

An AI-based approach to predict delivery outcome based on measurable factors of pregnant mothers.

Michael Owusu-Adjei^{1, 2}, James Ben Hayfron-Acquah,³ Abdul-Salaam Gaddafi,⁴ TwumFrimpong

¹Department of Computer Science, Kwame Nkrumah University of Science and Technology, Kumasi, Ghana.

²Department of Computer Science, Kwame Nkrumah University of Science and Technology, Kumasi, Ghana.

³Department of Computer Science, Kwame Nkrumah University of Science and Technology, Kumasi, Ghana.

⁴Department of Computer Science, Kwame Nkrumah University of Science and Technology, Kumasi, Ghana.

Corresponding author. E-mail: mowusuadjei@st.knust.edu.gh;

ABSTRACT

The desire for safer delivery mode that preserves the lives of both mother and child with minimal or no complications before, during and after childbirth is the wish for every expectant mother and their families. However, the choice for any particular birth delivery mode is determined primarily by a number of factors that feed into the ultimate decision of choice. Some of these include maternal birth history, maternal and child health conditions prevailing before and during labor onset. Predictive modeling has been used extensively to determine important contributory factors influencing delivery choice. However, missing among a myriad of features used in various research studies for this determination is maternal history of spontaneous, threatened and inevitable abortion(s) and how its inclusion can impact delivery outcomes. This research work therefore takes measurable maternal features that include real time information on administered partographs to predict delivery type. This is achieved by adopting effective feature selection technique to estimate variable relationship with the target variable. Three supervised learning techniques are used and evaluated for performance. Prediction accuracy score of area under the curve obtained show Gradient Boosting classifier achieved 91% accuracy, Logistic Regression 93% and Random Forest 91%. Balanced accuracy score for the same learning techniques obtained were; Gradient Boosting 82.73%, Logistic Regression 84.62% and Random Forest 83.02%. Correlation statistic for variable independence among input variables showed that delivery type as an output is associated with gestational age and the progress of maternal cervix dilatation during labor onset.

Keywords: Dilatation, cervix, delivery, maternal factors, models, pregnant, healthcare, antenatal care.

INTRODUCTION

It is an inalienable right for women irrespective of race, geographical region, social status, economic circumstance or ethnic orientation to a determination of choice of childbirth delivery mode. However, the right to a determination of any particular choice is largely dependent on assessment of balance of risks in individual circumstances together with eventual benefits that is to be derived from the choice made [1]. This is particularly important because, for the multiparous, a previous experience backed by delivery process history is a strong determinant for any particular delivery choice made. But for the nulliparous, overcoming the fear of uncertainties of any particular delivery outcome poses a greater challenge. Dimensions to delivery mode choice among expectant mothers vary from one to another. To the nulliparous with recurrent spontaneous abortions, the determination of delivery mode for a successful pregnancy (full term pregnancy) is of prime importance due to heightened expectations and increased anxiety [2]. This anxiety among nulliparous [3] often leads to requests for a particular delivery mode as compared to the nulliparous with no history of spontaneous abortions. Similarly, to the multiparous with particular delivery mode(s), requests for any particular delivery mode are informed by previous delivery experience and current medical conditions of both mother and child. But to the medical personnel, child birth delivery remains an outcome of a series of processes, assessments, procedures and evaluations based upon which a final determination of the appropriate delivery form is recommended and is informed by the desire to reduce and adequately manage pregnancy related complications before and during labor with the ultimate goal of preserving the lives of both mother and child. Other dimensions to childbirth delivery choice include religious and cultural beliefs and practices of expectant mothers and their families which also tend to influence the choice of particular childbirth delivery type(s)[4]. Further dimensions include the level of care and quality of interactions between healthcare providers and expectant mothers, and in this regard, World Health Organization's (WHO) guidelines on intrapartum care for a positive childbirth experience enumerate certain challenges that must be overcome. One such challenge is the medicalization of childbirth processes that undermine one's ability to give birth. Labor interventions are beyond the reach of many, increasing health equity gap thereby negatively impacting the childbirth experience [5]. Maternal care experiences and expectations, interactions with healthcare personnel, labor interventions and medicalization processes on childbirth delivery are summarized in a graphical presentation shown in Fig 1

Maternal interactions

Maternal interaction with various entities during the process of childbirth as shown in Fig 1 illustrates interactive components at play during the delivery process. Experiences with these components form the basis for maternal expectations (positive or negative). A positive feedback together with patient involvement in critical decision-making serves as an important end point for women in labor. This feedback reflects fulfilment of personal expectations including religious and socio-cultural beliefs and practices [5]. To achieve this positive feedback from interventions and personnel interactions, various research studies have examined individual levels of interactions and how these interactivities affect childbirth delivery outcome. A cohort study that compares long-term reproductive and obstetric outcome in women treated for fear of childbirth and those without this fear for all nulliparous[6] concluded that fear of childbirth among this group of women persists even in subsequent pregnancies. Additionally, the prevalence and impact of fear of childbirth and associated risk factors in another research study [7] reported incidence of fear of childbirth among nulliparous women. A focus study on labor interventions to address fear of childbirth and reduce negative birth experience [8] through a midwife-led psycho-education counseling to improve maternal and perinatal outcome proposed an evaluation of counseling interventions in clinical practice. However, a hypothetical assessment of outcome status on methods of delivery with mother's age [9] concluded that delivery outcome is independent of the maternal age. Among studies associated with delivery mode identified, maternal state (nulliparous or multiparous). Interaction related studies that examined the role of both healthcare personnel and the effectiveness of labor interventions and medication processes that meet maternal expectations for positive feedback have also been undertaken. Delivery expectations among pregnancy types remain varied and this is significantly due to differences in pregnancy circumstances. For the nulliparous with few or recurrent spontaneous abortions, childbirth delivery anxiety remains a greater challenge. It is therefore important to describe the various pregnancy types in any meaningful estimation of childbirth delivery mode for better and proper evaluation. Additionally, risk assessment factors that influence delivery type may also consider the number of successful pregnancies, number of spontaneous abortions for both multiparous and nulliparous, number of fetal deaths including still births, maternal age, gestational age and other measurable metrics with direct consequences on delivery outcome such as progress of maternal cervix dilatation over a given time interval, maternal blood pressure, fetal heart rate, maternal haemoglobin count, fetal weight and maternal temperature together with amniotic fluid index. This research work focuses on artificial intelligence-based techniques using measurable metrics such as maternal blood pressure at the onset of labor, fetal heart rate, maternal pulse rate, maternal haemoglobin count, maternal cervical dilatation count at the onset of labor over three hours interval, gestational age, fetal weight together with a number of successful pregnancies, abortions and fetal deaths, to estimate output variable independence using a unique feature selection technique called Chi-square correlation statistic test to help discover relationships between measurement of these characteristics and potential delivery type.

