Predicting the Most Suitable Delivery Method for Pregnant Women by Using the KGC Ensemble Algorithm in Machine Learning

Pusarla Sindhu, Parasana Sankara Rao
Department of Computer Science and Engineering
GITAM (Deemed to be)University, Visakhapatnam-530045, India

***Abstract*—Maternal and neonatal mortality rates pose a significant challenge in healthcare systems worldwide. Predicting the childbirth approach is essential for safeguarding the mother's and child's well-being. Currently, it is dependent on the judgment of the attending obstetrician. However, selecting the incorrect delivery method can cause serious health complications both in mother and child over short-time and long-time. This research harnesses machine learning algorithms' capability to automate the delivery method prediction process. This research studied two different stackings implemented in machine learning, leveraging a dataset of 6157 electronic health records and a minimal feature set. Stack1 consisted of k-nearest neighbors, decision trees, random forest, and support vector machine methods, yielding an F1-score of 95.67%. Stack 2 consisted of Gradient Boosting, k-nearest neighbors, and CatBoost methods, which yielded 98.84%. This highlights the superior effectiveness of its integrated methodologies. This research enables obstetricians to ascertain the delivery method promptly and initiate essential measures to ensure the mother's and baby's safety and well-being.**

Keywords*—*Delivery method; stacking; neonatal mortality; KGC ensemble algorithm

I. INTRODUCTION

Maternal and neonatal mortality can be significantly reduced by carefully selecting the most appropriate delivery method based on the mother's health, pregnancy complications, and fetal conditions [1]. Mode of childbirth has become a significant issue for obstetricians, health authorities, and mothers. Over the current years, there has been a global rise in cesarean delivery rates, contrary to the World Health Organization's guidelines advocating for rates below 15% [2]. However, determining an appropriate cesarean delivery rate and the impact it has on maternal and infant well-being remains a topic of debate within the obstetrics community [3]. Progressive medicalization can be attributed partly to the ever-rising rates of cesarean sections [4].

Medical technology advancements such as elective or planned Caesarean, emergency cesarean section, forceps delivery, and vacuum extraction have resulted in a safer experience during childbirth. In a cesarean delivery, also known as a C-section, surgeons perform a laparotomy (an abdominal incision in the mother) followed by a hysterotomy (an incision in the uterus) to facilitate the birth of the baby [5]. Broadly, experts classify cesarean sections into three categories: elective or planned, emergency, and cesarean on demand [6].

Obstetricians opt for cesarean delivery or C-section if the fetus inside the mother's womb is in an unusual position, has very little amniotic liquid, possessing multiple fetuses or many other reasons. Choosing the wrong delivery technique may impose immediate and long-term health complications on both mother and child. Seeking the help of technology can help obstetricians make decisions accurately. Machine learning can guide the obstetrician in predicting the probable chances of the type of delivery, which can educate the mother on a safer mode of birth.

The elective cesarean is a prepared delivery when maternal or fetal indications arise in the antepartum period opted before the ongoing labor. On the other hand, medical professionals perform an emergency cesarean when they observe symptoms that emerge during labor, requiring immediate medical intervention. Healthcare providers perform a cesarean on demand when women specifically request it, as indicated by its name. Cesarean delivery is performed for patients in conditions like when the mother possesses more than one fetus when the baby is in a breech position, when the mother has severe health conditions like diabetes or pre-eclampsia, when labor does not progress any further, when elderly primigravida or dystocia or maternal HIV occurs, C-sections can be a life savior. Assisted deliveries like forceps and vacuum extraction can provide extra support during childbirth for the protection of both mother and baby. Forceps provide traction on the baby's head and vacuum extraction using suction to assist delivery.

A well-chosen birthing technique-whether vaginal birth, cesarean section, or other interventions can help avoid potential risks, ensuring safer childbirth outcomes. The gynecologist chooses the delivery method by evaluating various obstetric characteristics like the number of fetuses and the medical history of the pregnant woman, such as diabetes and blood pressure. Personalized care, guided by healthcare professionals, ensures that the delivery method aligns with maternal and fetal needs, reducing the likelihood of complications and improving overall survival rates for mothers and newborns.The gynecologist selects the delivery method by evaluating various biological factors of the mother, such as age, medical history, and other relevant health conditions.

In this research, the proposed algorithm is an ensemble of K-Nearest Neighbors (KNN), CatBoost (CB), and Gradient Boosting (GB). The algorithms have been wisely chosen for their complementary strengths. KNN is helpful for less noisy datasets because of its simplicity and interoperability, while CatBoost handles categorical data efficiently, even on imbalanced datasets.GB excels in capturing complex relationships and delivering high predictive accuracy. The stack of these classifiers offers flexibility to explore a diverse learning approach to optimize model performance with hyperparameter tuning.

