

C-Section Prediction using Machine Learning Algorithms

Sanjam Kaur Bedi, Harleen Kaur

Abstract - The safety of the mother and the child largely depends on predicting the correct mode of delivery. Prediction is based on the opinion of the physician in charge of the delivery. If a wrong method is chosen, it could pose a great risk to the health of both the mother and the infant. Nowadays, many expecting mothers go for caesarean delivery without knowing whether it is suitable for them or not. In many developing countries, childbirth complications are a major delivery issue [1]. This study is divided into two parts: to identify the possible features used for determining the mode of childbirth and exploring and applying various machine learning algorithms on the best possible features predicting the mode of childbirth. To attain the objective, five different machine learning algorithms (KNN, Random Forest, Decision Tree, SVM, AdaBoost) have been applied. The dataset used for this research includes data of 4 different hospitals of Spain and contains a total of 6157 birth records and 161 features.

Keywords: Childbirth, Vaginal, Caesarean delivery, Supervised Machine Learning, Ensemble methods, AdaBoost, Random Forest, Decision Tree, SVM, KNN, SMOTE

1. Introduction

Mode of delivery these days is one amongst the foremost relevant issues that worry gynaecologists, mental health authorities and mothers. At the end of pregnancy, a baby leaves the womb of a mother either through the normal delivery, also called the vaginal delivery which is the ancient method of delivery, not involving any medical tools or through c-section delivery, which is the modern technology involving childbirth through surgical methods[3]. Vaginal delivery is thought-about the most popular technique of delivery involving lower morbidity and mortality rates than c-section. C-section delivers the baby through an incision made in the womb of the mother[2]. Both of these well-known deliveries have their own pros and cons and it may happen that the mode of delivery chosen for a particular woman might not suit the characteristics of that woman. This is happening in many developing countries such as Bangladesh, where the incorrect mode of delivery in the case of a woman causes temporary or permanent complications in her body. Incorrect methods of delivery pose various risks such as fatal termination, internal bleeding and respiratory issues in the baby in some cases.

C-section is performed often when a mother faces some complications in her delivery. This method is preferred in order to protect the health of the baby when normal delivery poses risks in the mother or the child. C-section helps in the cases when the labour stops or is very slow and not progressing, the baby is very large, the baby is in a breech position or the baby isn't able to get enough oxygen[4]. A lot of times a mother's health is taken into consideration while choosing the mode of delivery. If a mother has an infection like HIV or genital herpes, then C-section is chosen as vaginal delivery could transfer the infection to the baby[4]. A mother having high blood pressure or diabetes can have difficult normal delivery, therefore c-section is preferred in such cases. Normal delivery in these cases can damage the organs and cause other problems.

World right now is going through lots of developments in various fields. AI has taken over the majority of sectors. In such conditions, C-section, being the modern technology and regarded as a less painful procedure is widely pursued for. Many countries with less developed medical sectors are pushing their people towards c-section, not knowing that c-sections create loads of complications to both the mother and the kid. In Bangladesh itself, the rate of avertible c-section rate increased by 51% throughout 2016-2017 and 77% of all the c-sections in 2018 were medically unnecessary. During 2017 in Bangladesh, the maternal death rate was 173 over 100000 live births, which was higher than most of the developed countries. The same year, the United States saw 17 deaths per 100000 live births[3]. Also, according to the Institute of Public Health Nutrition's National Low Birth Weight Survey Bangladesh, in 2015, the caesarean section rate was 35% that exceeded the WHO's suggested range of 10-15%. Compared to normal delivery, maternal mortality and morbidity saw an increase to approximately twice the rate for c-section delivery[3].

