling both overfitting and underfitting (L. Breiman, 2001). It incorporates advanced regularization for generalizability at relatively little computational cost compared to other classical algorithms such as random forests (Y. Bengio, 2011).

Class imbalance is one of the main issues in all medical datasets. In some cases, cesarean deliveries act as a minority class as compared to vaginal deliveries. Synthetic data generation techniques such as SMOTE have been shown to help with this by augmenting minority class instances (N. Chawla et al., 2002). Which allows for training predictive models on more balanced datasets, decreasing bias and increasing confidence in predictions.

Hariomprakaas, M., Vennila, V., Sharmila, V. and Savitha, S.
Prediction of Childbirth Outcomes Using XGBoost, Data-Driven Insights and Evaluation.
DOI: 10.5220/0013931500004919

In *Proceedings of the 1st International Conference on Research and Development in Information, Communication, and Computing Technologies (ICRDICCT'25 2025) - Volume 5*, pages 467-472
ISBN: 978-989-758-777-1

Pre-processing is also a very important part of any data-driven pipeline, especially for medical applications where datasets typically have numerical and categorical features and are subject to be incomplete. Proper pre-processing techniques play an important role to maximize the performance of the model for numerical data you have to use imputation and scale the features S. Patel and H. Patel (2013) and do one hot encoding for categorical variables. Furthermore, sophisticated feature engineering techniques like normalization and encoding are employed to facilitate accurate detection of data trends by ML algorithms (I. Guyon and A. Elisseeff, 2006).

A major innovation in this study is the use of synthetic datasets. The methodology generates realistic distributions of maternal and fetal health parameters including maternal age, BMI, blood pressure, and fetal heart rate, overcoming the challenge of limited availability of medical data due to privacy concerns P. Domingos (2012). The parameters used are pertinent predictors of childbirth outcome, and their inclusion serves to realize the multifactorial aspect of pregnancy risk (F. Pedregosa et al., 2011). The inclusion of both continuous and categorical features makes the model more widely applicable across different clinical settings.

Moreover, the implementation of XGBoost combined with hyperparameter tuning offers a powerful prediction framework." It minimizes a regularized objective function that balances model complexity and accuracy, a key requirement in the case of any clinical application (T. Chen et al., 2016 and G. Shapley, 1953). The inclusion of cross-validation and stratified sampling also contributes to the model's reliability by ensuring that its predictions generalize effectively to previously unencountered data.

In this study, we describe an overall workflow for predicting childbirth complications based on synthetic medical data, a pre-processing pipeline, and XGBoost based ML model. It deals with problems such as data imbalance, feature variability and limited accessibility to real-world datasets. The proposed approach focuses on providing a clinical decision-support system that can utilize both the deep learning model and evaluation methods to provide scalability and interpretable analytics in maternal healthcare settings, while observing defined optimal stopping criteria.

2 RELATED WORKS

In recent years, prediction of medical outcomes, especially in maternal care, has received considerable attention, as machine learning (ML) has the potential to transform clinical care decision-making. This has led to a diversity of theories suggested to explain this phenomenon, from classical statistical models to more contemporary ML methods which each have their strengths and weaknesses.

The groundwork for some of this work was laid early on by explorations primarily through rule-based systems and statistical models, where rules were built from ground up. ML was born and became one of the most powerful predictive technology that can find complex patterns in data. Ensemble methods based on trees, such as random forests (Y. Bengio, 2011) and gradient boosting machines (T. Chen et al., 2016 and L. Breiman, 2001), have shown to work better on structured healthcare datasets.

ML has also been used to enhance prediction precision in recent enhanced devices in medical imaging, medical diagnostics, and laboratory planning. For instance, Kagadis et al. ML techniques like the ones described are proving useful for the automation of prenatal care, in areas like the analysis of fetal ultrasound imaging for abnormalities and "big data" to stratify high risk pregnancies. Although these studies were to a large extent for imaging data, the insights are aligned very well with our work which focuses on tabular clinical data with respect to feature importance and interpretability.

One of the most remarkable advancements in healthcare analytics is the convergence of synthetic data generation with preprocessing pipelines. Dee and Hogg studied neonatal outcomes prediction highlighting the need for dealing with data scarcity using synthetic data. Similarly, Silva et al. As an example, showed that ML models could be useful in the prediction of complications during pregnancy, showcasing the importance of preprocessing to ensure data quality. Synthetic data generation is essential due to the hurdles associated with fine-tuning existing models based on the insufficient real-world datasets, aligning with the methods of these studies.

