IJTMGH International Journal of Travel Medicine and Global Health

Int J Travel Med Glob Health. 2023 Dec;11(4):391-401
http://ijtmgh.com

doi 10.30491/IJTMGH.2023.402983.1367

Original Article Open Access

Machine Learning-Based Algorithms for Determining C-Section Among Mothers in Bangladesh

Sohani Afroja[1]*, Mohammad Alamgir Kabir[1], Arif Bin Saleh[2]

[1] Department of Statistics, Jahangirnagar University, Savar, Dhaka 1342, Bangladesh.
[2] Physician, Jersey Shore University Hospital, New Jersey, USA.

***Corresponding Author:** Sohani Afroja, Department of Statistics, Jahangirnagar University, Savar, Dhaka 1342, Bangladesh, Email: sohaniafroja@gmail.com.

Received 2023-06-18; **Accepted** 2023-02-19; **Online Published** 2023-12-01

Abstract

Background: C-section prevalence has increased drastically over the past few decades across the globe. This growth has been caused by an array of factors, including maternal, socio-demographic, and institutional factors, and it is a global concern in both developed and developing countries. Therefore, the objective of this study is to identify relevant risk factors for the delivery type, and find a more accurate ML-based model for identifying cesarean women.

Methods: The study is based on 5139 delivery cases from the Bangladesh Demographic Health Survey (BDHS) 2017-18. The number of C-sections performed in the nation has increased to at least 45 percent in the two years prior to 2022. Because of this, we have used multiple logistic regression and machine learning algorithms to determine cesarean delivery and identify the socio-demographic risk factor among mothers in Bangladesh.

Results: An independent $\chi 2$ test was performed before we considered six popular machine learning (ML) algorithms to predict C-section among women in Bangladesh, including logistic regression (LR), random forest (RF), support vector machine (SVM), k-nearest neighbor (KNN), naive Bayes (NB), and decision tree (DT). Model evaluation criteria included accuracy, mean absolute error (MAE), Cohen's kappa, precision, specificity, area under the curve (AUC), and F1 score value. Bivariate analysis results revealed that higher educated mothers and fathers, the richest family, overweight mothers, and hospital delivery had a higher percentage of cesarean babies. With an accuracy of 83.74%, NB (naive Bayes) outperforms the other five classifiers. We can get more precise information than accuracy from the ROC curve and the AUC. Depending on the AUC value, we can see that among all classifiers, Logistic Regression (LR) and Random Forest (RF) provide the most accurate classification for determining c-section among Bangladeshi women.

Conclusion: Our findings contribute to a better understanding of how to categorize C-section intentions among Bangladeshi women. The technique will be useful in identifying the women who are most likely to undergo a C-section in the healthcare system. As a result, the government can launch an effective public awareness campaign.

Keywords: Cesarean delivery, Machine Learning (ML) Algorithm, Performance indicator, Bangladesh.

Citation: Afroja S, Kabir M A, Saleh A B. Machine Learning-Based Algorithms for Determining C-Section Among Mothers in Bangladesh. Int J Travel Med Glob Health, 2023;11(4):391-401. doi: 10.30491/IJTMGH.2023.402983.1367

Introduction

Cesarean section, also known as C-section or cesarean delivery, is the surgical procedure by which one or more babies are delivered through an incision in the mother's abdomen, often performed because vaginal delivery would put the baby or mother at risk such as obstructed labor, twin pregnancy, high blood pressure in the mother, breech birth, problems with the placenta or umbilical cord[1]. Cesarean sections are absolutely critical to save lives in situations where vaginal deliveries would pose risks, so all health systems must ensure timely access for all women when needed. The WHO stated, in 2021, that not all the cesarean sections carried out at the moment are needed for medical reasons. Unnecessary surgical procedures can be harmful, both for a woman and her baby [2]. Bangladesh Health and Demographic Survey (BDHS) revealed that 34 caesarean sections were performed in 2017–2018, the majority of which were unnecessary. According to the World Health Organization, the "ideal rate" for C-sections has been seen as being between 10 and 15 percent since 1985 [3].

Prior to the 1980s, rates of cesarean section were generally less than 10%. These rates, however, have risen such that they have reached over 30% in the last decade in many developed countries [4]. This rise has been even greater in countries with rapidly industrializing economies, such as Brazil and China, where cesarean section rates are now around or over 50% [5]. The high cesarean birth rates have become a matter of concern to international public health agencies [6]. Bangladesh is facing a massive boom in the number of medically unnecessary C-sections - between 2016 and 2018 the number of operations increased by 51 per cent, new figures released by Save the Children reveal. Since the current C-section rate in Bangladesh is 31%, a minimum excess (unnecessary C-sections) of 8% exists if we use the conservative end (15%) of the WHO range. It said 7.7 out of 10 C-sections in Bangladesh are 'unnecessary'. It said Bangladesh saw an estimated 860,000 of these unnecessary operations last year.