Supervised Machine Learning Techniques and Types:

Supervised learning is one in which labeled examples of data (input and output) are used to train an algorithm to correctly predict or classify an output label from the input data. It is predominantly useful in real-world applications such as fraud detection, spam filtering detections, disease diagnosis, financial risk assessments, sentiment analysis, image and document classifications, etc. In data mining, supervised learning can be used in two problem domains: output classification, where the output variables are categorical, such as true or false, yes or no, etc., and regression, where relationships between dependent and independent variables are investigated for their impact on the output variable. Regression enables us to answer the question of variable importance in predictive analysis. A supervised learning algorithm's task is to find an appropriate mapping function to map an input variable (independent variable) (x) into an output variable (dependent variable) (y). Supervised learning types and the various algorithms they represent are simplified in graphical representation shown in Fig 2. Supervised learning techniques can be grouped into two main categories based on its use; Regression and Classification.

Related Research Works.

Current and previous research studies on interactions with healthcare personnel, labor interventions, variable importance and other relevant works that address patient expectations and enhance delivery experience with supervised learning technique(s) focusing on patient characteristics is examined. An exploratory study to understand women's expectations with respect to personnel interactions during labor came up with the following themes: provider match, safety/risks, decision making and care satisfaction [10]. For those with expectations that matched provider services, they expressed positive experiences; ironically, on interventions, there were mixed feelings about their use and appropriateness. This was attributed to bureaucratic and complicated processes. This expression of mixed feelings and growing concerns about commonly used childbirth intervention (induction of labor, Augmentation of labor, Artificial rupture of membranes, Episiotomy in vaginal births, Caesarean section etc) is echoed in another study [11] which considered the impact of variations in childbirth interventions in high-income countries for multiparous and nulliparous women. The importance and impact of its use is greatly underscored. But its use routinely in healthy women is estimated to cause maternal and neonatal harm hence the challenge to address ideal rates of use of interventions. Additional interactions that involves family members (spouses, partners, family members or friends) accompanying expectant mothers as companion of choice during labor is also estimated to improve childbirth delivery outcomes [12]. One identifying setback in this interaction is its implementation as many healthcare facilities in most countries lack clear policy direction in this endeavor. Addressing policy decisions as a challenge can lead to improvements in childbirth delivery outcomes. Childbirth delivery interventions are implemented to improve delivery outcomes, but one great challenge to its implementation is the required skill and knowledge by healthcare managers to implement these interventions and how they can impact on important clinical decisions [13]. This study concludes that critical thinking skills and appropriate communication skills were important ingredients necessary for clinical decision making by midwives but also identified variations in levels of knowledge by primary care midwives influencing clinical decisions on childbirth delivery intervention use. Further studies [14] to establish or determine the relationship of factors influencing maternal decision for a choice between caesarean section and vaginal delivery conducted statistical evaluations that identified relationships between three factors: culture, lifestyle,

and perception as the most important variables in the decision for a choice between caesarean section and normal delivery by mothers.

Expectations of every expectant mother is a process that guarantees both safety and reduces risks of pain and other related complications especially in the use of medicalization and labor interventions. In this endeavor, dealing with the psychological effects resulting from the use of interventions such as (episiotomies, forceps or vacuum extraction, C-sections, induction, etc.) could also address maternal expectations and lead to a more positive childbirth experience [15]. In view of the psychological and other negative effects associated with labor interventions, a recent focus on childbirth delivery with minimal interventions has emerged. The desire to limit intervention use is a shared concern by both healthcare providers and expectant mothers [16]. If this desire is to be achieved, then the need to identify patterns of change that necessitate these interventions is of utmost importance. Machine learning approaches with predictive modeling has the potential to identify these patterns of change. In the investigation of machine learning performance and its applicability in identifying pregnancy related complications using electronic medical records, medical images and biological markers, to identify two perinatal complications (pre-eclampsia and prematurity), etc. The prediction of prematurity from medical images in the review of perinatal complications with support vector machines yielded an accuracy score of 95.7% and the prediction of neonatal mortality with XGBoost technique produced an accuracy score of 99.7% [17]. Further studies to predict mode of delivery in view of prevailing high rates of caesarean section exceeding recommended limits using Support vector machines, Multilayer Perceptron, and Random Forest techniques to develop clinical decision support systems for the prediction of mode of delivery specific to three categories: caesarean section, eutocic vaginal delivery and, instrumental vaginal delivery. With an estimated sample population of 25,038 records consisting of 48 attributes, this study [18] use women with singleton pregnancies and concludes that the performance of the three algorithms were similar with above 90% classification accuracy for caesarean section and vaginal delivery and 87% between instrumental and eutocic delivery types. The high prevalence of caesarean section rates as reported in [18] are also echoed in a related article with the objective of determining sub-types of women at higher risk of caesarean section delivery[19]using demographic, clinical and organizational variableswith classification tree analysis. The conclusions drawn indicate that clinical variables were important predictors of caesarean delivery. Further predictive modeling [20] of emergency cesarean section using machine learning models such as logistic regression, random forest, Support Vector Machine (SVM), gradient boosting, extreme gradient boosting (XGBoost), light gradient boosting machine (LGBM), k-nearest neighbors (KNN), Voting, and Stacking concluded that using variables such as maternal age, height, weight at pre-pregnancy, pregnancy-induced hypertension, gestational age, and other ultrasound findings about the fetus showed logistic regression accuracy score of 0.78. It therefore identified Clinical and Sonographicfindings obtained at term as best predictors for emergency caesarean section need. An assessment of the possibility of vaginal delivery after a caesarean section delivery [21] study found limitations in the implementation of calculators into clinical practice. The study therefore centered on assessing the feasibility of machine learning models in addressing these limitations. Study conclusions showed that applying machine learning algorithms that assigns individual risk score for every successful vaginal delivery after caesarean section may assist in future decision making for a delivery outcome.

