Predicting child delivery mode using data mining by considering maternal and fetal health conditions

Fatsuma Jauro1*, Baththama Bello Alhassan1, Aliyu Garba1, Shehu Muhammad Toro2

1 Department of Computer Science, Ahmadu Bello University, Zaria, Nigeria
2 Faculty of Medicine, Ahmadu Bello University, Zaria, Nigeria

ABSTRACT
In maternity care, deciding which method to use sometimes depends on the interest of the mother. In other cases, the mode of delivery is decided based on the observed health condition of the mother and the fetus. Predicting mode of delivery before term would help reduce the excessive and insignificant usage of operative procedures. In this work, data mining classification models have been used to predict the mode of delivery in obstetrics by considering both maternal and fetal factors. Particularly, K-Nearest Neighbor (KNN), Naïve Bayes (NB), Support Vector Machine, and Decision Trees models were used. The data used is a real dataset obtained from the maternity unit of Ahmadu Bello University Teaching Hospital, Zaria. All the used models were found to be efficient in predicting the mode of delivery as none has less than 90% accuracy. However, NB was found to be the best amongst all with an accuracy of 99.78% and KNN being the least but still with an accuracy of 91.41%.

ARTICLE INFO
Article history:
Received 29 January 2019
Received in revised form
27 April 2019
Accepted 29 April 2019
Published 10 Aug 2019
Available online 10 Aug 2019

KEYWORDS
Maternity
Delivery
Prediction
Models
Accuracy

1. Introduction

Data mining is being used by many organizations, including financial institutions, educational organizations, healthcare organizations, and many others. In healthcare, data mining can be applied in resource usage optimization, cost efficiency, decision making based on evidence, identification of high-risk patients, and taking necessary actions (Pradhan, 2016).

As a part of the healthcare sector, the maternity unit is responsible for providing prenatal/antenatal care to ensure good health for both mother and fetus from pregnancy to delivery. Any signs of prenatal or delivery risks associated with the mother or the child may be detected during antenatal care by the Obstetricians.

Different modes of child delivery exist in maternity care; Normal delivery is referred to as Spontaneous Vaginal Delivery (SVD), Caesarian section (CS), Vacuum extraction, and Forceps delivery. Normal birth takes place naturally without any assistance from doctors. For some women, normal birth becomes difficult or impossible due to some factors associated either with them or with the fetus. Caesarian section or C-section is the delivery of a baby through a surgical procedure in the mother’s abdomen and uterus (Breth, 2015). In some cases, a C-section could be elective and sometimes carried out as an emergency due to unforeseen circumstances. Vacuum extraction and forceps delivery are operative procedures used during normal delivery to assist the baby out of the birth canal using vacuum or forceps, respectively. Deciding which
method of delivery to use sometimes depends on the choice of the mother. In other cases, the method of delivery is determined by observing health conditions or risk factors associated with the mother or the child. Early prediction of the possible and ideal mode of delivery would help in deciding those women in need of surgeries or assistance with operative procedures, hence minimizing the unnecessary usage of operative procedures with the insignificant benefits (Simpson and Thorman, 2005). Assisting the maternity doctors in the decision-making process would reduce malpractice and carelessness, hence improving quality in maternity care and providing better healthcare to pregnant women (Pereira et al., 2015a).

The goal of this paper is to use machine learning classification models to predict the mode of delivery in obstetrics by considering maternal and fetal risk factors. Specifically, K-Nearest Neighbor (KNN), Naive Bayes (NB), Support Vector Machine (SVM), and Decision Trees (DT) are used. The data used is a real dataset obtained from the maternity unit of Ahmadu Bello University Teaching Hospital (ABUTH), Zaria.

1.1. Modes of delivery and their indicators

1.1.1. Caesarian section

Cesarean section is the most common obstetric surgical procedure. It is used to improve fetal and maternal safety, especially when used appropriately. Sometimes, the procedure is performed based on maternal choice with no medical indications. While in some cases health risk factors associated with the mother or fetus indicate the need for the procedure. Such indicators include Previous Cesareans, Obstructed labor, Fetal Distress, Failure to progress, Breech presentation, Hypertensive disorders with pregnancy, Antepartum hemorrhage (placenta previa, placental abruption, and vasa previa), Postdates pregnancy, twins, cord prolapse, transverse lie, maternal health condition (such as diabetes), and fetal health conditions (such as weight, congenital disease) (Shamshad, 2008; Zoe and Ghaem-Maghami, 2001).