Mothers who choose to undergo C-sections do so for a variety of reasons [7–10]. Furthermore, C-sections have been associated with obesity [7], maternal age, education, wealth index, mothers' employment status, and place of delivery [9–10], but studies in Bangladesh have not used machine learning algorithms to identify women who will undergo C-sections. This analytical study's aim is to pinpoint the maternal socio-demographic variables that influence delivery methods in Bangladesh with the help of the Bangladesh Demographic Health Survey (BDHS), 2017–18. Using the multiple logistic regression (MLR)-based model with a p-value (p-value < 0.05) and odds ratio (OR), potential risk factors for women having C-section deliveries are identified. Six different machine learning (ML)-based classifiers used to predict whether women would give birth through C-section, including logistic regression (LR), random forest (RF), support vector machine (SVM), k-nearest neighbor (KNN), naive Bayes (NB), and decision tree (DT). The accuracy, mean absolute error (MAE), Cohen's kappa, precision, specificity, area under the curve (AUC), and F1 score value were the criteria used to assess the model's performance. The overall layout of this paper is as follows: materials and methods are described in section 2 along with a dataset description, statistical analysis, feature selection techniques, and machine learning algorithms, results are shown in section 3, the detailed arguments are shown in part 4 as discussion, and the conclusion is then offered in section 5.

2.Materials and methods

The data was obtained from the BDHS in 2017-2018. The potential risk factors were then extracted using a multiple logistic regression model. Third, we used six ML-based classifiers (LR, RF, SVM, KNN, NB, and DT) to predict C-section among mothers and tuned their hyperparameters. Finally, the performance of these classifiers was evaluated using eight evaluation parameters, namely, accuracy, MAE, MSE, RMSE, precision, sensitivity, specificity, and F1 Score, to determine the best predictor.

2.1. Data sources

Data from the most recent Bangladesh Demographic and Health Surveys (BDHS) from 2017-2018 are used in this study. This study does not require ethical approval because it is based on secondary data. In the BDHS 2017-18, a two-stage stratified sampling method was used [11]. The survey included 20127 completed interviews of ever-married women aged 15-49 years, with 7374 (36.64%) from cities and 12753 (63.36%) from rural areas. The adjusting dataset for this analysis included 5139 respondents chosen for the final research, with 1745 (34%) urban and 3394 (66%) rural respondents.

2.2. The dependent and predictor variables

The response variable in this study was "delivery by cesarean section," which was classified as "No or Yes" because the main purpose of this study was to predict mother's delivery by C-section among women aged 15-49 years. If the respondent was delivered via cesarean section, she is in the "Yes" group; otherwise, she is in the "No" group.

The selected predictor variables are the division (Barisal, Chittagong, Dhaka, Khulna, Mymensingh, Rajshahi, Rangpur, Sylhet), type of residence (urban, rural), mother's age (15-19 years, 20-24 years, 25-29 years, 30 years and above), mother's education level (no education, primary, secondary, higher), father's education level (no education, primary, secondary, higher), wealth index (poorest, poorer, middle, richer, richest), mother's working status(no, yes), media access (no, yes), mother's body mass index (underweight (< 18.5), normal (18.5–24.99), overweight (≥ 25), place of delivery (home, hospital), and ever had a terminated pregnancy (no, yes). These are potential factors that determine the c-section among mothers in Bangladesh [12-14].

2.3. Statistical analyses

Statistical analyses were performed using SPSS version 23 (SPSS Inc Chicago, USA), R version 3.5.2 (Bell Laboratories, New Jersey, USA), and Python 3.4.0. The frequency distribution was used to describe the background characteristics of the respondents. For bivariate analysis, cross-tabulations were executed to attain the women's delivery by cesarean section for selected co-variables and to identify statistically significant determinants using the Pearson chi-square ($\chi 2$) test. The test was two-tailed, and a significant factor with a p-value<0.05 was considered. We've implemented logistic regression (LR) to identify the critical covariates for determining the C-section among mothers in Bangladesh using p-value (p-value <0.05) and odds ratio (OR). We have applied six ML-based classifiers such as-logistic regression (LR), random forest (RF), support vector machine (SVM), k-nearest neighbors (KNN), naive Bayes (NB), and decision tree (DT) for determining the C-section among mothers based on the covariates.

2.3.1. Logistic Regression (LR)

Let Y is a binary variable i.e. mode of delivery, which has a Bernoulli distribution of the parameter π; then, the logistic regression model is log[π/(1-π)=β_0+β_1 X_1+...〖+β〗_p X_p β], where β_0,β_1,...〖,β〗_p are the unknown coefficients or parameters [15].

2.3.2. Random Forest (RF)

RF is a classification and regression technique built on the combination of many decision trees. It is specifically a group of trees built from training data and internally validated to produce a forecast of the response given the predictors for upcoming observations [16].