Preterm deliveries are a worldwide health concern especially to expectant mothers and their immediate families due to associated complications in its management and the attendant deaths resulting from

these complications. Predictive algorithm use with improved accuracy based on the use of important variable features is a challenge that must be overcome. The adoption of entropy feature selection strategy is viewed as a means of overcoming this challenge [22] using three classifiers namely; decision tree (DT), logistic regression (LR), and support vector machine (SVM). SVM generates a prediction accuracy of 90.9% as the highest accuracy rate. An inclusion and exclusion criteria was adopted in the sampling process. Sampled criteria included gestation age of 28 weeks or older, women who delivered live births, those registered with antenatal clinic attendance. Exclusion criteria were; women with multiple gestation (twin gestation), women with still births and women referred to other hospitals.

In its establishment [23] of labor risk scores for maternal and neonatal unfavourable delivery outcomes using machine learning techniques, dataset characteristics of mean gestational age 39.35 ± 1.13 weeks, mean maternal age 26.95 ± 6.48 years and mean parity of 0.92 ± 1.23 are used. This study achieved different accuracy scores at different cervical dilatations. At a cervical dilatation of 4 centimeters (4cm) an accuracy score of 75% was achieved and at cervical dilatation of 10cm, 89% accuracy score was achieved. A systematic review [24] of pregnancy outcomes with machine learning for optimal delivery mode, showed that the use of unsupervised learning techniques together with deep learning algorithms for prediction, resulted in the determination of reasons of maternal complications previously unknown.

In increasing use of artificial intelligence and predictive techniques in various fields and its extension to the healthcare system for disease diagnosis and prediction of diseases, it is important to consider ethical issues relating to data generation, use and acquisition mechanisms [25] as a result of its use in both therapeutic and diagnostic decisions. The increasing use of artificial intelligence and machine learning in healthcare applications is also underscored in a related study [26] that considered reliable prediction model for maternal care decision support systems based on data collected on antenatal signs and symptoms (enriched data) to predict mode of childbirth delivery before term. Conclusions in this study suggests that the use of “enriched data” contributed to the high model performance in sensitivity, specificity, F1-score and receiver operating characteristic curve score (auc). Prediction accuracy scores achieved by the various learning techniques; k-nearest neighbor was 84.38%, bagging was 83.75%, random forest was 83.13%, decision tree was 81.25%, and AdaBoostM1 was 80.63%. Socio-cultural dynamics as a health determinant differ in different geographical regions of the world. A study to determine the effect of socio-demographic effects of caesarean section delivery [27] identified close relationships between a woman’s level of education, income level, habitat and health conditions such as hypertension for both primiparous and multiparous women.

Summary of Related Works

Many of the concepts identified in the related research works can be linked to maternal interaction diagram in Fig 1. Concepts involving labor interventions such as identifying predictors of childbirth delivery, healthcare personnel interactions with patients, knowledge and skills of healthcare personnel in administering labor interventions, medicalization processes (identified medical conditions) that requires care and support, impact of variations for childbirth intervention for both multiparous and nulliparous women have been examined in various studies. Some of the problems identified and addressed in related works above include; prediction of vaginal delivery outcome after a caesarean

section [27], Estimation of labor risk scores for maternal and neonatal delivery outcome [23], A systematic review [24] of pregnancy outcome with machine learning for optimal delivery mode, The adoption of entropy feature selection strategy as a means of overcoming the challenges of use of important variable features[22], determining sub-types of women at higher risk of caesarean section delivery [19], predictive modeling [20] of emergency cesarean section as a delivery outcome, psychological and other negative effects associated with labor interventions[16], identifying pregnancy related complications [17, understanding women's expectations with respect to healthcare personnel interactions during labor [10], impact of variations in childbirth interventions in high-income countries for multiparous and nulliparous women[11]and many others using various features both demographic and medical records of patients. In determining childbirth delivery outcome for an expectant woman either by healthcare personnel or through maternal or patient request, an understanding of pregnancy history relating to number of unsuccessful pregnancies (spontaneous abortions or otherwise), number of stillbirths (fetal deaths if any) in addition to other known factors such as maternal age, blood pressure, haemoglobin count, maternal pulse, fetal heart rate etc may provide useful insight into critical decisions and the underlying reasons for which patient requests are made. It is the non-inclusion of these factors (unsuccessful pregnancies (spontaneous abortions or otherwise), number of stillbirths (fetal deaths if any) that is identified as a research gap that must be addressed. This research therefore, includes these factors in the predictive modeling for delivery outcome to help bring to the fore the impact of these factors in the design of any decision support system for delivery outcome with efficient AI-based techniques.

Research Hypothesis

Null Hypothesis: No relationship exists between delivery outcome and measurable metrics taking into account incidence of abortion(s) and fetal deaths that may have occurred.