This research focuses on tackling the existing literature gap by developing a predictive model that integrates multiple clinical features to define the mode of childbirth. The approach utilizes a stacked ensemble of ML algorithms to optimize performance through hyperparameter tuning, providing more accurate predictions. The novelty of this study lies in diverse classifiers, which significantly enhance the model's performance metrics.

The main contributions of this research include

- A detailed literature survey on the type of algorithms used in the recent studies.
- Model can handle large datasets with more features accurately.
- The study employs three classifier algorithms stacked to ensure more precise and reliable decisions.

The paper is structured as follows: Section II provides a review of related literature, Section III presents the methods and materials used, Section IV describes the experimental process Section V analyzes the results obtained, and finally, Section VI concludes the study and suggests potential future research directions.

II. Literature Survey

Recent studies have witnessed the growing importance of machine learning by various authors for pregnancy outcomes and optimizing medical decisions.Fernández et al. [3] investigated algorithms like support vector machines(SVM), Random forest(RF), and the Multilayer Perceptron (MLP) to forecast the delivery type among three classes, namely C-section, euthocic vaginal, and instrumental deliveries. HGSORF, which applies the Henry Gas Solubility Optimization algorithm with Random Forest was designed to predict C-sections, demonstrating the potential of decision-making [7]. Khan et al. [8] investigated how machine learning techniques transform gynecological healthcare, aiming to enhance diagnostics and the challenges faced. Lestari et al. [9] conducted a comprehensive review on anticipating pregnancy-related complications. Tiruneh et al. [10] organized a broad review to compare Pre-Eclampsia Prediction using regression models and machine learning algorithms.

Islam et al. [11] performed a systematic review on ML uses to predict pregnancy outcomes, identifying gaps in the existing approaches and proposing a research agenda for future work.Kolasa et al. [12] have done a diverse review of the usage of ML algorithms in health care. Mas-Cabo et al. [13] studied algorithms like Multilayer Perceptron(MLP) and Artificial Neural Network(ANN) to forecast anticipated labor in women with early labor symptoms through the analysis of electrohysterogram (EHG) signals of the uterus. ANN was employed to estimate the success of labor induction, analyzing the uterine EHG signals [14]. The authors achieved different objectives with ML, such as classifying placenta cells [15], developing pregnancy disorder in the first-trimester prediction [16], and evaluating cesarean delivery risk in term nulliparous.

XGBoost (Extreme Gradient Boosting) has been extensively used to solve predictive modeling tasks by constructing a series of decision trees to rectify previous errors and avoid overfitting. To illustrate, Sultan [17] experimented with different algorithms to identify the most suited algorithm to classify cesarean section deliveries. Xi [18] worked towards predicting the large gestational age neonates in parturients exposed to radiation using machine learning.Yu [19] evaluated using machine learning algorithms for preterm birth forecast in singleton pregnancies through time-series data [20].

Stacking has been popular machine learning, widely used in various applications like the detection of thyroid diseases [21], software bug prediction [22] predicting childbirth approach, preterm birth prediction [23] used for various applications. Islam et al. [4] focused on the features suited best for the prediction of the delivery technique using algorithms like RF, SVM, Decision Tree (DT), K-Nearest Neighbors (KNN), and stacking classifier (SC). Yang and Shami [24] explored hyperparameter optimization across most machine learning algorithms, emphasizing its critical role in enhancing predictive performance.

III. Methods and Materials

In this study, we assessed the appropriateness of utilizing various ML algorithms for anticipating the mode of childbirth across six classifications: CES Programmed(Elective Cesarean), Emergency Cesarean, Eutocic delivery, Forceps delivery, vacuum extraction, and Epistiomy.

The dataset comprised instances of women presented for childbirth at four public hospitals across three distinct autonomous regions in Spain in 2014 [25]. One hundred and sixty-one personal health and medical features were recorded from each mother and the fetus; few details were gathered in antepartum and remaining in intrapartum.

A. Dataset Description

The Target variable Type of birth was noted as the significant result of the labor. The Target Variable has been classified as one of the classes CES Programmed, Elective Cesarean, Eutocic delivery, Forceps delivery, vacuum extraction, and Epistiomy. Medical circumstances surrounding the mother and fetus were given utmost priority in deciding the type of delivery technique for their well-being.

B. Data Preprocessing

This process is crucial for ensuring the accuracy and effectiveness of the models built using the data and preprocessing, including data reduction, preparation, and balancing techniques. The dataset contained 6157 data records with 161 attributes. Data reduction is a crucial preprocessing approach that can decrease the size and complexity of a dataset while preserving its essential information. It involves removing or merging redundant or irrelevant features, removing noise and outliers, and transforming the data into a more convenient representation.