2.3.3. Support Vector Machine (SVM)

The SVM proposed by Vapnik has been studied extensively for classification, regression and density estimation. SVM seeks to distinguish between classes by maximizing geometric margin and minimizing classification error [17].

2.3.4. K-Nearest Neighbors (KNNs)

k-NN is a robust and adaptive classification algorithm that is part of the supervised ML family. The decision boundary of the algorithm depends on a few input points and their particular positions. Thus, the classification of new cases is based on a similarity or the use of observations in the training set that are closest in metric space [18]

2.3.5 Naive Bayes (NB)

The naive Bayes algorithm is a classification algorithm that is based on Bayes' theorem, which is a way of calculating the probability of an event based on its prior knowledge. The algorithm is called "naive" because it makes a simplifying assumption that the features are conditionally independent of each other given the class label [19].

2.3.6 Decision Tree (DT)

Decision Tree classifier is the regression model which is represented in the form of tree structure. The purpose of the Decision Tree classifier is to break down the dataset into smaller subset. The tree consists of decision nodes and leaf nodes. The node which is present as the top most of the decision node acts as a predictor which is called the root node. The node which cannot be further divided is known as leaf node [20].

2.4 Model Evaluation

Following seven evaluation parameters were taken into consideration.

2.4.1 Accuracy

Accuracy calculates the proportion of classifications a model successfully predicts to all predictions produced [21]. It may be calculated as:

$$Accuracy = \frac{(True\ positive + True\ negative)}{(True\ positive + True\ negative + False\ positive + False\ negative)}$$

2.4.2 Mean Absolute Error (MAE)

MAE measures the average magnitude of the errors in a set of predictions, without considering their direction. It's the average over the test sample of the absolute differences between prediction and actual observation where all individual differences have equal weight.

2.4.3 Cohen's Kappa

Dealing with multiclass and unbalanced classification issues requires the use of Cohen's Kappa statistics. It involves analyzing the actual taxonomy and comparing the expected and actual classifications in a data collection. Cohen's Kappa range is 1. Cohen's Kappa values of 0, according to Landis and Koch, denote no agreement, 0 to 0.20, slight, 0.21 to 0.40, fair, 0.41 to 0.60, moderate, 0.61 to 0.80, and significant, as well as 0.81 to 1 practically perfect agreement [22].

2.4.4 Precision

It indicates what proportion of all the points that the model expected to be positive are in fact positive [23].

$$Precision\ = \frac{True\ Positive}{True\ Positive\ +\ False\ Positive}$$

2.4.5 Specificity

In order to accurately identify negative ratios, specificity is measured. This can also be shown as a false positive rate. Mathematically, specificity can be calculated as follows [24]:

$$Specificity = \frac{True\ Negative}{True\ Negative\ +\ False\ Positive}$$

2.4.6 Area under the ROC Curve (AUC)

It indicates how well the model can distinguish between classes. AUC close to 1, which indicates a high level of separability, is a sign of an excellent model. An inadequate model has an AUC close to 0, which indicates that it has the weakest separability [24].

2.4.7 F1-score

It's the combination of both metrics precision and recall where recall measures how the model identifies events in the positive class. Mathematically, F-1 score is given as follows [24]:

$$F1 - score = 2 * \frac{Precision * Recall}{Precision + Recall},$$

$$where\ Recall = \frac{True\ Positive}{True\ Positive + False\ Negative}.$$

3.Results

3.1 Descriptive statistics for the participants' socio-demographic characteristics.

The socio-demographic characteristics of the participants are shown in Table 1, with cesarean sections accounting for 32.7% of total births. Chittagong Division has the highest percentage of respondents (16.5%). Rural areas are home to more than two-thirds (66.0%) of the respondent population. Almost all mothers are literate, with only a few exceptions (6.4%). The majority of the participants (35.6%) are between the ages of 20 and 24. The majority of participants (approximately 62%) had normal BMI. Only 19.7% of those come from the wealthiest families, 19.9% from the richer class, and the rest from the middle, poorer, and poorest classes. Fathers have a lower education level than mothers, with most mothers finishing secondary school (47.8%), but fathers only finishing primary school (34.1%). Sixty-three percent of mothers are unemployed. The majority of people prefer hospitals (50.2%) to give birth at home. More than half of those responded to the survey (63.6%) have access to the media. Almost all respondents (83%) had not terminated their pregnancy.

Table 1: Socio-demographic characteristics of the participants.