Alternate Hypothesis: Taking into account the history of abortion(s) and any fetal deaths that has occurred in the lives of a pregnant woman (spontaneous or otherwise), relationship exists between delivery outcome and real-time measurable metrics obtained from the partograph. To determine the impact of these metrics on delivery outcome, feature selection technique with Chi-square correlation statistic for variable independence is performed for each feature selected to determine its impact on delivery outcome. The determination of best prediction accuracy using balanced accuracy from the evaluation of three (3) machine learning techniques namely; Logistic Regression, Random Forest and Gradient boosting classifiers are used in this context.

Research Materials and Methods

Formal request for permission to use healthcare facility for this exercise was made through a correspondence dated 10th December, 2019 referenced Ds24/2019. Correspondence confirming grant of permission was received on 5th January, 2020 referenced KGHR210/2020.

A qualitative research approach was adopted with focus on maternal history that includes number of (spontaneous abortions or otherwise), number of stillbirths (fetal deaths if any) to understand how history of previous pregnancy outcome could influence the selection or adoption of birth delivery mode. This could lead to a better understanding of contextual parameters with significant contributions to child birth delivery type phenomenon. To achieve this objective, purposive sampling technique for the following category of participants was adopted, namely; those who have not given birth (irrespective of the outcome of previous pregnancies-nulliparous), those in their first pregnancy (primigravida), those who had given birth once (primiparous), those who had been pregnant more than once (multigravida), those who had given birth more than once (multiparous), those who had given birth to five or more infants with gestational age of 24 weeks or more (grand multipara), those who had been pregnant for more than five times (grand multigravida) and finally, those with seven or more deliveries beyond 24 weeks of gestation (great grand multipara). Partograph records of 842 patients were purposively sampled from patients who had delivered at Kwahu Government Hospital from January 2020 to September 2020. The only exclusion criteria used was to eliminate partographs that had not been filled properly or had empty spaces (unfilled sections). Partograph [28] is seen as an essential tool for skill management of the delivery process recording labor progress, maternal and child conditions in real-time for decision making hence its use together with other medical records as the main source of data in this research work. Average gestational period was 38.76 weeks, maximum gestational period was 46 weeks and minimum gestational period was 24 weeks. Minimum age recorded was 14years and maximum age was 45years. Participants 23years were at the 25th percentile, which means 25% of all participants lie below 23years, 50% participants were below and above 28years (50th percentile) and 75% were younger than 32years (75th percentile). Total number of sample population between the ages of 23 years and 32 years were 421. The patient with the highest pregnancies had 17 (gravida 17) with 11 deliveries (para 11) at age 40 years. Age distribution and density spread and other exploratory statistics including study sample population characteristics are shown in Table 1.

Ethical approval and Consent: Ethical approval to conduct this research was obtained from The Clinical coordination team of Kwahu Government Hospital for the use of Electronic healthcare record dataset with approval notice referenced KGHR210/2020 on 5th January, 2020.

Feature Selection and Feature Categories

Subset selection for most relevant features from collected electronic health record of patient information gathered from antenatal hospital attendance is achieved with supervised feature selection technique to ensure that redundant, irrelevant and noisy features are excluded. Supervised Feature selection techniques consider the target variable which in this instance is delivery outcome and can be used for the labeled datasets. The specific feature selection technique adopted in this research work was chi-square correlation statistic which is part of filters method for supervised feature selection.

Chi-square Test: Chi-square test, a filter feature selection method used to determine feature relationships between categorical variables. The chi-square value is calculated between each feature

and the target variable for which the desired number of features with the best chi-square value and alpha value of less than 0.05 is selected. Presentation of feature selection techniques based on four methodologies namely; filter methods, wrapper methods, embedded methods and hybrid methods is illustrated in Fig 3. Among feature categories considered in this research are nulliparous, primigravida, primiparous, multigravida, multiparous, grand multipara, grand multigravida and great grand multipara.

Results

The study sample of 842 participants with varying characteristics of pregnancy considered the impact of including counts of abortion(s) and fetal deaths on delivery outcome decisions. First time pregnancies without abortion(s) were 186 with age range of 14-37, total patients with the history of abortion(s) were 215 and fell within the age range of 17-43 made up of abortion(s) with live births 168, with age range of 19-43 and abortion(s) with no live births 47 age range was 17-38. 251 patients with more than one pregnancy but zero parity (no births) were also recorded and the age range for this category was from 14-38. Fig 4 has three (3) boxplot images that describe skewness of data used, (gravida, parity and abortions) to show data distribution or spread. It can be seen that the median value for gravida is 1 shown by the line in the box and one extreme outlier of value. Both parity and abortion boxplots have only one whisker which means either their minimum values in both instances are equal to that of the lower quartile or the maximum values are equal to the upper quartile. Abortions boxplot has no median line which means that the median value is equal to either the lower or upper quartile. Additional exploratory analysis with scatter to determine relationships in the collected data is also shown in Fig 5. Three features are identified in Fig 5 (gravida, parity and abortions) and shows the density of occurrence in each plot. Recorded abortions under 1 reported were fewer than those between 1 and 3. Number of deliveries (parity) recorded were higher between 0-4 than above 4. Number of pregnancies recorded from 1 to 6 were high than above 6. Fig 6 describes the various processes and demonstrates potential areas of data collection for predictive modeling purposes. It also includes sub-processes which in this instance served as major source of data collection. Among the objectives of this research work was the determination of variable independence on the prediction outcome, Table 2 describe results obtained from performing Chi-square correlation statistic test with collected features to determine feature relationship with the dependent variable. Two dependent variables are identified as those with correlations to the prediction outcome and these are gestational age (in weeks) and progress of maternal cervical dilatation. Fig 7, 8 and 9 are confusion matrices of the three algorithms used namely; Fig 7 (Logistic Regression), Fig 8 (GradientBoosting) and Fig 9 (Random Forest) and each contains descriptions of predicted values for true positive predictions, false positive predictions, true negatives and false negatives. These are also represented as True negatives (TN), True Positives (TP), False Negatives (FN) and False Positives (FP). In Table 3, predicted probability scores for each algorithm is displayed together with balanced accuracy scores in each instance. One of the key objectives was to evaluate model prediction accuracy performance with balanced accuracy scores as real world applications contain imbalanced datasets for which contributions from the minority class is overlook by the majority class. Using balanced accuracy scores instead of roc_auc scores will help address this challenge. In Figure 10, a display of roc_auc score curve shows the following scores obtained by each machine learning algorithm used; Logistic regression has 93%, Random Forest has 91% and Gradientboosting has 91%.