1.1.2. Forceps delivery

Forceps are operative tools designed to assist in the delivery of the baby by applying traction to the fetal head (Ross and Beall, 2017). The person in charge must have an understanding of the safe limits of the procedure. Indicators for this procedure include: Prolonged second stage of labor, Suspicion of immediate or potential fetal compromise in the second stage of labor, Shortening of the second stage of labor to benefit the mother (in cases of exhaustion, bleeding, cardiac or pulmonary disease, and history of spontaneous pneumothorax) (Ross and Beall, 2017).

1.1.3. Vacuum extraction

Vacuum and forceps are both delivery instruments but because vacuum extraction has limited risk compared to the forceps, it has progressively replaced forceps as the instrument of choice for many practitioners. Vacuum extraction has the same general indicators as forceps delivery (Garrison, 2017).

In this paper, data mining has been used to predict the mode of delivery in obstetrics by considering some indicators (pregnancy factors) associated with the mother and the fetus.

2. Related Works

A study on obstetrics’ risk factors and delivery techniques was conducted by Pereira et al. (2015a) where real data was used to prove the viability of using data mining model (DMM) to predict which delivery mode should be adopted, through the pregnancy characteristics of the pregnant patients. DM classification algorithms were adopted to predict mode of delivery using four different techniques: DT, generalized linear models (GLMs), SVM, and NB, for their practicality in predicting the most likely outcome in this kind of study. But, only 26 features were used for prediction. We considered more number of features (52) to gain more precise prediction. Again, fetal factors were not taken into consideration despite being a contributing factor to delivery mode.

In Pereira et al. (2015a), the authors studied the maternal risk factors associated with preterm delivery. DMMs were used to predict preterm deliveries taking into consideration the identified risk factors. A sensitive metric to predict preterm deliveries was developed, assisting physicians in the decision-making process regarding the patients at risk and alerting close observation preventing possible complications, hence avoiding preterm birth. Four different algorithms were used: DT, GLMs, SVM, and NB.

In Salisu et al. (2018), the authors compared the strength of two techniques used by decision tree algorithms one that uses information gain and the other that uses Gini index criteria. The later was found to be the best in predicting childbirth delivery type but was not compared with other DMMs.

Other studies have been conducted to determine the mode of delivery in obstetrics by considering different health issues. For example, Hannah et al.
was the first to include obstetricians, ophthalmologists, and their trainees in conducting a survey on the recommended mode of delivery for pregnant females with risk factors of Rhegmatogenous Retinal Detachment (RRD). The outcome of the research recommends that obstetricians concerned about probable RRD in pregnant patients may be unnecessarily recommending operative management.

3. Data Mining Models

Different DMMs are in existence and are used for different purposes. In this study, four DMMs have been used for predicting the mode of delivery in obstetrics. KNN is used for classification and regression. It is one of the simplest classification algorithms and can give a highly competitive result. The closest neighbor (NN) rule distinguishes the classification of an unknown data point on the basis of its closest neighbor whose class is already known.

DT used for classification and regression using a tree structure. A given dataset is broken down into smaller subsets while developing an associated decision tree. DT handles the categorical and numerical dataset.

SVM is known to be used for learning classification, regression, or ranking function. They are developed based on statistical learning theory and structural risk minimization with the aim of determining the location of decision boundaries that produce the ideal separation of classes (Han et al. & Pei, 2011).

NB techniques are a set of supervised learning algorithms based on Baye’s theorem and independent assumptions between predictors. NB is easy to develop without complex iterative parameter estimation, hence making it useful for large dataset. It is widely used because it does well and outperforms more sophisticated techniques (Nikam, 2015).