Data reduction aims to make the dataset more manageable and accessible without sacrificing important patterns or relationships. Dimensionality reduction helps machine learning algorithms function more effectively and predict better. We discarded features with minimal or no impact (e.g. OLIGOAMNIOS, Isoimmunization, MIOMECTOMY, and many more). Other aspects, such as the sex of the fetus and fetal admission to the ICU, are only identifiable after the delivery process. Islam and his colleagues conducted structured interviews with 111 features to obtain information on their relevance [4]. The weighted average score, which ranges from 1 to 5, was calculated across 111 features by averaging each score and considering their assigned weight. The importance or significance of the feature can determine the weight. Thirty-two features scored above 1.5 on average, while 79 scored below 1.5. Setting a threshold of 1.5, the initial set comprised 32 features, with the remaining 79 considered less significant. Table I contains the 32 attributes of the dataset selected after preprocessing.

Data reduction was achieved by removing 18 duplicate records from the dataset. We replaced missing values for the numerical features with their mean, and the mode was substituted for categorical features. Data preparation involves initializing input data to make it suitable for effective analysis by algorithms. The MinMaxScaler transforms the dataset's features, allowing each feature to have an equal impact on the predictive capability by scaling to an interval of 0 to 1. Subsequently, the normalize function applies L2 normalization, which modifies the feature vectors to have a unit norm, improving the model's capacity to handle different magnitudes across the data.

The dataset suffered from a strong inequality between the superior class Eutocic and the minority class Epistiomy. To mitigate overfitting and prevent the model from favoring the majority class excessively in terms of accuracy and frequency over the minority class, random oversampling was employed to balance the class distribution until achieving a 1:1 ratio. Random oversampling duplicates the samples from the minority class and introduces them to the superior class to achieve a balanced representation of majority and minority samples in the dataset.

C. Algorithms

1) K-Nearest Neighbors: This non-parametric algorithm assigns labels or predicts values depending on the majority class of its closest neighbors in the feature domain.

2) Gradient Boosting: This machine learning ensemble technique constructs predictive models by merging weak learners, which are decision trees, into a strong ensemble model.

3) CatBoost: Categorical boosting, also known as CatBoost, stands out for its efficient handling of categorical features. This algorithm automatically handles categorical variables without extensive preprocessing, making it convenient for real-world datasets with common categorical features.

IV. Methodology

The algorithms were created, trained, and evaluated using Jupyter Notebook, a scientific program development and open-source development platform implemented in Python utilizing the Scikit-learn package. From the data, 80% is randomly utilized for the trained model, and the remaining 20% is employed for testing and evaluating each model.

A. Experimental Setup

Each algorithm is tuned for the hyperparameters using Grid search with five-fold cross-validation and accuracy as the performance metric for tuning the hyperparameters. Through an exploratory research approach, the hyperparameters have been selected after testing over a broad range of configurations. This gives the flexibility in identifying optimal parameters based on the dataset's performance rather than predetermined theoretical values.

1) K-Nearest Neighbors: Several configurations of the KNN algorithm were used as it doesn't build the model explicitly during training. A general rule doesn't exist that predicts the optimal value of the parameters, which in turn is dependent on the dataset characteristics and must be found empirically. A grid search is employed as a tuning technique to identify the optimal values of the hyperparameters. Grid Search has been applied to the K-Nearest Neighbors (KNN) algorithm, tuning the hyperparameters across the following ranges: the number of neighbors (n_neighbors) from 1 to 20, the weighting metrics ("uniform" and "distance"), and the distance metrics ("Euclidean", "Manhattan", "Minkowski").

- n_neighbors: This parameter dictates the number of neighboring data points examined during prediction, influencing the model's decision boundary flexibility. A smaller "n_neighbors" value can make the model more susceptible to noise, potentially causing overfitting. Conversely, a larger "n_neighbors" value can make the model too generalized, possibly overlooking local patterns in the data.
- metric: This parameter defines the distance metric used to measure the distance between points in the feature space. Various distance metrics lead to varied notions of proximity between points, which can affect the performance of the K-nearest neighbors algorithm.
- distance: The weights parameter steers the decision of data point classification in KNeighborsClassifier. Exploring various weight configurations can boost the model's effectiveness, mainly when working with imbalanced datasets or fluctuating feature significance.

The best parameters the classifier finds with grid search for this algorithm are {"n_neighbors": 5, "metric": "manhattan", "weights": "uniform"}.

2) Gradient Boosting: Grid Search has been applied to the algorithm, tuning the hyperparameters across the following ranges considering the number of estimators from 10 to 200, learning rate from 0.01 to 1.0, and maximum depth of trees from 1 to 20.

- n_estimators (Number of Estimators): This parameter dictates the number of boosting trees built in the ensemble. A higher number of trees generally leads to a more expressive model, potentially capturing intricate patterns in the data.