Discussion

This section begins with emphasis on whether research objectives have been achieved. One of the key components set out in the research hypothesis was a determination of impact of used features taking into account the inclusion of abortion(s) and fetal deaths. Results from performing feature selection for variable importance with Chi-square correlation statistic as shown in Table 2 indicate p-values for abortions and fetal deaths as 0.974 and 0.179 respectively which are greater than p-value of alpha (0.05 confidence interval) therefore the test failed to reject H_0 in both instances. However, p-values for two other features (fetal gestational age and progress of cervical dilatation) showed statistical significance with P-value of alpha for fetal gestational age obtained was **0.00** and for cervical dilatations p-value of alpha was **4.43e-63**. These results have been highlighted for emphasis. The two features therefore rejected H_0 indicating correlation in delivery outcome. They prove to be significant predictors of delivery outcome. They were also identified as significant predictors of post partum hemorrhage in a risk prediction modeling research [29]. Fetal gestational age as a predictor of delivery outcome is also shared in a related study for the prediction of labor outcome [30] which among other factors mentioned gestational weeks of 37 and 38 as significant correlated variables to delivery outcome. Gestational age as a significant factor is underscored in other related works such as by [31]. Among factors predicting vaginal delivery as an output after labor induction, gestational age ≤ 39 weeks is listed as a significant predictor by [32]. Progress of Cervical dilatation as a predictor of delivery outcome is also underscored in a related study that used advanced cervical dilation as a predictor for low emergency caesarean section delivery [33]. The use of both features (fetal gestational age and cervical dilatation) as predictors in various research studies therefore gives credence to the predictive capabilities of the factors involved. Evaluation metrics such as area under the receiver operating characteristic curve (auc_roc) and prediction accuracy score shown in Figure 10 (roc_auc score graph) and balanced accuracy score shown in Table 3, indicate high performing traditional machine learning models. An roc_auc score of 91% obtained with random forest is comparable and even higher than results obtained in similar research settings [34] score of 86%. The justification for its use is also underscored in similar research findings [35] conducted to predict the risk of birth asphyxia and in the prediction of intrauterine growth restriction which used deep learning techniques and obtained an roc_auc score of 91% [36].

Model evaluation performance from table 3 show predicted probabilities for True Negative Rates as; Logistic Regression: 73.68%, Gradientboosting: 68.42% and Random Forest: 69.74%. Prediction of an outcome (positive-csection, negative-svd:1, 0) for each algorithms probability accuracy is determined by its percentage. Prediction of svd as a delivery outcome by logistic regression will be 73.68% accurate, 68.42% accurate for Gradientboosting and 69.74% accurate for Random forest therefore Logistic regression has a lower prediction error score than the other two algorithms. This makes Logistic Regression the algorithm of choice. Machine learning modeling of interactions as shown in the maternal interactions flowchart in Fig 1 and the delivery process flowchart in Fig 6 provides a clearer and better understanding of what is required for a successful delivery outcome and an appropriate delivery mode which is based on sound clinical judgment that takes into account the objective of preserving the lives.

Research Contribution:

In this research work, patient's history of previous abortion(s) and fetal deaths have been added to already known variables predominantly used in determining childbirth delivery outcomes in known related research works. Prediction scores obtained with roc_auc for these traditional modeling techniques such as random forest are comparable and in this instance competitively higher than those obtained with advanced techniques as stated in the discussions section. Graphical display of maternal interaction flowchart diagram in this work simplifies childbirth delivery process for enhanced understanding. Real-World applications such as in medical fields have unequal dataset class distribution (imbalanced dataset) problems therefore model evaluation metrics used for performance assessment may take into account minority class contributions. The disparity in output class distributions is discounted by most machine learning techniques giving an erroneous impression of a relatively high prediction accuracy score performance (if prediction accuracy is the focus) in such studies. The use of balanced accuracy score obtained from computed predicted true negative values, true positives values, false negative values and false positive values will lead to the determination of best model performance in instances where minority class determination is a major priority such as healthcare systems.

Strengths and Limitations

Strengths identified in this study are two-fold, one is in the determination of feature correlation. Chi-square correlation statistic showed two feature correlations which are subject of related research studies and therefore confirms the validity of our research results. A second novelty is the use of balanced accuracy in the performance evaluation of our models. We have brought clarity to bear on the use of this evaluation metric and assigned reasons for its use. We have also included in this research work the issue of number of different types of abortions to investigate its impact on delivery outcome decisions. This work has obtained balanced accuracy scores that are significantly high as compare to other related works in this domain. However, this work is limited in certain respects, firstly, is the issue of data size or sample size, most machine learning algorithms work best with large datasets, our work is limited in the size of data collected. This limitation also includes the non-inclusion of patient body mass index (BMI) in the features collected. Observations made and conclusions drawn from electronic health records without personal interactions with patients is another limitation that would have helped clarified certain issues of concern for a contextual understanding.

Mitigating measures: The only mitigating measure considered was the issue of sample size. This is addressed by the use of traditional machine learning algorithms that work best with small sample size for efficient results. It is in this light that we used Logistic Regression together with Random Forest and an ensemble model GradientBoosting classifier for comparative analysis of performance.