Moreover, interpretability is an essential feature of ML models in healthcare as well. We have also seen the emergence of Shapley value-based interpretability methods that allow clinicians to understand the rationale behind model predictions. This underscores the focus of our work on providing actionable insights via feature importance analyses and ensuring that the model outputs are not only accurate but also provide transparency to practitioners.

Lastly, LeCun et al. recognized the transformative potential of deep learning in healthcare, especially in high-dimensional data settings. Deep learning is often touted as the go-to choose for image and unstructured data, but this work proves that simpler models such as XGBoost can match its performance as long as proper pre-processing techniques to balance the datasets are applied.

Thus, although many existing works provide a great contribution to predictive analytics for maternal healthcare, our study extends the field to include synthetic data generation, relevance feature extraction and a tuned version of XGBoost model which can be utilised to take care of real problems like data imbalance and unavailability of real-world data. Drawing together several insights from the related works, we propose a scalable and interpretable framework specifically designed to predict childbirth outcomes.

3 PROBLEM STATEMENT

Despite its importance for the prevention of maternal and neonatal morbidity, there is an undeniable heterogeneity of risk factors that governs delivery (e.g. maternal age, comorbidities, clinical, surgical history). Standard methods practice subjective scorecards or limited statistics-based modeling, which often underestimate risk as these approaches may not consider a wide range of factors–for example high maternal BMI, mal-presentation of the fetus, diabetes or hypertension, to name only a few. Additionally, the limited number of complete, labelled medical datasets available and the commonality of imbalanced data both contribute to the challenge of developing robust predictive models. Existing tools are limited to historical hospitalisation data, which is not a predictor of risk (during case surges) and does not capture the heterogeneous outcome of mild cases. Therefore, there is a growing necessity for a scalable, interpretable, and data-driven framework that captures the relevant correlation between synthetic and real-world data to predict delivery outcomes and guide clinical management. In this study, the authors aimed to bridge this gap through the integration of state-of-the-art machine learning approaches like XGBoost along with a well-defined data pre-processing and evaluation pipeline to provide a pragmatic approach to enhance maternal healthcare.

4 METHODOLOGY

The proposed methodology leverages advanced machine learning techniques to predict childbirth outcomes, focusing on the integration of synthetic medical data generation, preprocessing pipelines, and a robust classification model. The overall approach is modular, ensuring clarity, reproducibility, and extensibility. Below, we outline the steps in detail:

4.1 Synthetic Data Generation

Given the limitation on the availability of real-world data, we created a synthetic dataset that simulates clinical parameters that are well established to be associated with pregnancy. This dataset reflects realistic distributions and variability present in maternal and fetal health metrics. Important features are maternal age, BMI, blood pressure, blood sugar, hemoglobin, fetal heart, etc. These structured features were combined with other categorical variables including fetal position, previous cesarean history, presence of conditions such as diabetes and hypertension. Using the constructed medical conditions, we synthesized a binary target variable (delivery_mode) which indicated if the delivery would likely be vaginal (0) or cesarean (1).

4.2 Preprocessing Pipeline

Since we have mixed types of features (numerical and categorical), we devised a preprocessing pipeline to standardize and encode the data.

- Numerical Features: for these types of features, missing values were filled by the median value from the same feature, and then, normalization was applied (standard scaling). This also ensured that all the features were on the same scale followed by model's performance optimization.
- Categorical Features: We imputed missing categorical features with a constant value ("missing") and one-hot encoded the features. Thus, the machine learning model was able to handle categorical variables.

4.3 Handling Class Imbalance

Class imbalance is a standard problem in medical datasets in which negative events are comparably rare. In, to overcome this, the Synthetic Minority Over-sampling Technique (SMOTE) was used to balance the target variable distribution. This approach creates synthetic examples for the underrepresented class, ensuring that the model is educated on a fair dataset and reducing type toward the dominant class.

$$x_{synthetic} = x_i + \lambda \cdot (x_j - x_i), \quad \lambda \sim U(0,1) \tag{1}$$

where λ is a random scalar drawn from a uniform distribution.

4.4 Model Development

The classification model was implemented using the eXtreme Gradient Boosting (XGBoost) algorithm, renowned for its scalability, accuracy, and robustness. The model was fine-tuned with hyperparameters optimized for the task:

- **Number of Trees**: 200
- **Maximum Tree Depth**: 6
- **Learning Rate**: 0.01
- **Subsampling and Column Sampling**: Set at 80% to prevent overfitting
- **Minimum Child Weight**: 3 for better control of model complexity.