Demographic characteristics	Frequency (5139)	Percent (%)
Delivery by cesarean section		
No	3457	67.3
Yes	1682	32.7
Division		
Barisal	546	10.6
Chittagong	846	16.5
Dhaka	741	14.4
Khulna	525	10.2
Mymensingh	622	12.1
Rajshahi	543	10.6
Rangpur	573	11.2
Sylhet	743	14.5
Type of residence		
Urban	1745	34.0
Rural	3394	66.0
Mother's Age		
15-19 years	891	17.3
20-24 years	1827	35.6
25-29 years	1338	26.0
30-49 years	1083	21.1
Mother's educational level		
No education	331	6.4
Primary	1437	28.0
Secondary	2458	47.8
Higher	913	17.8
Father's education level		
No education	714	13.9
Primary	1751	34.1
Secondary	1696	33.0
Higher	978	19.0
Wealth index		
Poorest	1122	21.8
Poorer	1054	20.5
Middle	929	18.1
Richer	1023	19.9
Richest	1011	19.7
Mother's working status		
No	3236	63.0
Yes	1903	37.0
Media Access		
No	1871	36.4
Yes	3268	63.6
Mother's Body Mass Index		
Underweight	818	15.9
Normal	3153	61.4
Overweight	1168	22.7
Place of delivery		
Home	2558	49.8
Hospital	2581	50.2
Ever had a terminated pregnancy		
No	4264	83.0
Yes	875	17.0

Table 2 displays the percentage distribution and correlation between some covariates of interest. Each covariate had a statistically significant relationship with the likelihood ($\chi2$ test) of a cesarean section being chosen as the mode of delivery. The proportion of mothers with an intention to deliver by cesarean section is found to be higher for the Khulna division (44.2%), mothers living in an urban area (42.6%), unemployed mothers (37.2%), mothers in the age group 30 and above (34.4%), mothers with the richest wealth status (63.2%), mothers with higher education (61.1%), fathers with higher education (60.8%), overweight mothers (51.2%), hospitals as a place of delivery (65.2%), mothers with media access (41.6%), and mothers who had a terminated pregnancy (37.8%).

Table 2: Percentage distribution and association between selected covariates and women's delivery by cesarean section in Bangladesh

Selected co-variates	Delivery by cesarean section		χ^2- value	P - value
	No	Yes		
Division				
Barisal	73.1	26.9	126.627	**0.000***
Chittagong	72.6	27.4		
Dhaka	56.1	43.9		
Khulna	55.8	44.2		
Mymensingh	72.2	27.8		
Rajshahi	61.9	38.1		
Rangpur	68.6	31.4		
Sylhet	75.0	25.0		
Type of residence				
Urban	57.4	42.6	116.400	**0.000***
Rural	72.3	27.7		
Mother's Age				
15-19 years	71.0	29.0	8.156	**0.043***
20-24 years	66.2	33.8		
25-29 years	67.6	32.4		
30 - 49 years	65.6	34.4		
Mother's educational level				
No education	84.9%	15.1%	532.078	**0.000***
Primary	82.5%	17.5%		
Secondary	66.6%	33.4%		
Higher	38.9%	61.1%		
Father's education level				
No education	83.9	16.1	537.380	**0.000***
Primary	78.1	21.9		
Secondary	65.3	34.7		
Higher	39.2	60.8		
Wealth index				
Poorest	87.3%	12.7%	695.862	**0.000***
Poorer	78.0%	22.0%		
Middle	69.1%	30.9%		
Richer	62.8%	37.2%		
Richest	36.8%	63.2%		

Selected co-variates	Delivery by cesarean section		χ^2- value	P - value
	No	Yes		
Mother's working status				
No	62.8	37.2	79.529	**0.000***
Yes	74.9	25.1		
Media Access				
No	82.8	17.2	321.876	**0.000***
Yes	58.4	41.6		
Mother's Body Mass Index				
Underweight	78.1	21.9	247.884	**0.000***
Normal	71.3	28.7		
Overweight	48.8	51.2		
Place of delivery				
Home	100.0	0.0	2478.094	**0.000***
Hospital	34.8	65.2		
Ever had a terminated pregnancy				
No	68.3	31.7	13.015	**0.000***
Yes	62.1	37.9		

* p-value < 0.05.

Table 3 shows the results of a multiple logistic regression analysis performed between cesarean section delivery and other covariates. Mothers aged 30 and up are 1.364 times (OR = 1.364, 95% CI: 1.011-1.839, p = 0.042) more likely to have a cesarean section than mothers aged 15 to 19. Higher educated fathers are 1.589 times (OR = 1.589, 95% CI: 1.075-2.348, p = 0.02) more likely to attend cesarean section delivery than uneducated fathers. Women who have access to media are nearly 1.35 times more likely (OR = 1.35, 95% CI: 1.087-1.675, p = 0.007) than those who do not have access to media. Overweight mothers are nearly 1.712 times (OR = 1.712, 95% CI: 1.267-2.313, p = 0.007) more likely to have a cesarean section than underweight mothers. Respondents who have ever terminated a pregnancy are nearly 1.273 times (OR = 1.273, 95% CI: 1.01-1.606, p = 0.041) more likely to undergo cesarean delivery than those who have never terminated a pregnancy.