Generally, the selection of the mode of delivery is that of the medical skilled in charge. A maternal aid app will assist the doctors in predicting the mode of delivery and facilitate in reducing the complications of childbirth. A lot of studies have been conducted in order to predict pregnancy results. Some of the

clinical decision support systems have also been developed. The research work already done involved developing a personalized prediction tool for normal birth after going through caesarean delivery using different ML algorithms. One such work included designing data mining classification models in order to predict modes of delivery using the factors associated with obstetric risk in real time. Other studies on this research are introduced in the later sections of this report. However, there is a need for further research to predict whether the characteristics of a particular will be suitable for a particular mode of delivery and which algorithms will provide best accuracy for predicting the c-section probability in a mother[6]. Therefore, this research has been conducted keeping certain objectives in mind:

1. Prioritizing features that are most important in predicting mode of childbirth
2. Exploring the machine learning algorithms to check which model will give best accuracy in the predictions. The later sections of this report are organized as follows:

The Literature Survey is briefly introduced in Chapter 2. The Proposed Solutions have been presented in Chapter 3. The Results of the accuracy of the applied machine learning models is presented in Chapter 4. Finally, the Discussion and Conclusion are presented in Chapter 5.

2. Literature review

A systematic and detailed literature review was carried out for this research project. Muhammad Nazrul Islam, et. al [3] conducted a study to find best machine learning algorithms to predict the mode of childbirth. They used algorithms such as Decision tree, random forest, KNN and SVM to predict the modes of childbirth. The accuracy of all the models ranged from 74% to 97% when applying the models on different features of low importance to high importance. The outcome showed the accuracy of all the applied models and the important factors that determine the mode of childbirth. It classified the factors into various categories from less important to most important. However, the research carried out was not easily understandable. It opened the door to many types of deliveries and the terms used are less known to people and more to medical practitioners.

Campillo-Artero et al[5]. did a study in order to determine the factors leading to emergency c-section using the ML classification Algorithms- Logistic regression, Classification tree and Random forest. The accuracy of the models in this research ranged from 74% to 81%. The result of this research focused on the mothers who were older than 35 and had higher body mass index and were more likely to have gone through c-section before.

S.A. Abbas et al.[6] did research to relate the probability of c-section with the age of the mother. The result of the study concluded that the mothers who were younger than 20 years and those who were older than 35 years have higher probability of going forward with c-section. The study also concluded that those women who go through c-section have higher blood pressure than those with normal deliveries[3].

Research by Md. Sakib Bin Alamet et al.[1] focused on the bagging ensemble classifiers. The study concluded that the bagging ensemble classifiers had performed better than the traditional machine learning algorithms in the domain of predicting childbirth.

Nafiz Imtiaz Khan et al.[7] conducted another study in which they used supervised ensemble machine learning models: Adaboost, Cat Boost and XGBoost. This study did not show very good accuracy results as only 11 features were considered in order to predict the c-section delivery.

Jeong Ha Wie et al.[8] did the study on prediction of emergency c-section using machine learning algorithms. Among the nine machine learning models used, logistic regression had the best performance. The study concluded that the pre-gestational maternal weight was associated with the c-section, whereas maternal weight at the time of delivery did not affect the prediction. It also concluded that the risk of c-section increased with increase in neonatal birth weight. The study also showed some limitations. The main limitation was that the study was a retrospective cohort study.

Mona M. Jambool[9] conducted research on data mining in health care to predict caesarean delivery operations. It resulted in the conclusion that naive Bayes approach outperformed all the other models used. The results were realistic due to the CV approach used while evaluating the performance of the classifier.

Thus, to summarize, a lot of concerns were observed throughout the literature review. Quite a number of studies focused on the prediction of c-section using different types of traditional ML algorithms, while some tried to find relation between various features of a woman and chances of c-section regarding those

features and some tried to find out the best features that are most important for determining the mode of childbirth. Therefore, keeping in mind the literature review, this study focuses on the prediction of caesarean section in a pregnant woman using specific machine learning algorithms as well as ensemble machine learning techniques to achieve a better accuracy.

3. Proposed Methodology

The research work to build the proposed model has been carried out in four phases namely Data Collection, Data Pre-processing, Training the model and Performance analysis of the model. The above Phases have been briefly described in the following subsections. The pictorial representation of the proposed model is shown in Fig 1.