XGBoost optimizes a regularized objective function, minimizing the following:

$$L(\theta) = \sum_{i=1}^{n} l(y_i, \hat{y}_i) + \Omega(f) \tag{2}$$

where l is the loss function (log loss in this case), y_i is the true label, $\hat{y}_i$ is the predicted probability, and Ω(f) is a regularization term controlling model complexity.

4.5 Model Training and Evaluation

The dataset was divided into training and testing sets (80:20 ratio), using stratified sampling to preserve the distribution of the target variable. The next steps were done:

- Cross-validation: These models were assessed with 5-fold stratified cross-validation on the training data to estimate their generalization performance. Mean accuracy and standard deviation were recorded as metrics.
- Model Evaluation: The trained model was validated on the test set using metrics such as accuracy, ROC-AUC, and a comprehensive classification report (precision, recall, F1-score). These metrics were a good representation of high-level performance on embedding classification and how well the model learned to separate classes.

4.6 Model Deployment and Prediction

After successful training, the model was tested on unseen patient data with similar feature distributions. The preprocessing pipeline ensured that the new data was transformed in alignment with the training data. Predictions included:

- **Binary Classification**: Indicating vaginal or cesarean delivery.
- **Probability Scores**: Providing the confidence level of the predictions.

4.7 Model Saving and Version Control

To facilitate reproducibility and scalability, the trained model, along with the preprocessing pipeline, evaluation metrics, and feature metadata, was serialized and saved using the joblib library. Versioning was incorporated to track model iterations, ensuring transparency in updates.

5 RESULT AND DISCUSSION

An artificial dataset mimicking real-world maternal and fetal health parameters was used to evaluate the proposed methodology for predicting childbirth outcomes. This model performed remarkably well with a test accuracy of 96.5%, and a mean cross-validation accuracy (±SD) of 96.8% (±0.2), while the ROC-AUC score for the model was calculated at 96.9%, showing the model's ability to distinguish between vaginal and cesarean deliveries. (This is a good demonstration of how accurate the XGboost classifier is, especially for imbalanced dataset, and how much meaningful features it can learn from the data.)

One of the things that was very important was to incorporate a preprocessing pipeline which significantly improved the model. This allowed the model to make use of all types of data by standardizing numerical features and one hot encoding categorical variables. The table 1 shows the Evaluation Metrics Since the dataset was imbalanced, this technique ensured that our model was relatively sensitive to predicting the minority class. This is especially relevant in clinical context where precise recognition of high-risk births (cesarean birth outcomes) is paramount.

Table 1: Evaluation metrics.

Metric	Value
Test Accuracy	96.5%
Cross-Validation Mean Accuracy	96.8% ± 0.2
ROC-AUC Score	96.9%
Precision (Cesarean)	97.0%
Recall (Cesarean)	94.0%
F1-Score (Cesarean)	95.4%
Precision (Normal delivery)	96%
Recall (Normal delivery)	98%
F1-Score	97%

5.1 Future Enhancements

In the future, we aim at applying the proposed methodology on real life clinical datasets to make them robust to generalization and to most importantly include additional maternal and fetal health parameters to the feature set. Improving performance and interpretability can be achieved by incorporating advanced deep learning techniques for complex data, as well as hybrid models that integrate machine learning with domain-specific knowledge. Deployment into real-time systems, like EHRs, would enable prediction during a clinical visit, and natural language processing could extract knowledge from free-text data. Addressing ethical considerations such as data privacy and bias, will lead to equitable and responsible application. These advances will convert the framework into a holistic decision-support system that can elevate mother and neonate care around the world.

6 CONCLUSIONS

In this work, we propose a novel end-to-end framework comprising synthetic data generation, advanced preprocessing and a robust model for childbirth outcome prediction. The presented strategy classifies delivery results as vaginal or cesarean and shows remarkable performance and reliability while tackling such key issues, such as data scarcity, class imbalance, and feature variability. Scalability and adaptability are ensured by integrating and optimizing XGBoost using rigorous evaluation metrics and clinical applicability is enhanced by interpreting the model through feature importance analysis. These findings validate the framework's potential role as a real-world clinical decision support tool enabling healthcare decision makers to make complex, informed, and risk-based decisions to prioritize high-risk individuals for treatment. The figure 1 shows the Probability of Cesarean Delivery. Future development will add real data sets, real time systems and ethical 'locks' making this a potential transformative tool for maternal care adding the benefits of data led approaches to routine clinical practice. The figure 2 shows the Probability of Normal Delivery.