Table 3: Factors of women's delivery by cesarean section in Bangladesh using multiple logistic regression (MLR).

Factors	Odds Ratio \`(OR)	95% Confidence interval (CI)		p-value
		Lower	Upper	
Division				
Barisal (Ref)	-	-	-	-
Chittagong	0.634	0.442	0.908	**0.013***
Dhaka	1.248	0.863	1.805	0.24
Khulna	1.368	0.936	1.999	0.105
Mymensingh	1.149	0.781	1.691	0.481
Rajshahi	1.319	0.902	1.931	0.153
Rangpur	0.877	0.602	1.278	0.496
Sylhet	0.777	0.533	1.134	0.191
Type of residence				
Urban (Ref)	-	-	-	-
Rural	1.392	1.141	1.698	**0.001***

Factors	Odds Ratio `(OR)	95% Confidence interval (CI)		p-value
		Lower	Upper	
Mother’s Age				
15-19 years (Ref)	-	-	-	-
20-24 years	1.198	0.937	1.53	0.149
25-29 years	1.106	0.845	1.446	0.464
30 - 49 years	1.364	1.011	1.839	**0.042***
Mother’s educational level				
No education (Ref)	-	-	-	-
Primary	0.748	0.455	1.231	0.253
Secondary	0.885	0.544	1.439	0.623
Higher	1.136	0.668	1.933	0.638
Father's education level				
No education (Ref)	-	-	-	-
Primary	1	0.72	1.389	0.999
Secondary	1.107	0.787	1.556	0.56
Higher	1.589	1.075	2.348	**0.02***
Wealth index				
Poorest (Ref)	-	-	-	-
Poorer	1.306	0.95	1.796	0.10
Middle	1.281	0.922	1.781	0.14
Richer	1.204	0.86	1.687	0.28
Richest	2.255	1.529	3.326	**0.00***
Mother’s working status				
No (Ref)	-	-	-	-
Yes	0.762	0.628	0.925	**0.006***
Media Access				
No (Ref)	-	-	-	-
Yes	1.35	1.087	1.675	**0.007***
Mother’s Body Mass Index				
Underweight (Ref)	-	-	-	-
Normal	1.131	0.876	1.462	0.345
Overweight	1.712	1.267	2.313	**0.00***
Place of delivery				
Home (Ref)	-	-	-	-
Hospital	21.79	0.00	22.25	0.978
Ever had a terminated pregnancy				
No (Ref)	-	-	-	-
Yes	1.273	1.01	1.606	**0.041***

$Ref = Reference\ category,$ * p-value < 0.05.

The logistic regression classifier has an accuracy of 82.88%, according to Table 4. The fitted model's precision and specificity were 67.72% and 80.23%, respectively, with an F1 score of 76.74%. The AUC was calculated to be 0.91. A random forest prediction performance result with an accuracy of 81.71% was displayed. The random forest classifier's precision, specificity, and F1 score in this case were 71.71%, 86.97%, and 71.09%, respectively. In this instance, the AUC was 0.91. The support vector machine classifier's overall accuracy was 82.72%, with a precision of 64.87% and a specificity of 74.63%. In this case, the F1 score and the AUC value were 78.69% and 0.90, respectively. Using the k-nearest neighbor algorithm, the accuracy was 80.16%, with precision and specificity of 69.04% and 85.60%, respectively, and an F1 score of 68.79%. The AUC in this case was 0.85. These values are 83.74% (accuracy), 68.93% (precession), 81.14% (specificity), 77.79% (F1 score), and 0.90 (AUC) for naive Bayes classifiers. Finally, we used a decision tree and obtained an accuracy of 82.72%. Other parameters such as precession, specificity, F1 score, and AUC are 64.87%, 74.63%, 78.69%, and 0.89, respectively. NB (naive Bayes) performed the best among the six classifiers in terms of accuracy (83.74%) and MAE (16.27). Cohen's kappa value is 0.6529.

Table 4: Performance indicators of all six machine learning algorithms to determine delivery by cesarean section among mothers in Bangladesh

Model	Accuracy	MAE	Cohen's kappa	Precision	Specificity	AUC	F1
LR	82.88	17.12	0.6357	67.72	80.23	0.91	76.74
RF	81.71	18.29	0.5772	71.71	86.97	0.91	71.09
SVM	82.72	17.28	0.6524	64.87	74.63	0.90	78.69
KNN	80.16	19.84	0.5424	69.04	85.60	0.85	68.79
NB	83.74	16.26	0.6529	68.93	81.14	0.90	77.79
DT	82.72	17.28	0.6524	64.87	74.63	0.89	78.69