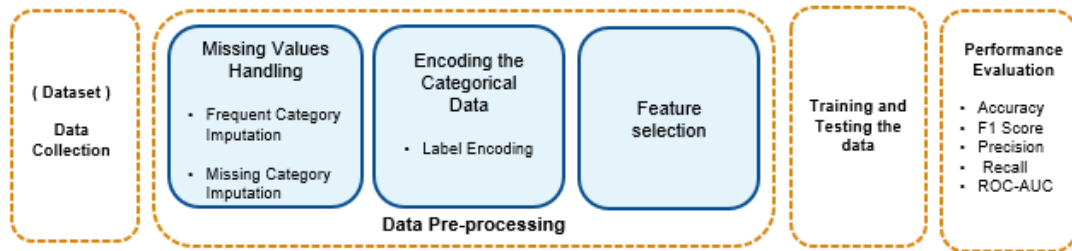


Fig 1. Proposed Model

Phase 1: Data Collection

The dataset used in this study was prepared by Campillo-Artero et al. [5], in order to build effective and efficient ML models. The dataset consists of birth records of four public hospitals of Spain in the year 2014. It consists of 6157 singleton birth records and a total of 161 features which includes 141 categorical and 19 numerical and 1 unnamed feature. The dataset also contained some columns which were unnamed and thus were removed from the dataset. It consists of some pre-delivery as well as post-delivery attributes which are useful in determining various predictions related to mother as well as child birth like, Amniotic Liquid, Oxytocin, Previous Caesarean, Age, Height, Weight, Foetal Intrapartum pH, Type of Birth etc.

Phase 2: Data Pre-processing

In this phase, collected data was synthesized. Duplicate data as well as noise were removed from the dataset. Null values were removed by applying different missing value imputation techniques like mean, median imputation for numerical features and frequent values as well as filling null values with a category called missing were performed. Machine Learning models can work upon numerical data only, due to this reason categorical encoding was performed. For this purpose, Label encoding which labels the categories into a numeric form was implemented. Since the objective of the study was to predict the C-section, therefore features which were irrelevant with the study were also removed. 32 important features among the left features were selected for the further analysis and model building purpose. The features used in the proposed model have been taken according to the reference paper [3].

The dataset collected consists of unbalanced data, grouped by type of birth. In Order to build the more effective and accurate classification Synthetic Minority Oversampling Technique (SMOTE) was implemented. This algorithm helps to overcome the imbalance data problem by synthetically generating the new instances using an interpolation method.

Phase 3: Training the Model:

Under this phase different well-known machine learning algorithms were selected which includes Decision Tree (DT), K-Nearest Neighbour (KNN), Support Vector Machine (SVM) to train the model. Ensemble machine learning techniques like Adaboost and Random Forest have also been implemented to classify the subject i.e. type of birth into C-section or non-C-section. Ensemble methods are supervised ML techniques that combine the predictions of several base models so as to produce one optimal model for prediction. This gives a better predictive performance compared to implementing a single model. These models are implemented using Scikit Learn which is a python module integrated with a wide range of Machine Learning algorithms [1].

3.1 Decision Tree (DT)- Decision tree is a supervised machine learning technique useful in both classification as well as regression. In this study the decision tree is used for classification purposes. It creates a predictive model in which features are interpreted in the form of tree structure. Prediction from the decision tree are the results of feature-based splits. Decision trees consist of root as well as leaf nodes. Root nodes decide which feature has to be selected for splitting the tree. It helps in evaluating the feature which best splits the data. On the other hand, leaf nodes are the terminating nodes of the tree. Root nodes or the deciding nodes of the tree can be selected using several methods like feature having high Entropy, feature having largest Information Gain, Gini index [10].

Steps followed in Decision Tree are

1. Input the dataset
2. Set the hyperparameters like criterion (entropy, gini)
3. Select the root node and split the data on the basis of gain mentioned above
4. Repeat the process until reach the stopping point (leaf node)
5. Train and test the model

The block diagram for the decision tree is given in Fig 2.