TABLE I. FEATURES OF THE DATASET

Feature	Description	Type
AGE	Age of the individual	Numerical
ALCOHOL	Alcohol consumption during pregnancy	Categorical
AMNIOCENTESIS	A medical procedure involving the extraction of amniotic fluid for various diagnostic purposes, such as genetic testing.	Categorical
AMNIOTIC LIQUID	Characteristics of amniotic fluid	Categorical
ANESTHESIA	Administration of anesthesia during childbirth or related medical procedures.	Categorical
ART	Presence of assisted reproductive technology in the conception of the pregnancy.	Categorical
ART MODE	Methods to Assisted Reproductive Technology	Categorical
BMI	Body Mass Index of the individual	Numerical
CARDIOTOCOGRAPHY	Monitoring the fetal heartbeat and uterine contractions during labor.	Categorical
COMORBIDITY	Presence of one or more additional disorders or diseases alongside the pregnancy	Categorical
COMPLICATIONS	Medical issues or difficulties during pregnancy or childbirth.	Categorical
EPISIOTOMY	A surgical incision made during childbirth to widen the opening of the vagina.	Categorical
FETAL INTRAPARTUM PH	Measurement of the acidity or alkalinity of the fetal blood during labor.	Categorical
GESTATIONAL AGE	The age of the fetus in weeks, calculated from the beginning to end of the menstrual period.	Numerical
HEIGHT	Height of the individual	Numerical
INDUCTION	Method of labor induction	Categorical
KG INCREASED PREGNANCY	Increase in weight during pregnancy	Numerical
MATERNAL EDUCATION	Level of education of the mother	Categorical
MISCARRIAGES	Number of miscarriages	Numerical
NUMBER OF PREV CESAREAN	Number of previous cesarean deliveries	Numerical
OBSTETRIC RISK	Risk factors associated with pregnancy	Categorical
OXYTOCIN	Use of oxytocin during labor	Categorical
PARITY	The count of a woman giving birth to a fetus crossing the gestational age of 24 weeks or more.	Numerical
PREINDUCTION	Usage of medical interventions to initiate labor before it starts spontaneously.	Categorical
PREVIOUS CESAREAN	Indication of whether the mother had a cesarean section in a previous pregnancy.	Categorical
PREVIOUS PRETERM PREGNANCIES	Number of previous preterm pregnancies	Numerical
PREVIOUS TERM PREGNANCIES	Number of previous full-term pregnancies	Numerical
ROBSON GROUP	Robson classification group	Categorical
SMOKING	Smoking status during pregnancy	Categorical
START ANTENATAL CARE	Timing of initiation of medical care and attention during pregnancy.	Categorical
SUBSTANCE ABUSE	History of harmful substances (e.g., drugs or alcohol) during pregnancy.	Categorical
WEIGHT	Weight of the individual	Numerical

- learning_rate (Learning Rate): The learning rate controls the contribution of each tree to the final prediction. A higher learning rate speeds convergence but risks overshooting, while a lower rate requires more iterations but enhances generalization.
- max_depth (Maximum Depth of Trees): This parameter determines the maximum depth allowed for each tree in the ensemble. Deeper trees can capture more detailed data features but may also result in overfitting as they learn the noise in the training data.

The hyperparameters by the grid search for this algorithm are “n_estimators”: 90, “learning_rate”: 0.5, “max_depth”: 10.

3) CatBoost: A systematic exploration of different parameter combinations was conducted to optimize the training of CatBoost models involving varying key hyperparameter ranges, namely, the number of estimators (“n_estimators”) from 10 to 200, the depth (“depth”) from 1 to 16, and the learning rate (“learning_rate”) from 0.01 to 1.0.

- n_estimators (Number of Estimators): This defines the number of trees (boosting iterations) to be built during training. Increasing the number of estimators can lead to a more complex model, potentially improving performance, but it may also increase training time and the risk of overfitting
- learning_rate (Learning Rate): The learning rate controls the step size at each iteration during the gradient descent optimization process. It adjusts the model weights in response to the error gradient.
- max_depth (Maximum Depth of Trees): This parameter sets the maximum depth allowed for each tree in the ensemble. Deeper trees can capture more complex relationships in the data but may also lead to overfitting.

The hyperparameters by the grid search for this algorithm are “n_estimators”: 100, “learning_rate”: 0.5, “max_depth”: 10.

4) Stacking: An ensemble learning method combines to generate a new training set for a meta-classifier based on the predictions of multiple classifiers. Each classifier is trained on the entire training set individually, and the meta-classifier is learned from the predictions made by the base models. Fig. 1 illustrates the stacking classifier’s architecture. The initial training data (X) had 6157 samples and 2 features. Three M different models (M = 3) are trained on X, and their predictions (y) are combined to generate a data set X^2 for the level 2 model. A strong SC was proposed in which KNN, GB, and CB with the hyperparameters were the base classifiers.

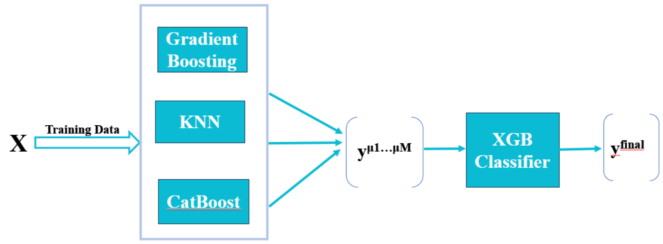

Fig. 1. Proposed KGC ensemble architecture.