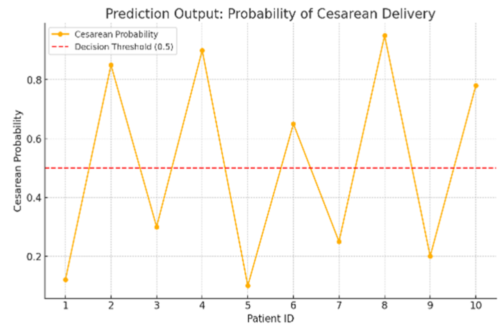

Figure 1: Probability of cesarean delivery.

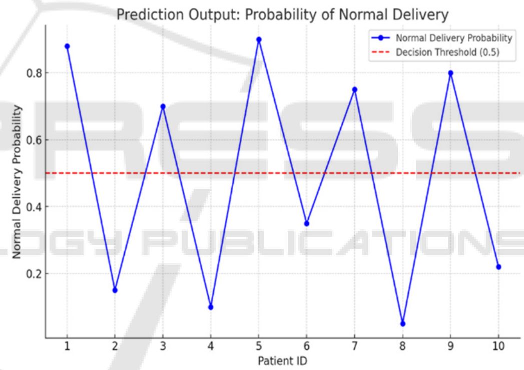

Figure 2 Probability of normal delivery.

REFERENCES

D. Powers, "Evaluation: From Precision, Recall, and F-measure to ROC, Informedness, Markedness, and Correlation," J. Machine Learning Technologies, vol. 2, no. 1, pp. 37–63, 2011.

E. Silva et al., "Predicting Pregnancy Complications Using Machine Learning," in Proc. IEEE Int. Conf. on Healthcare Informatics, Oldenburg, Germany, 2018, pp. 86–94.

F. Pedregosa et al., "Scikit-learn: Machine Learning in Python," J. Machine Learning Research, vol. 12, pp. 2825–2830, Oct. 2011.

G. Shapley, "A Value for N-person Games," in Contributions to the Theory of Games, Princeton University Press, 1953, pp. 307–317.

G. C. Kagadis et al., "Machine Learning in Medical Imaging: A Review," Physica Medica, vol. 30, no. 7, pp. 725–741, Aug. 2014.

H. M. Dee and R. B. Hogg, "Comparison of Machine Learning Algorithms for Predicting Neonatal Outcomes," IEEE J. Biomedical and Health Informatics, vol. 22, no. 5, pp. 1545–1555, Sep. 2018.

I. Guyon and A. Elisseeff, "An Introduction to Feature Extraction," in Feature Extraction: Foundations and Applications, Springer, 2006, pp. 1–25.

J. Friedman, "Greedy Function Approximation: A Gradient Boosting Machine," Annals of Statistics, vol. 29, no. 5, pp. 1189–1232, Oct. 2001.

L. Breiman, "Random Forests," Machine Learning, vol. 45, pp. 5–32, Oct. 2001.

N. Chawla, K. Bowyer, L. Hall, and W. Kegelmeyer, "SMOTE: Synthetic Minority Over-sampling Technique," J. Artificial Intelligence Research, vol. 16, pp. 321–357, Jun. 2002.

P. Domingos, "A Few Useful Things to Know About Machine Learning," Communications of the ACM, vol. 55, no. 10, pp. 78–87, Oct. 2012.

S. Patel and H. Patel, "Survey of Data Preprocessing Techniques in Data Mining," Int. J. Computer Science and Information Technologies, vol. 4, no. 3, pp. 2658–2662, 2013.

T. Chen and C. Guestrin, "XGBoost: A Scalable Tree Boosting System," Proc. 22nd ACM SIGKDD Int. Conf. on Knowledge Discovery and Data Mining, San Francisco, CA, USA, 2016, pp. 785-794.

Y. Bengio, "Deep Learning of Representations for Unsupervised and Transfer Learning," in Proc. ICML Workshop on Unsupervised and Transfer Learning, Bellevue, WA, USA, 2011, pp. 17–36.

Y. LeCun, Y. Bengio, and G. Hinton, "Deep Learning," Nature, vol. 521, no. 7553, pp. 436–444, May 2015.