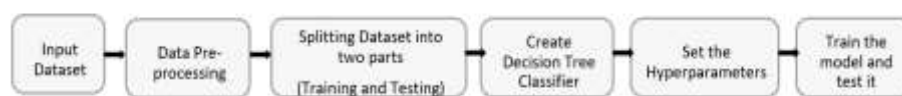


Fig 2. Flow diagram of Decision tree classifier

3.2 Random Forest (RF) - Random forest is an ensemble machine learning method which works on bagging techniques which involves dividing the complete data into several randomly generated subsets and executing base model in each randomly generated subset parallelly. Random Forest works well for both classifications furthermore as regression. It gives predictions on the basis of prediction of several base models. It is created from the subsets of data and the output at the end is based on the average or majority ranking. Decision tree is taken as the base model in case of random forest. It randomly selects the observations, builds a decision tree over there and the majority result is taken in case of classification. Each individual tree predicts the class and the majority class at the end is taken as the final prediction. This algorithm is comparatively slower than the decision tree due to multiple tree generation [11].

Steps followed in Random Forest are

1. Input the dataset
2. Set the hyperparameters like criterion (entropy, gini) and number of samples
3. Construct decision tree for each sample
4. Train and get the prediction from each decision tree
5. For each predicted result voting will be performed and majority voted prediction result is considered to be the final result

The block diagram for the random forest is shown in Fig 3.



Fig 3. Flow diagram of Random forest classifier

3.3 K-Nearest Neighbour (KNN) - KNN is a non-parametric lazy learner algorithm as it does not have any explicit assumptions about the data distribution and it takes less work for training and more work for making predictions. KNN imputes the similarity between the new data point and the available cases and puts the new data point into the category which is most similar to the available categories. This algorithm can be used for both classification as well as regression. KNN calculates the distance between the new data point and the other points present in the proximity and the shortest among them is filtered out. Distances between two data points are calculated using different distance measures like Euclidean

distance, Manhattan distance, Chebyshev distance. Due to these reasons KNN is termed as a distance-based algorithm [12].

Steps Followed in KNN are

1. Input the dataset
2. Set the hyperparameters like number of neighbours (k), distance
3. Calculate the distance between the new point and others using distance measures mentioned like Euclidean, Manhattan, Chebyshev
4. Sort the distances in ascending order and select first K entries
5. Get the labels of selected entries and return the mode of them
6. Train and test the model

The block diagram of KNN is shown in Fig 4.

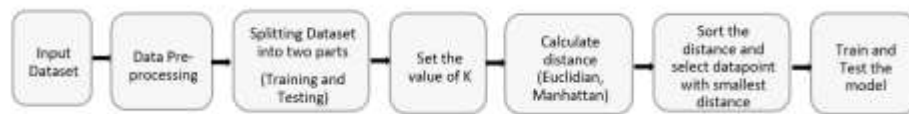


Fig 4. Flow diagram of KNN classifier

3.4 Support Vector Machine (SVM) - SVM algorithm is unique as compared to other machine learning algorithms as it has the capability to easily implement as well as easily handle multiple continuous and categorical variables. SVM can be used for both regression as well as classification purposes. In classification, the main aim of the SVM algorithm is to generate a best line or decision boundary which can segregate spaces into classes. This best decision boundary is called a hyperplane [3]. SVM technique finds a hyperplane which separates two hyperplanes with maximum margin. Sometimes dimensions of hyperplane are changed from 1 to n if the differentiation of classes is not simple and is called kernel.