5) Proposed Algorithm: KGC Ensemble Algorithm: A KGC ensemble algorithm has been developed using the stacking of KNN, GB, and CatBoost algorithms with the hyper-parameters found in the GridSearch. This algorithm excels beyond the standard ML algorithms with default parameters (Stack 1). The primary goal of the stacking is to minimize data variance and optimize its suitability for machine learning models.

K-Nearest Neighbors (KNN) starts the process by classifying data points based on the proximity to other points in the dataset. In KNN, each data item is assigned to the class frequency among k nearest numbers where k is predetermined.

$$h(x) = \text{majority\_label}\left(\{y_i \mid x_i \in N_k(x)\}\right)$$

where:

- x is the data point to classify,
- $N_k(x)$ represents the k-nearest neighbors of x,
- and *majority_label* denotes the most frequent label among these neighbors.

The algorithm classifies based on the majority class contribution among neighbors, effectively performing similarity-based classification.

Further, the dataset undergoes similar training under Gradient Boosting (GB). GB builds an ensemble of decision trees that sequentially trains every tree to rectify the errors of its predecessor. The process continuously minimizes a loss function, making the model more accurate. Prediction is a weighted sum of the predictions of all trees, and the function h can be expressed as:

$$h(x) = \sum w_i f_i(x)$$

where:

- $f_i(x)$ are the individual decision trees,
- and w_i are the weights assigned to each tree's prediction.

Finally, CatBoost, an advanced boosting algorithm, is applied. CatBoost builds an ensemble of trees, similar to GB, but uses ordered boosting and permutation techniques to ensure unbiased and robust learning. The hypothesis function h in CatBoost can be represented as:

$$h(x) = \sum \alpha_i T_i(x)$$

where:

- $T_i(x)$ are the individual trees,
- and α_i are the weights assigned to each tree's output.

The ensemble of these models KNN, GB, and CatBoost-work to improve the overall performance, reduce variance, and provide a robust solution for classification tasks. The pseudocode of the ensemble is shown in Algorithm 1.

Algorithm 1 KGC Ensemble Algorithm:

Input: Initial training dataset $D_s = \{(x_i, y_i)\}_{i=1}^{k}$
Output: An ensemble classifier H
$B_1, B_2, \ldots, B_L$: Base classifiers
M: Meta classifier
p_{il}: Predicted values output by base classifiers B_i for data sample x_i
Step 1: for $l \leftarrow 1, 2, \ldots, L$ do
Train base classifier B_l on D_s to obtain h_l
Step 2: for each data sample x_i in D_s do
for $l \leftarrow 1, 2, \ldots, L$ do
$p_{il} = h_l(x_i)$ *(Predict using base model h_l)*
Step 3: Form the new dataset D_{new} by augmenting D_s with the predictions p_{il}
$D_{\text{new}} = \{(x_i, p_{i1}, p_{i2}, \ldots, p_{iL}, y_i)\}_{i=1}^{k}$
Step 4: Train the meta classifier M on D_{new}
return $H(x) = M(x, h_1(x), h_2(x), \ldots, h_L(x))$

6) Performance evaluation of algorithms: The following metrics are evaluated at each stage of the experimentation.

- Precision: Precision calculates the number of predictions made for a class that belongs to the class.
- Recall: Recall calculates the number of estimates made for a class amongst all the cases of that class present in the dataset
- F1-Score: The F1-score is the weighted average of precision and recall, typically ranging from 0 to 1.

V. Results and Discussion

Two stack ensembles are constructed to evaluate the performance of the different model configurations. The first stack consisted of the classifiers DT, KNN, RF, and SVM balanced using the sampling technique ADASYN. In contrast, the second stack contained KNN, GB, and Catboost, each tuned with the optimized hyperparameters with the random oversampling technique. The experimental setup involved comparing the predictive performance of both stacks on the same validation dataset, with results being analyzed to assess the overall effectiveness of the ensemble.

Results provided by the proposed stack showcase an excellent predictive capability to classify the type of delivery method correctly. Even though the individual methods performed well, the proposed algorithm combines the pitfalls of each method and showcases good results compared to the individual models. Fig. 2 represents the stack's performance comprising the models SVM, DT, KNN, and RF balanced using the ADASYN balancing technique, and the metrics of the proposed KGC Ensemble balanced using random oversampling are depicted in Fig. 3.

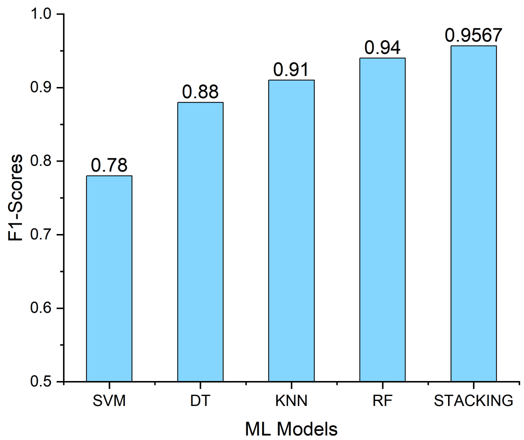

Fig. 2. Metrics of ADASYN sampling.