Among the six classifiers, we obtain the best performance of NB (naive Bayes) with an accuracy of 83.74%. Although accuracy is a parameter for evaluating performance, we estimate model performance based on the ROC (receptor performance) curve and the AUC (area under the ROC curve) value. Because the overall accuracy is based on a cut point, the ROC curve tries all the cut points and plots the false positive rate and true positive rate. If we try to interpret the model performance depending on accuracy, we only consider a particular cut point. But overall accuracy varies with different cut points, which are taken into account when drawing the ROC curve. Furthermore, AUC is the measure of separability that indicates the model's capability in distinguishing between classes. Thus, in practice, the ROC curve and the AUC can give us more accurate information than accuracy. Depending on the AUC value (Figure 1), we can see that Logistic Regression (LR), Random Forest (RF) produces a distinction between C-section and normal delivery among all classifiers. Figure2 shows that the location of delivery is an important factor in determining cesarean delivery.

Figure 1: Performance evaluation of various classification techniques using ROC curve.

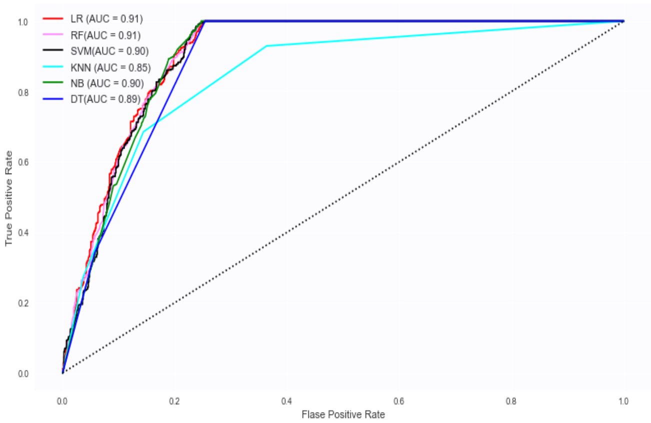
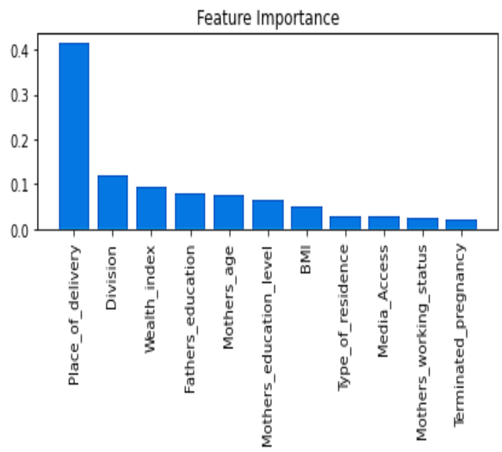


Figure 2: Visualization plot for feature importance for determining C-section.



4. Discussion

The study proposed a machine learning-based approach for more accurate risk assessment of women undergoing cesarean sections using BDHS, 2017–18 data. To determine the C-section risk factors, multiple logistic regression was used. Division, region, mother's age, father's education, wealth index, mother's employment status, media access, mother's body mass index, and aborted pregnancy were shown by MLR results to be the most significant risk factors for C-section. This finding was consistent with previous research. In our study, we observed that the prevalence of C-section was more among the 30-49 age group. Different studies also found that the probability of C-section was higher with increased age [13].

Mothers and fathers who are educated and wealthy have shown a significant rise of C-section which is quite alarming for a low- and middle-income country like Bangladesh where resources are scarce and problems are multiple. In Bangladesh some important social determinants of health, particularly female education has improved considerably during the last couple of decades, [25]. Results of this study indicate that educational gain was greater for mothers than their husbands. Studies in other settings have also explored the influence of maternal education on the use of maternity care services especially CS delivery [26]. In Bangladesh it can be said that with increasing average income and higher coverage of private facilities, the CS rate will continue to rise unless we take some strategic moves and impose some controlling provisions [13].

Preference of the performance of C-section depends on different medical emergencies that include high maternal age, obesity of mother, place of delivery [27]. These are the factors that are considered as risk factors and induce a preference for C-section. In this study mothers aged 30 or more have greater chances of CS delivery than younger mothers. Awareness developed due to increased exposure to different media like TV. Therefore, chances of CS are high for their delivery.

This is the first study that uses machine learning classifiers to identify C-sections among mothers in Bangladesh. The primary goal of this study is to identify cesarean deliveries among Bangladeshi mothers. To achieve the research objectives, six well-known machine learning algorithms, including logistic regression, random forest, support vector machine, k-nearest neighbor, naive Bayes, and decision tree classifier, are used. We trained all models using the training data set and tested their performance using the test data set. Using the $\chi 2$ test, all covariates are significantly related to the outcome variables.

Considering that many authors have made comparisons based on accuracy, the prediction performance of these six machine learning algorithms is compared with accuracy [28]. According to the accuracy, the naive Bayes algorithm generated the best results. Despite this, a number of researchers have demonstrated that AUC is a better method than accuracy in terms of both experience and form [29]. The logistic regression and random forest algorithms produced the best results based on the ROC curve area.