Steps followed in SVM Technique

1. Input the dataset
2. Choose the kernel function like rbf, linear, polynomial
3. Train the models based on these kernels function
4. Make the prediction and test the model

The block diagram for SVM is shown in Fig 5. SVM

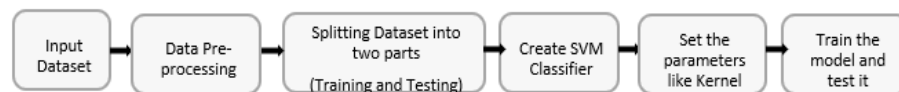


Fig 5. Flow diagram of SVM classifier

3.5 Adaboost (AB) - The AdaBoost is an ensemble machine learning technique which trains and deploys the trees in order to increase the efficiency of the classifiers. In order to build a strong predictive model from weak models, an iterative approach is used to learn from the mistakes of the weak classifier models. For every iteration, miss classified data points are identified and their weights are increased. On the other hand, weights of the correct points are decreased so that extra attention can be given to the wrongly classified points by the next following classifier and get them correct. In this technique, instances are drawn into the subsequent data sets from an iteratively updated sample distribution of the training data and the classifiers are combined through weighted majority voting [3].

Steps followed in Adaboost Technique are

1. Input the dataset

2. Set the hyperparameters like base estimator, number of estimators, learning rate and random state
3. Build the model and make predictions
4. Update the weights, assign higher weights to the misclassified points and lower weights to correctly classified points
5. Build the next model and make predictions
6. Repeat step 4 and 5 for the number of estimator's times
7. Using weighted average of individual model, make the ultimate model
8. Train and test the model

The block diagram for Adaboost is shown in Fig 6.



Fig 6. Flow diagram of Adaboost method

Phase IV: Performance Evaluation

Cross Validation technique is a statistical technique for performance evaluation of machine learning models. In this study Stratified K fold Cross Validation technique which works well for the imbalance dataset. In this the whole dataset is divided into k folds of equal size having the same ratio of instances of C-section and non-C-section.

F1 score - F1-score is a measure of a model's accuracy on a dataset. It is the way of combining precision and recall of a model and is the harmonic mean of the model's precision and recall [13]. F1 score is a good choice for comparing different models predicting the same thing.

Precision - Precision of a model refers to the percentage of its results which are relevant [14]. It is the ratio of correctly classified positive examples to the total positively classified examples.

Recall - Recall of a model refers to the percentage of the total relevant results correctly classified by the algorithm [15]. It is the ratio of correctly classified positive examples to the total actual positive examples.

ROC-AUC - Receive operating characteristics curve (ROC) - Area under the curve (AUC) is the performance measure at various threshold settings for the classification problem. ROC is a probability curve on the other hand AUC is a measure of separability and the graph from true positive rate to false positive rate.

Performance of every prediction model is measured in terms of accuracy, precision, recall, F1 score and ROC-AUC curve. ROC-AUC score is used for the final model selection. Results are briefly described in the following subsections.

4. Results

In this section, a detailed description of the observations of various regression methodologies is provided. We collected the dataset [5] and the features considered for experimentation are provided in below table 1.

Table 1. The features used in experimentations from the data set [10]

Previous Caesarean	Episiotomy	Start Week of Antenatal Care	Weight
Complications	Oxytocin	ART	BMI
Robson Group	Obstetric Risk	Previous Term Pregnancies	Age

ART Mode	Parity	Amniotic Liquid	Cardiotocography
Previous Preterm Pregnancies	Fatal Intrapartum Ph	Number of Miscarriages	Maternal Education
Amniocentesis	Comorbidity	Anaesthesia	Substance abuse
Preinduction	Number of Previous Caesarean	Gestational Stage	Smoking
Induction	Weight Increased during pregnancy	Height	Alcohol

4.1 Decision Tree: Below table 2 shows the results of Decision Tree considering different hyperparameters like criterion for splitting (entropy and gini index) and number of folds (3, 5, 10) for stratified K fold cross validation folds. From the below table it can be seen that the highest F1 score obtained for the Decision Tree is 0.94218 when entropy is considered as criterion for splitting and 10 folds cross validation is considered. Highest ROC-AUC is 0.89232 is obtained when the combination of gini index as criterion and 5 folds is considered as hyperparameters for training the model.