Conclusions, Recommendations and Future Work

We have shown in this research how related study results are connected to the maternal interaction model shown in Fig 1. We have also shown the effect of including history of various types of abortions as an input variable and established variable correlations between the input variables and the output. Our predictive features with outcome correlations have shown to be subjects of related research works which confirms our result validity. We have achieved prediction accuracy scores that are comparable to related research works and even much better when compared with the use of prediction accuracies instead of balanced accuracies within this domain. It is our determination to gather large volumes of data for further predictive modeling in this regard.

Sample Population Count (conceptions, deliveries, abortions, fetal deaths and age range)

Description	Count	Age range
1 st time pregnancy with no abortion(s)	186	14-37
Patients with abortions	215	17-43
Abortions with no live birth	47	17-38
Abortions with live births	168	19-43
More than one pregnancy, No births but more than (0) abortions	47	17-38
More than one pregnancy, No births	251	14-38

Table 1. Statistical distribution of sample population counts

Feature Dependency Statistics

Feature	Chi-square value	p-value	Relationship with output
parity	8.24	0.99	failed to reject H_0
gravida	14.19	0.97	failed to reject H_0
pa_age	35.64	0.99	failed to reject H_0
abortions	5.68	0.97	failed to reject H_0
Fet_deaths	11.43	0.18	failed to reject H_0
Mat_bp_systolic	140.82	0.88	failed to reject H_0
Mat_bp_diastolic	141.76	0.31	failed to reject H_0
mat_temp_ini	37.12	0.60	failed to reject H_0
mat_Hb	81.37	0.99	failed to reject H_0
cerx_dil	366.93	4.43e-63	rejected H_0
ges_age	73.00	0.00	rejected H_0
mat_pul_ini	111.60	0.97	failed to reject H_0
fhr_ini	70.20	0.98	failed to reject H_0
anc_vis	20.92	0.98	failed to reject H_0
f_gender	2.62	0.62	failed to reject H_0
f_weight	56.94	0.98	failed to reject H_0

Table 2 Correlation Statistic for variable dependence.

Model Evaluation Performance Results

Model	FNR	TNR	FPR	PPV	NPV	TPR	F1	Balanced Accuracy
Gradient boosting	2.96%	68.42%	31.58%	84.52%	92.86%	97.04%	90.34%	82.73 %
Logistic regression	4.44%	73.68%	26.32%	86.58%	90.32%	95.56%	90.85%	84.62 %
RandomForest	3.7%	69.74%	30.26%	84.97%	91.38%	96.30%	90.28%	83.02 %

Table 3 Performance scores.

Table 3 is a display of performance scores from model evaluations showing false negative rates (FNR), true negative rates (TNR), false positive rates (FPR), predicted positive values (PPV), negative predicted values (NPV), true positive rates (TPR), f1-score and balanced accuracies for the three models used.

$$\text{FNR} = \frac{\text{FN}}{\text{FN} + \text{TP}}$$

where FN = False Negatives and TP = True Positives

$$\text{TNR} = \frac{\text{TN}}{\text{TN} + \text{FP}}$$

Where TN = True Negatives, FP = False Positives

$$\text{FPR} = \frac{\text{FP}}{\text{FP} + \text{TN}}$$

Where FP = False Positives, TN = True Negatives

$$\text{PPV} = \frac{\text{TP}}{\text{TP} + \text{FP}}$$

Where TP = True Positives, FP = False Positives

$$\text{NPV} = \frac{\text{TN}}{\text{TN} + \text{FN}}$$

Where TN = True Negatives, FN = False Negatives

$$\text{TPR} = \frac{\text{TP}}{\text{TP} + \text{FN}}$$

TP + FN Where TP = True Positives, FN = False Negatives

Display of (probability) predicted scores aggregated at threshold points referred to as the auc score by the individual machine learning techniques is shown in figure 10. Prediction accuracy scores of each model is indicated as Random Forest 0.91, Gradient Boosting Classifier 0.91 and Logistic Regression 0.93.

Features used and their descriptions

Feature	Description
pa_age	Age of patient
gravida	Number of pregnancies
parity	Number of deliveries
abortions	abortions
fet_deaths	Fetal deaths
mat_Bp_systolic_ini	Maternal systolic blood pressure
mat_Bp_diastolic_ini	Maternal diastolic blood pressure
mat_temp_ini	Maternal body temperature
mat_Hb	Maternal haemoglobin count
cerx_dil	Cervical dilatation
ges_age	Fetal gestational age
mat_pul_ini	Maternal pulse
fhr_ini	Fetal heart rate
anc_vis	Number of antenatal visits
f_gender	Fetal gender
f_weight	Fetal weight
del_type	Delivery type

Table 4 Variable feature selections used.

Declarations:

Authors Contributions: Conceptualization, Michael Owusu-Adjei,

Methodology: Michael Owusu-Adjei, Twum Frimpong, Gaddafi Abdul-Salaam

Supervision: James Ben Hayfron-Acquah.

Funding: This publication is part of an academic research that is not funded by any organization or individual. It is self-funded by the student without external funding.

Ethics approval and consent to participate; Not applicable

Consent for publication: Approved

Availability of data and Materials: Dataset used for analysis available through upload upon request.

Conflicts of Interest: The author(s) declare that there is no conflict of interest regarding the publication of this paper.

Acknowledgments: Appreciation goes to the management and staff of Kwahu Government Hospital for their immense support and assistance especially the Medical Director Dr. Kobina Awotwe Wiredu, Dr. Nana Osei (clinical coordinator).