5. Conclusion

The excessive use of cesarean delivery in Bangladesh is a critical public health issue that risk factors for C-section women were identified. We used six ML-based classifiers to predict C-sections in women. Our experimental results revealed that the NB (naive Bayes) classifier provided the highest accuracy while the logistic regression and random forest classifiers provided the highest AUC. In this study, we discovered that the location of delivery is the most important factor in cases of C-section. This research will assist health care providers and policymakers in developing a system for implementing necessary interventions as well as care practices in order to avoid severe complications and the burden on the healthcare system.

Highlights

What do we already know about this topic?

Cesarean deliveries are associated with a number of health risks for both mothers and children. Several socioeconomic and demographic reasons are to blame for Bangladesh's rising C-section births. Multiple statistical approaches (Binary Logistic Regression analysis) have been used to determine the significant factors of c-section in married women.

What does this research contribute?

Machine learning techniques were not investigated to develop prediction models for c-section in married women. As a result, we included a machine learning approach for better prediction.

Acknowledgments

The authors are grateful to NIPORT, Dhaka, Bangladesh, the Government of Bangladesh, and MEASURE DHS, Calverton, MD, USA for granting permission to use the BDHS and BMMS datasets.

Conflict of Interest

There were no potential conflicts of interest disclosed by the authors in relation to the research, authorship, and/or publication of this article.

Funding Source

This research received no specific funding from public, commercial, or non-profit organizations.

Authors' Contributions

AS designed the methodology, and drafted the manuscript. While SAB offered his experience in c-sections and assisted with manuscript preparation, KMA and AS handled the data preparation and statistical analysis. The manuscript was edited by KMA. The finished manuscript was read and approved by all writers. Each author further attests that none of the content in this or any similar work has been submitted to or will be published in any other publication.

Consent For Publication

The authors declare that they have consent for publication.

Ethics approval

The study maintained the confidentiality of the participants' personal information. All information that would identify the participants was excluded from the analysis and reporting of the data.

Reference:

1. Molina, G., Weiser, T. G., Lipsitz, S. R., Esquivel, M. M., Uribe-Leitz, T., Azad, T., ... & Haynes, A. B. (2015). Relationship between cesarean delivery rate and maternal and neonatal mortality. Jama, 314(21), 2263-2270.doi:10.1001/jama.2015.15553.

2. World Health Organization. WHO Statement on Caesarean Section Rates, 2021.

3. Walker, J. (2016). Caesarean section: a global perspective. A textbook of caesarean section. Oxford University Press, Oxford, 191-200. doi: 10.1093/med/9780198758563.003.0014.

4. World Health Organization, Human Reproduction Programme. WHO statement on caesarean section rates. Geneva: WHO; 2015.

5. Pasko, D., Subramaniam, A., & Tita, A. T. (2016). The epidemiology of caesarean delivery. Textbook of Caesarean Section, 37. doi: 10.1093/med/9780198758563.003.0004.

6. Lumbiganon P, Laopaiboon M, Gülmezoglu AM, Souza JP, Taneepanichskul S, Ruyan P, et al. Method of delivery and pregnancy outcomes in Asia: The WHO global survey on maternal and perinatal health 2007–08. Lancet 2010; 375(9713): 490–9. doi: 10.1016/S0140-6736(09)61870-5.

7. Darmasseelane, K., Hyde, M. J., Santhakumaran, S., Gale, C., & Modi, N. (2014). Mode of delivery and offspring body mass index, overweight and obesity in adult life: a systematic review and meta-analysis. PloS one, 9(2), e87896. doi: 10.1371/journal.pone.00878960.

8. Begum, T., Rahman, A., Nababan, H., Hoque, D. M. E., Khan, A. F., Ali, T., & Anwar, I. (2017). Indications and determinants of caesarean section delivery: Evidence from a population-based study in Matlab, Bangladesh. *PloS one*, *12*(11), e0188074. doi: 10.1371/journal.pone.0188074.

9. Gondwe, T., Betha, K., Kusneniwar, G. N., Bunker, C. H., Tang, G., Simhan, H., ... & Haggerty, C. L. (2019). Maternal factors associated with mode of delivery in a population with a high cesarean section rate. Journal of Epidemiology and Global Health, 9(4), 252. doi: 10.2991/jegh.k.191017.001.

10. Amiri, M., Rostami, M., Sheidaei, A., Fallahzadeh, A., & Ramezani Tehrani, F. (2023). Mode of delivery and maternal vitamin D deficiency: an optimized intelligent Bayesian network algorithm analysis of a stratified randomized controlled field trial. Scientific Reports, 13(1), 8682. doi:10.1038/s41598-023-35838-6.