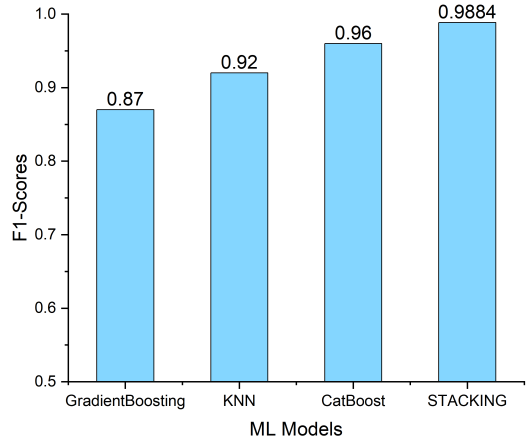

Fig. 3. Metrics of random oversampling.

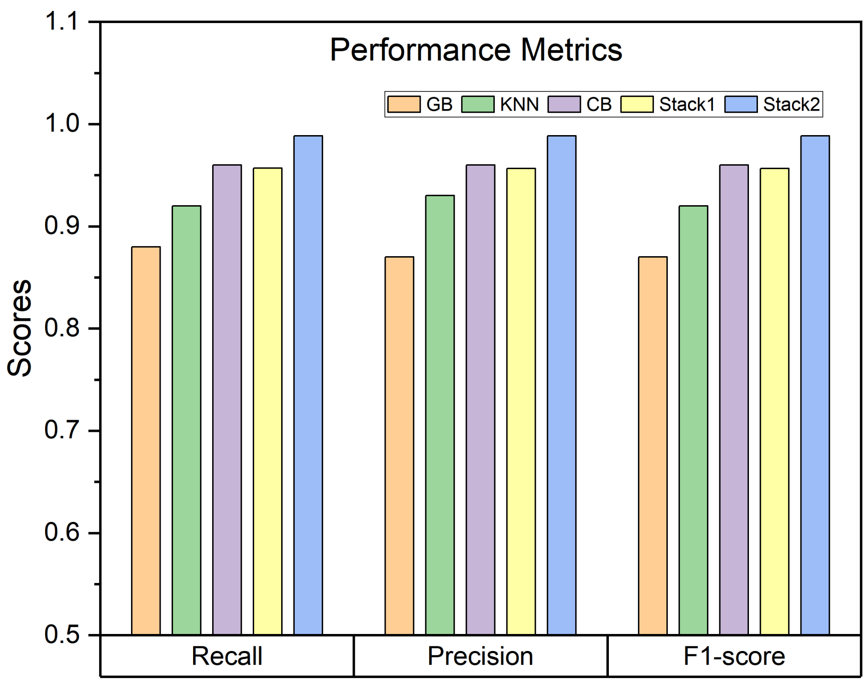

Fig. 4. Comparison of the results obtained by each stack.

Developing decision support systems is complex and aims to maximize performance measures like precision and recall to reduce false positives and negatives. While high and balanced values are ideal, clinical criteria, which can vary between hospitals, should determine which measure to prioritize. Therefore, hospital systems should be flexible, allowing clinicians to adapt protocols and prioritize specific performance measures, such as cesarean, vaginal, and assisted vaginal delivery decisions. The overall performance metrics of both stacks are represented in Table II with numerical values and in Fig. 4 through bar plots.

This study's findings exceed previous works aimed at predicting delivery methods and assessing cesarean risks, considering antepartum and intrapartum factors. In [3], a cohort of 25,038 patients with single pregnancies from the Service of

TABLE II. PERFORMANCE METRICS FOR DIFFERENT MODELS

Model	Recall	Precision	F1-Score
GradientBoosting	0.88	0.87	0.87
KNN	0.92	0.93	0.92
CatBoost	0.96	0.96	0.96
STACK1	0.957	0.9567	0.9567
STACK2	0.9885	0.9886	0.9884

Obstetrics and Gynaecology of the University Clinical Hospital evaluated the feasibility of using algorithms, namely, SVM, RF, and MultiLayer Perceptron, to predict the delivery method among cesarean, eutocic and assisted vaginal deliveries. The algorithms displayed an accuracy of 87%-90%. The study by Sultan et al. [17] and by Hasan et al. [26] on a sample containing 692 cesarean and 5465 non-cesarean samples collected from 4 hospitals in Spain worked on cesarean prediction. The SVC, XGB, and RF ensemble has achieved an F1-score of 96%. Concerning [27], a cohort of 13527 was prospectively assessed to predict cesarean deliveries. 32 classifiers were assessed, and the Quadratic discriminant analysis achieved an accuracy and F1-score of 97.9%.