Table 2. Observations of Decision Tree method

Decision Tree		Precision	Recall	Accuracy	F1 Score	ROC-AUC
Criterion	K-Folds					
Entropy	3	0.81007	0.82340	0.92919	0.93308	0.88431
	5	0.79870	0.83902	0.93991	0.93861	0.89066
	10	0.80915	0.80878	0.93276	0.94218	0.88977
Gini Index	3	0.78305	0.80389	0.92919	0.93227	0.88783
	5	0.79651	0.81268	0.93422	0.93276	0.89232
	10	0.79586	0.80677	0.93421	0.93877	0.88898

4.2 Random Forest: Below table shows the results of Random Forest considering different hyperparameters like criterion for splitting (entropy and gini index), number of estimators (20, 30, 40, 50) and number of folds (3,5,10) for stratified K fold cross validation folds. From the below table it can be seen that the highest F1 score obtained for the Random Forest is 0.95355 when entropy as criterion, 50 estimators and 10 folds cross validation is considered. Highest ROC-AUC is 0.98513 is obtained when the combination of entropy as criterion, 50 estimators and 5 folds is considered as hyperparameters for training the model. The recorded observations are provided in below table 3.

Table 3. Observations of Random forest classifier method

Random Forest			Precision	Recall	Accuracy	F1 Score	ROC-AUC
Criterion	Number of Estimators	K_Folds					
	20	3	0.85336	0.85366	0.95176	0.95030	0.97976
		5	0.85811	0.86146	0.95160	0.95257	0.98154
		10	0.86301	0.84975	0.95030	0.95046	0.98141
	30	3	0.85407	0.85366	0.95160	0.95274	0.98119
		5	0.85544	0.86146	0.95322	0.95111	0.98251

Entropy		10	0.85271	0.85856	0.95485	0.95225	0.98258
	40	3	0.85711	0.84487	0.95095	0.95290	0.98354
		5	0.86522	0.86049	0.95128	0.95241	0.98400
		10	0.85792	0.86341	0.95338	0.95338	0.98451
	50	3	0.85595	0.85952	0.95241	0.95257	0.98259
		5	0.85387	0.86341	0.95241	0.95274	0.98261
		10	0.85808	0.86542	0.95306	0.95355	0.98513
Gini Index	20	3	0.85217	0.84682	0.95030	0.94786	0.98035
		5	0.86662	0.85171	0.95160	0.95209	0.97829
		10	0.85747	0.85371	0.95111	0.95030	0.98112
	30	3	0.85535	0.85074	0.95225	0.94998	0.98152
		5	0.85570	0.85854	0.95274	0.95290	0.98058
		10	0.85671	0.86726	0.95160	0.95241	0.98173
	40	3	0.85148	0.85756	0.95209	0.95127	0.98248
		5	0.85651	0.85659	0.95339	0.95209	0.98249
		10	0.85343	0.87223	0.95225	0.95062	0.98395
	50	3	0.86032	0.85269	0.95192	0.95063	0.98262
		5	0.85409	0.86829	0.95144	0.95306	0.98355
		10	0.85369	0.86828	0.95225	0.95160	0.98340

4.3 K Nearest Neighbor: Below table shows the results of KNN considering different hyperparameters like P (1 for Manhattan distance measure and 2 for Euclidean), number of neighbours (10, 15, 20, 25) and number of folds (3,5,10) for stratified K fold cross validation folds. From the below table it can be seen that the highest F1 score obtained for the KNN is 0.91278 when Manhattan as distance measure with 10 estimators and 10 folds cross validation is considered. Highest ROC-AUC is **0.96037** is obtained when the combination of Manhattan as distance measure with 25 neighbours and 5 folds is considered as hyperparameters for training the model. The observed results are given in below table 4.