11. The Dhs Program - Bangladesh, *The DHS Program – Bangladesh: Standard DHS, 2017-18 dhsprogram.com*, 2017, https://dhsprogram.com/methodology/survey/survey-display-536.cfm.

12. Rahman, M., Shariff, A. A., Shafie, A., Saaid, R., & Tahir, R. M. (2015). Caesarean delivery and its correlates in Northern Region of Bangladesh: application of logistic regression and cox proportional hazard model. *Journal of Health, Population and Nutrition*, *33*, 1-11. doi: 10.1186/s41043-015-0020-2.

13. Rahman, M. M., Haider, M. R., Moinuddin, M., Rahman, A. E., Ahmed, S., & Khan, M. M. (2018). Determinants of caesarean section in Bangladesh: Cross-sectional analysis of Bangladesh Demographic and Health Survey 2014 Data. *PloS one*, *13*(9), e0202879. doi: 10.1371/journal.pone.0202879.

14. Haider, M. R., Rahman, M. M., Moinuddin, M., Rahman, A. E., Ahmed, S., & Khan, M. M. (2018). Ever-increasing Caesarean section and its economic burden in Bangladesh. *PloS one*, *13*(12), e0208623. doi: 10.1371/journal.pone.0208623.

15. Bitew, F. H., Sparks, C. S., & Nyarko, S. H. (2022). Machine learning algorithms for predicting undernutrition among under-five children in Ethiopia. Public Health Nutrition, 25(2), 269-280. doi: 10.1017/S1368980021004262.

16. Boulesteix, A. L., Janitza, S., Kruppa, J., & König, I. R. (2012). Overview of random forest methodology and practical guidance with emphasis on computational biology and bioinformatics. Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery, 2(6), 493-507. doi: 10.1002/widm.1072.

17. Vapnik, V. (1999). The nature of statistical learning theory. Springer science & business media.

18. Hastie T, Tibshirani R & Friedman J (2009) The Elementsof Statistical Learning: Data Mining, Inference, andPrediction, 2nd ed. New York: Springer. doi: 10.1007/978-0-387-21606-5.

19. Onan, A., Korukoğlu, S., & Bulut, H. (2016). A multiobjective weighted voting ensemble classifier based on differential evolution algorithm for text sentiment classification. Expert Systems with Applications, 62, 1-16. doi: 10.1016/j.eswa.2016.06.005.

20. Sarker, F., Tiropanis, T., & Davis, H. C. (2014, November). Linked data, data mining and external open data for better prediction of at-risk students. In 2014 International Conference on Control, Decision and Information Technologies (CoDIT) (pp. 652-657). IEEE. doi:10.1109/CoDIT.2014.6996973

21. Olson, D. L., & Delen, D. (2008). Advanced data mining techniques. Springer Science & Business Media.

22. Yu, W., Liu, T., Valdez, R., Gwinn, M., & Khoury, M. J. (2010). Application of support vector machine modeling for prediction of common diseases: the case of diabetes and pre-diabetes. BMC medical informatics and decision making, 10(1), 1-7. doi: 10.1186/1472-6947-10-16

23. Narkhede, S. (2018). Understanding auc-roc curve. Towards Data Science, 26(1), 220-227.

24. Jain, D. (2020). Performance measures of models. Towards Data Science.

25. Häger, R. M., Daltveit, A. K., Hofoss, D., Nilsen, S. T., Kolaas, T., Øian, P., & Henriksen, T. (2004). Complications of cesarean deliveries: rates and risk factors. *American journal of obstetrics and gynecology*, *190*(2), 428-434. doi: 10.1016/j.ajog.2003.08.037.

26. Gupta, S., Saluja, K., Goyal, A., Vajpayee, A., & Tiwari, V. (2022). Comparing the performance of machine learning algorithms using estimated accuracy. *Measurement: Sensors*, *24*, 100432. doi: 10.1016/j.measen.2022.100432.

27. Ling, C. X., Huang, J., & Zhang, H. (2003). AUC: a better measure than accuracy in comparing learning algorithms. In *Advances in Artificial Intelligence: 16th Conference of the Canadian Society for Computational Studies of Intelligence, AI 2003, Halifax, Canada, June 11–13, 2003, Proceedings 16* (pp. 329-341). Springer Berlin Heidelberg. doi: 10.1007/3-540-44886-1_25.

28. Cleland, J. G., & Van Ginneken, J. K. (1988). Maternal education and child survival in developing countries: the search for pathways of influence. Social science & medicine, 27(12), 1357-1368. doi: 10.1016/0277-9536(88)90201-8.

29. Béhague, D. P., Victora, C. G., & Barros, F. C. (2002). Consumer demand for caesarean sections in Brazil: informed decision making, patient choice, or social inequality? A population-based birth cohort study linking ethnographic and epidemiological methods. Bmj, 324(7343), 942. doi: 10.1136/bmj.324.7343.942.