The proposed KGC ensemble model consisting of KNN, Gradient Boosting, and CatBoost achieved a high F1 score of 98.84%, demonstrating its strong predictive capability. Gradient Boosting and CatBoost enhance the model's ability to handle complex, nonlinear relationships, while KNN adds the advantage of capturing local data patterns. The blend strikes a balance between bias and variance, in turn improving the model's generalization. However, the ensemble has challenges, including increased computational complexity and longer training times. Moreover, the decision-making process becomes less interpretable due to the complexity of Gradient Boosting and CatBoost, making it harder to understand how the model arrives at its predictions than KNN.

VI. CONCLUSION

An alarming increase in cesarean section rates surpassing WHO recommended levels poses significant medical, financial, and organizational challenges. Typically in low-risk pregnancies and non-elective cesarean sections, the possibility of vaginal or c-section deliveries relies on medical conditions and initiatives taken. Machine learning can act as a driving force and aid obstetricians in predicting the best feasible delivery mode based on the medical conditions, ensuring the mother's and newborn's safety. Machine learning can assist obstetricians working night shifts and in varying localities by providing real-time, data-driven insights and recommendations on the optimal delivery mode, accounting for factors such as limited staff availability, resource constraints, and patient demographics.

This research has proposed an optimized KGC classifier combining the algorithms of GB, KNN, and CatBoost along with a random oversampling balanced dataset to increase the predictive capability of the delivery mode. The performance of the proposed classifier has been tested against different performance metrics such as F-measure, recall and precision. Our proposed KGC classifier has achieved an F1-score of 98.84%, outperforming stack1 composed of KNN, DT, SVM, and RF.

In addition,KGC algorithm provides better results compared to previous studies. Hence we can conclude an optimized and random oversampled-balanced KGC classifier can reliably predict the C-section and vaginal classes.While our research has provided valuable insights,it is important to understand the dataset was limited to 6157 records. To further validate the model's robustness, future work is aimed to target the larger datasets with more intrapartum details. Incorporating such demographic variables in future studies could enhance the predictive power and applicability of the model. Further , our classifier can be applied to skin cancer, Parkinson's, gestational diabetes and socio-demographic data. In the future, we would like to work with deep learning algorithms for better predictive capabilities to implement a possible computer decision support system model to be built for the benefit of obstetricians and pregnant women.

ACKNOWLEDGMENT

The authors would like to thank ChatGPT for assisting in the proof reading.

REFERENCES

[1] Z. S. Lassi, T. Mansoor, R. A. Salam, J. K. Das, and Z. A. Bhutta, "Essential pre-pregnancy and pregnancy interventions for improved maternal, newborn and child health," *Reproductive health*, vol. 11, pp. 1–19, 2014.

[2] World Health Organization, "Medical eligibility criteria for contraceptive use." https://www.who.int/publications/i/item/WHO-RHR-15.02, 2015.

[3] A. D. R. Fernández, D. R. Fernández, and M. T. P. Sánchez, "Prediction of the mode of delivery using artificial intelligence algorithms," *Computer Methods and Programs in Biomedicine*, vol. 219, p. 106740, 2022.

[4] M. N. Islam, T. Mahmud, N. I. Khan, S. N. Mustafina, and A. N. Islam, "Exploring machine learning algorithms to find the best features for predicting modes of childbirth," *IEEE Access*, vol. 9, pp. 1680–1692, 2020.

[5] F. Parenting, "Different childbirth methods you must know." https://parenting.firstcry.com/articles/different-childbirth-methods-you-must-know/, 2020.

[6] L. Seshadri and G. Arjun, *Essentials of Obstetrics*. Wolters Kluwer India Pvt Ltd, 2015.

[7] M. S. Islam, M. A. Awal, J. N. Laboni, F. T. Pinki, S. Karmokar, K. M. Mumenin, S. Al-Ahmadi, M. A. Rahman, M. S. Hossain, and S. Mirjalili, "Hgsorf: Henry gas solubility optimization-based random forest for c-section prediction and xai-based cause analysis," *Computers in Biology and Medicine*, vol. 147, p. 105671, 2022.

[8] I. Khan and B. K. Khare, "Exploring the potential of machine learning in gynecological care: a review," *Archives of Gynecology and Obstetrics*, pp. 1–19, 2024.

[9] D. Lestari, F. I. Maulana, S. F. Persada, and P. D. P. Adi, "Machine learning for perinatal complication prediction: A systematic review," in *International Conference on Information, Communication and Computing Technology*, pp. 789–803, Springer, 2023.

[10] S. A. Tiruneh, T. T. T. Vu, D. L. Rolnik, H. J. Teede, and J. Enticott, "Machine learning algorithms versus classical regression models in pre-eclampsia prediction: A systematic review," *Current Hypertension Reports*, pp. 1–15, 2024.