4. Methodology

4.1. Data understanding

In the course of conducting this research, real dataset was obtained from Obstetrics and Gynecology unit of ABUTH Zaria. It contains about 1,800 data instances consisting 52 variables signifying maternal information, including age, weight, parity, presentation, Twin, mode of delivery with a rationale behind any assisted delivery or CS, and other health conditions. Some of the variables represent information about the baby, such as sex, weight, and malformation.

4.2. Data preparation

In the pre-processing stage, sanitization of the data was performed by removing all the instances with some noise values. Inconsistent and unimaginable values were discovered, for example, zero was captured as a mother’s age and baby’s weight as a result of typographical error in the course of data capture. The most relevant features for the purpose of this research were extracted to be five. Both Vacuum Assisted and Vacuum are considered as Vacuum Assisted. One other feature that was not considered is Laparotomy. After the sanitization, before validating the models with the dataset for processing, the target variables were encoded into numeric values as shown in Table 1.

Table 2 shows the statistical measures of some of the most relevant features extracted in Python using Scikit-Learn library.

Fifty-one different features were considered and encoded as 1 or 0 in the presence or the absence of each, respectively. The 12 features shown in Table 2 were the highly contributing variables in the prediction. The expected predicted values are considered to be 1–6 as encoded in Table 1. Four different data models were employed to predict, compare, and select the best to utilize. In the dataset, 30% is used to train the model using cross-validation concept.

The considered target variables are (five types of deliveries) as follows: SVD, CS, Forceps, Breach Assisted, and Vacuum Assisted with the data distribution as shown in Figure 1. It can be seen that 73.6% of the deliveries are SVD followed by CS with about 24.4% then Breach Assisted with 2%, Vacuum with 0.6%, and forceps 0.2%.

4.3. Modelling

SVM, NB, DT, and KNN machine learning classification models were applied to the processed data. For all the models, a random sampling was used by taking 30% as training data set and 70% as testing data to avoid overfitting. Same variable combinations were used to predict the mode of delivery.

<table>
<thead>
<tr>
<th>Table 1. Encoding type of delivery.</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Type of delivery</strong></td>
</tr>
<tr>
<td>CS (CS)</td>
</tr>
<tr>
<td>SVD</td>
</tr>
<tr>
<td>Vacuum assisted</td>
</tr>
<tr>
<td>Breach assisted</td>
</tr>
<tr>
<td>Forceps</td>
</tr>
</tbody>
</table>
5. Results and Discussion

5.1. Evaluation

To evaluate the models, the following performance metrics estimated by the confusion matrix were computed and summarized as shown in Table 3:

Sensitivity = TP/TP + FN
Specificity = TN/TN + FP
Accuracy = (TP+TN)/(TP+TN+FP+FN)

where TP is number of True Positives, FP is False Positives, TN is True Negatives, and FN False Negatives.

As shown on Table 3, the best model that accomplished the best accuracy is NB with an accuracy of 99.78%, Sensitivity of 0.9, and Specificity of 1.0 followed by DT with an accuracy of 99.56%, Sensitivity of 1.0 which is higher than the NB, and Specificity of 0.99. Next is SVM with an accuracy of 97.14%, Sensitivity of 1.0 which is also higher than that of NB, and Specificity of 0.99. Finally, KNN with an accuracy of 91.41%, Sensitivity of 0.72 which is least amongst all, and Specificity of 1.0. During the analysis of the results, oversampling of data yielded lower performance metrics as compared to cross validation approach. The fact that there is a high challenge in determining a specific mode of delivery but being able to predict the most critical that requires emergency attention of the medical doctor to make a decision of what needs to be done, shows the usefulness of this research. Even with the low occurrences of Vacuum, Breach Assisted, and Forceps, with a percentage of 2%, 0.6%, and 0.2%, respectively, the models were able to predict if any is amongst.

6. Conclusion

The machine learning classification models applied on the real data set all proved to be efficient in predicting the mode of delivery as none has less than
90% accuracy. NB was found to be the best amongst all with an accuracy of 99.78% and KNN being the least but still with reasonable accuracy of 91.41%. It is recommended that the management of ABUTH and other hospitals should adapt and deploy this model in making the most appropriate decision on the mode of delivery. In the future, deep learning alongside NB which has proven to be the best model in this research can be explored.

References