References

- [1] B. Zewude, G. Siraw, and Y. Adem, "The Preferences of Modes of Child Delivery and Associated Factors Among Pregnant Women in Southern Ethiopia," *Pragmatic Obs. Res.*, vol. Volume 13, no. July, pp. 59–73, 2022, doi: 10.2147/por.s370513.
- [2] P. J. Teoh and S. M. Nelson, "Fear of childbirth according to parity, gestational age and obstetric history: Correspondence [1]," *BJOG An Int. J. Obstet. Gynaecol.*, vol. 116, no. 7, p. 1005, Jun. 2009, doi: 10.1111/J.1471-0528.2009.02153.X.
- [3] C. Nilsson *et al.*, "Definitions, measurements and prevalence of fear of childbirth: A systematic review," *BMC Pregnancy Childbirth*, vol. 18, no. 1, pp. 1–15, 2018, doi: 10.1186/s12884-018-1659-7.
- [4] P. Adatarara, J. Strumpher, E. Ricks, and P. P. Mwini-Nyaledzigbor, "Cultural beliefs and practices of women influencing home births in rural Northern Ghana," *Int. J. Womens. Health*, vol. 11, pp. 353–361, 2019, doi: 10.2147/IJWH.S190402.
- [5] World Health Organization, *Intrapartum care for a positive childbirth experience*. 2018. [Online]. Available: <http://apps.who.int/iris/bitstream/10665/260178/1/9789241550215-eng.pdf?ua=1%0Ahttp://www.who.int/reproductivehealth/publications/intrapartum-care-guidelines/en/>
- [6] L. Möller, A. Josefsson, C. Lilliecreutz, C. Gunnervik, M. Bladh, and G. Sydsjö, "Reproduction, fear of childbirth and obstetric outcomes in women treated for fear of childbirth in their first pregnancy: A historical cohort," *Acta Obstet. Gynecol. Scand.*, vol. 98, no. 3, pp. 374–381, 2019, doi: 10.1111/aogs.13503.
- [7] M. A. O'Connell, P. Leahy-Warren, L. C. Kenny, S. M. O'Neill, and A. S. Khashan, "The prevalence and risk factors of fear of childbirth among pregnant women: A cross-sectional study in Ireland," *Acta Obstet. Gynecol. Scand.*, vol. 98, no. 8, pp. 1014–1023, 2019, doi: 10.1111/aogs.13599.
- [8] J. Fenwick *et al.*, "Study protocol for reducing childbirth fear: A midwife-led psycho-education intervention," *BMC Pregnancy Childbirth*, vol. 13, pp. 1–9, 2013, doi: 10.1186/1471-2393-13-190.
- [9] A. O. Ebenezer, "An Analysis Of Delivery Methods And Outcome Of Child Birth: Case Of South-Western Nigeria," vol. 5, no. 05, pp. 289–292, 2016.
- [10] A. M. Bibeau, "Interventions during labor and birth in the United States: A qualitative

- analysis of women's experiences," *Sex. Reprod. Healthc.*, vol. 5, no. 4, pp. 167–173, 2014, doi: 10.1016/j.srhc.2014.10.003.
- [11] A. Seijmonsbergen-Schermers *et al.*, "Variations in childbirth interventions in high-income countries: Protocol for a multinational cross-sectional study," *BMJ Open*, vol. 8, no. 1, 2018, doi: 10.1136/bmjopen-2017-017993.
- [12] WHO, "Companion of choice during labour and childbirth for improved quality of care," *Publ. World Heal. Organ.*, no. 4, pp. 1–7, 2020.
- [13] D. C. Zondag, T. M. van Haaren-Ten Haken, P. M. Offerhaus, V. Y. F. Maas, and M. J. Nieuwenhuijze, "Knowledge and skills used for clinical decision-making on childbirth interventions: A qualitative study among midwives in the Netherlands," *Eur. J. Midwifery*, vol. 6, no. September, pp. 1–9, 2022, doi: 10.18332/ejm/151653.
- [14] . Rimin, E. Girsang, and S. L. R. Nasution, "Decision Making Model for Choosing Normal Maternity or Cesarean Section with Machine Learning Approach," no. Himbep 2020, pp. 88–95, 2021, doi: 10.5220/0010289200880095.
- [15] "Euro J Social Psych - 2020 - Hoffmann - Psychological aspects of childbirth Evidence for a birth-related mindset.pdf."
- [16] A. S. Bryant and A. E. Borders, "ACOG COMMITTEE OPINION Number 766:Approaches to Limit Intervention during Labor and Birth," *Obstet. Gynecol.*, vol. 133, no. 2, pp. E164–E173, 2019, doi: 10.1097/AOG.0000000000003074.
- [17] A. Bertini, R. Salas, S. Chabert, L. Sobrevia, and F. Pardo, "Using Machine Learning to Predict Complications in Pregnancy: A Systematic Review," *Front. Bioeng. Biotechnol.*, vol. 9, no. January, pp. 1–16, 2022, doi: 10.3389/fbioe.2021.780389.
- [18] A. De Ramón Fernández, D. Ruiz Fernández, and M. T. Prieto Sánchez, "Prediction of the mode of delivery using artificial intelligence algorithms," *Comput. Methods Programs Biomed.*, vol. 219, 2022, doi: 10.1016/j.cmpb.2022.106740.
- [19] E. Stivanello, P. Rucci, J. Lenzi, and M. P. Fantini, "Determinants of cesarean delivery: A classification tree analysis," *BMC Pregnancy Childbirth*, vol. 14, no. 1, pp. 1–8, 2014, doi: 10.1186/1471-2393-14-215.
- [20] J. H. Wie *et al.*, "Prediction of Emergency Cesarean Section Using Machine Learning Methods: Development and External Validation of a Nationwide Multicenter Dataset in Republic of Korea," *Life*, vol. 12, no. 4, 2022, doi: 10.3390/life12040604.
- [21] M. Lipschuetz *et al.*, "Prediction of vaginal birth after cesarean deliveries using machine learning," *Am. J. Obstet. Gynecol.*, vol. 222, no. 6, pp. 613.e1-613.e12, Jun. 2020, doi: 10.1016/J.AJOG.2019.12.267.
- [22] R. Raja, I. Mukherjee, and B. K. Sarkar, "A Machine Learning-Based Prediction Model for Preterm Birth in Rural India," *J. Healthc. Eng.*, vol. 2021, 2021, doi: 10.1155/2021/6665573.