[11] M. N. Islam, S. N. Mustafina, T. Mahmud, and N. I. Khan, "Machine learning to predict pregnancy outcomes: a systematic review, synthesizing framework and future research agenda," *BMC pregnancy and childbirth*, vol. 22, no. 1, p. 348, 2022.

[12] K. Kolasa, B. Admassu, M. Holownia-Voloskova, K. J. Kedzior, J.-E. Poirrier, and S. Perni, "Systematic reviews of machine learning in healthcare: a literature review," *Expert Review of Pharmacoeconomics & Outcomes Research*, vol. 24, no. 1, pp. 63–115, 2024.

[13] J. Mas-Cabo, G. Prats-Boluda, J. Garcia-Casado, J. Alberola-Rubio, R. Monfort-Ortiz, C. Martinez-Saez, A. Perales, and Y. Ye-Lin, "Electrohysterogram for ann-based prediction of imminent labor in women with threatened preterm labor undergoing tocolytic therapy," *Sensors*, vol. 20, no. 9, p. 2681, 2020.

[14] C. Benalcazar-Parra, Y. Ye-Lin, J. Garcia-Casado, R. Monfort-Ortiz, J. Alberola-Rubio, A. Perales, and G. Prats-Boluda, "Prediction of labor induction success from the uterine electrohysterogram," *Journal of Sensors*, vol. 2019, no. 1, p. 6916251, 2019.

[15] E. Rohith, V. Sowmya, and K. Soman, "Convolutional neural networks for placenta cell classification," in *2019 2nd International Conference on Intelligent Computing, Instrumentation and Control Technologies (ICICICT)*, vol. 1, pp. 1273–1277, IEEE, 2019.

[16] D. S. Maylawati, M. A. Ramdhani, W. B. Zulfikar, I. Taufik, and W. Darmalaksana, "Expert system for predicting the early pregnancy with disorders using artificial neural network," in *2017 5th International Conference on Cyber and IT Service Management (CITSM)*, pp. 1–6, IEEE, 2017.

[17] N. Sultan, M. Hasan, M. F. Wahid, H. Saha, and A. Habib, "Cesarean section classification using machine learning with feature selection, data balancing, and explainability," *IEEE Access*, vol. 11, pp. 84487–84499, 2023.

[18] X. Bai, Z. Zhou, Z. Zheng, Y. Li, K. Liu, Y. Zheng, H. Yang, H. Zhu, S. Chen, and H. Pan, "Development and evaluation of machine learning models for predicting large-for-gestational-age newborns in women exposed to radiation prior to pregnancy," *BMC Medical Informatics and Decision Making*, vol. 24, no. 1, p. 174, 2024.

[19] Q.-Y. Yu, Y. Lin, Y.-R. Zhou, X.-J. Yang, and J. Hemelaar, "Predicting risk of preterm birth in singleton pregnancies using machine learning algorithms," *Frontiers in big Data*, vol. 7, p. 1291196, 2024.

[20] Y. Zhang, S. Lu, Y. Wu, W. Hu, and Z. Yuan, "The prediction of preterm birth using time-series technology-based machine learning: retrospective cohort study," *JMIR Medical Informatics*, vol. 10, no. 6, p. e33835, 2022.

[21] G. Obaido, O. Achilonu, B. Ogbuokiri, C. S. Amadi, L. Habeebullahi, T. Ohalloran, C. W. Chukwu, E. Mienye, M. Aliyu, O. Fasawe, *et al.*, "An improved framework for detecting thyroid disease using filter-based feature selection and stacking ensemble," *IEEE Access*, 2024.

[22] P. Sindhu, G. S. Peruri, and M. Yalavarthi, "An empirically based object-oriented testing using machine learning," *EAI Endorsed Transactions on Internet of Things*, vol. 10, 2024.

[23] N. P, I. S. U, B. V, and H. R, "Preterm birth prediction using machine learning," in *2024 International Conference on Inventive Computation Technologies (ICICT)*, pp. 1–6, 2024.

[24] L. Yang and A. Shami, "On hyperparameter optimization of machine learning algorithms: Theory and practice," *Neurocomputing*, vol. 415, pp. 295–316, 2020.

[25] C. Campillo-Artero, M. Serra-Burriel, and A. Calvo-Pérez, "Predictive modeling of emergency cesarean delivery," *PloS one*, vol. 13, no. 1, p. e0191248, 2018.

[26] M. Hasan, M. J. Zobair, S. Akter, M. Ashef, N. Akter, and N. B. Sadia, "Ensemble based machine learning model for early detection of mother's delivery mode," in *2023 International Conference on Electrical, Computer and Communication Engineering (ECCE)*, pp. 1–6, 2023.

[27] M. Kowsher, N. J. Prottasha, A. Tahabilder, H. Kaiser, M. Abdur-Rakib, and M. S. Alam, "Predicting the appropriate mode of childbirth using machine learning algorithm," *International Journal of Advanced Computer Science and Applications*, vol. 12, no. 5, 2021.