- [23] S. A. Shazly, B. J. Borah, C. G. Ngufor, V. E. Torbenson, R. N. Theiler, and A. O. Famuyide, "Impact of labor characteristics on maternal and neonatal outcomes of labor: A machine-learning model," *PLoS One*, vol. 17, no. 8 August, pp. 1–13, 2022, doi: 10.1371/journal.pone.0273178.
- [24] 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 Childbirth*, vol. 22, no. 1, pp. 1–19, 2022, doi: 10.1186/s12884-022-04594-2.
- [25] K. H. Ahn and K. S. Lee, "Artificial intelligence in obstetrics," *Obstet. Gynecol. Sci.*, vol. 65, no. 2, pp. 113–124, 2022, doi: 10.5468/ogs.21234.
- [26] Z. Ullah, F. Saleem, M. Jamjoom, and B. Fakieh, "Reliable prediction models based on enriched data for identifying the mode of childbirth by using machine learning methods: Development study," *J. Med. Internet Res.*, vol. 23, no. 6, pp. 1–12, 2021, doi: 10.2196/28856.
- [27] T. Nimi, D. Costa, P. Campos, and H. Barros, "Sociodemographic determinants of caesarean delivery in the largest public maternity hospital in Angola," *Acta Med. Port.*, vol. 32, no. 6, pp. 434–440, 2019, doi: 10.20344/amp.10409.
- [28] G. Space, "The Partograph: An Essential Tool for Decision-Making during Labor," 2002.
- [29] A. Ranjbar, S. Rezaei Ghamsari, B. Boujarzadeh, V. Mehrnoush, and F. Darsareh, "Predicting risk of postpartum hemorrhage using machine learning approach: A systematic review," *Gynecol. Obstet. Clin. Med.*, vol. 3, no. 3, pp. 170–174, 2023, doi: 10.1016/j.gocm.2023.07.002.
- [30] D. G. Iliescu, S. Belciug, R. C. Ivanescu, R. C. Dragusin, M. L. Cara, and D. Laurentiu, "Prediction of labor outcome pilot study: evaluation of primiparous women at term," *Am. J. Obstet. Gynecol. MFM*, vol. 4, no. 6, p. 100711, 2022, doi: 10.1016/j.ajogmf.2022.100711.
- [31] A. Husby, J. Wohlfahrt, and M. Melbye, "Gestational age at birth and cognitive outcomes in adolescence: population based full sibling cohort study," *Bmj*, pp. 1–8, 2023, doi: 10.1136/bmj-2022-072779.
- [32] T. Kamlungkuea, J. Manonai, P. Suriyawongpaisal, and W. Hansahiranwadee, "Factors Predicting Successful Vaginal Delivery Following Induction of Labor in Term Pregnancy," *Int. J. Womens. Health*, vol. 14, no. January, pp. 245–255, 2022, doi: 10.2147/IJWH.S347878.
- [33] J. Breckenkamp *et al.*, "Advanced cervical dilatation as a predictor for low emergency cesarean delivery: A comparison between migrant and non-migrant Primiparae - Secondary analysis in Berlin, Germany," *BMC Pregnancy Childbirth*, vol. 19, no. 1, pp. 1–8, 2019, doi: 10.1186/s12884-018-2145-y.
- [34] V. Mehrnoush, A. Ranjbar, M. V. Farashah, F. Darsareh, M. Shekari, and M. S. Jahromi,

“Prediction of postpartum hemorrhage using traditional statistical analysis and a machine learning approach,” *AJOG Glob. Reports*, vol. 3, no. 2, pp. 1–10, 2023, doi: 10.1016/j.xagr.2023.100185.

- [35] F. Darsareh, A. Ranjbar, M. V. Farashah, V. Mehrnoush, and M. Shekari, “Application of machine learning to identify risk factors of birth asphyxia,” vol. 9, pp. 1–7, 2023.
- [36] E. Taeidi, A. Ranjbar, F. Montazeri, V. Mehrnoush, and F. Darsareh, “Machine Learning-Based Approach to Predict Intrauterine Growth Restriction,” *Cureus*, vol. 15, no. 7, 2023, doi: 10.7759/cureus.41448.

Figure Legend

Fig 1. **Maternal interactions flowchart.** Caption: Maternal care experiences and expectations, interactions with healthcare personnel, labor interventions and medicalization processes on childbirth delivery

Fig 2. **Supervised machine learning types.** Caption: Supervised learning types and the various algorithms they represent.

Fig 3. **Feature selection techniques.** Caption: Presentation of feature selection techniques based on four methodologies namely; filter methods, wrapper methods, embedded methods and hybrid methods.

Fig 4. **Boxplot images:** Caption. Boxplot images that describe skewness of data used, (gravidity, parity and abortions) to show data distribution or spread.

Fig 5 **Scatter plot:** Caption. Scatter plot to determine relationships in the collected data Three features are identified in the collected data are (gravidity, parity and abortions) and shows the density of occurrence in each plot.

Fig 6 **Areas for Feature collection:** Caption. Describe the various processes and demonstrates potential areas of data collection.

Fig 7 Logistic regression confusion matrix: Caption. Contain descriptions of predicted values for true positive predictions, false positive predictions, true negatives and false negatives.

Fig 8 Gradientboosting classifier confusion matrix: Caption. Contain description of predicted values for true positive predictions, false positive predictions, true negatives and false negatives.

Fig 9 Random forest classifier confusion matrix: Caption. Contain description of predicted values for true positive predictions, false positive predictions, true negatives and false negatives.

Fig 10. Receiver operating characteristic curve: Caption. Display of roc_auc score curve that shows scores obtained by each machine learning algorithm used; Logistic regression has 93%, Random Forest has 91% and Gradientboosting has 91%.