Table 4. Observations of KNN classifier method

K Nearest Neighbour			Precision	Recall	Accuracy	F1 Score	ROC-AUC
P value (Minkowski)	Number of Neighbours	K_Folds					
	10	3	0.69348	0.84098	0.91506	0.91018	0.94515
		5	0.69764	0.82732	0.91327	0.90937	0.94840
		10	0.69943	0.83417	0.91343	0.91278	0.94992
	15	3	0.66243	0.87122	0.89963	0.90190	0.95339
		5	0.65373	0.86537	0.90418	0.90109	0.95452

P=1 (Manhattan)		10	0.65067	0.87127	0.90271	0.90320	0.95515
	20	3	0.67103	0.86830	0.90206	0.90612	0.95716
		5	0.90921	0.87122	0.90921	0.90872	0.95902
		10	0.68313	0.87322	0.90758	0.91100	0.95799
	25	3	0.65111	0.88878	0.89654	0.90109	0.95995
		5	0.65088	0.88195	0.90255	0.90125	0.96037
		10	0.65184	0.88201	0.90320	0.90304	0.95947
P=2 (Euclidean)	10	3	0.58807	0.86439	0.87299	0.87526	0.93629
		5	0.58136	0.86537	0.87786	0.87397	0.93750
		10	0.58990	0.87612	0.87608	0.87494	0.93982
	15	3	0.54066	0.90147	0.85578	0.85951	0.94375
		5	0.54588	0.89268	0.86081	0.85903	0.94228
		10	0.55094	0.89567	0.86146	0.85903	0.94307
	20	3	0.54879	0.90048	0.86260	0.86048	0.94354
		5	0.55058	0.89268	0.85773	0.85951	0.94547
		10	0.56096	0.89079	0.86422	0.86422	0.94498
	25	3	0.52609	0.90342	0.84717	0.84717	0.94654
		5	0.52920	0.90341	0.85058	0.84847	0.94673
		10	0.53865	0.90446	0.85366	0.85301	0.94786

4.4 Support Vector Machine: Below table shows the results of SVM considering different hyperparameters like kernel (Linear, rbf, Polynomial) and number of folds (3,5,10) for stratified K fold cross validation folds. From the below table it can be seen that the highest F1 score obtained for the SVM is 0.94721 when the rbf kernel is used with 10 folds cross validation is considered. Highest ROC-AUC is **0.97358** is obtained when the combination of the rbf kernel and 10 folds is considered as hyperparameters for training the model [15]. The observed results are provided in below table 5.

Table 5. Observations of SVM method

Support Vector Machine (SVM)		Precision	Recall	Accuracy	F1 Score	ROC-AUC
Kernel	K-Folds					
Linear	3	0.76637	0.90730	0.93877	0.94234	0.97121
	5	0.77390	0.91220	0.97121	0.93714	0.97148
	10	0.77041	0.90726	0.93584	0.93682	0.97139
Rbf	3	0.80859	0.89952	0.94803	0.94721	0.97190
	5	0.80474	0.90049	0.94673	0.94575	0.97235
	10	0.80274	0.89464	0.94591	0.94543	0.97358

Poly	3	0.75292	0.90829	0.93373	0.93666	0.97044
	5	0.76400	0.90829	0.93341	0.93341	0.97098
	10	0.74731	0.91021	0.97098	0.93324	0.97040

4.5 Adaboost: Below table 6 shows the results of Adaboost considering different hyperparameters like Decision Tree and Random Forest as base estimators, learning rate (0.01, 0.1, 1), number of estimators (20,30,40,50) and number of folds (3,5,10) for stratified K fold cross validation folds. From the below table it can be seen that the highest F1 score obtained for the Adaboost is 0.95322 when the Decision Tree as base estimators with 0.01 as learning rate and 30 estimators along with 10 folds cross validation is considered. Highest ROC-AUC is 0.98532 is obtained when Random Forest is considered as base estimator along with the combination of 0.1 as learning rate, 30 as number of estimators and 10 folds is considered as hyperparameters for training the model[16].

Table 6. Observations of Adaboost Method

Adaboost				Precision	Recall	Accuracy	F1 Score	ROC-AUC
Base Estimator	Learning Rate	Number of Estimators	K-Folds					
Decision Tree	0.01	20	3	0.78566	0.80777	0.93325	0.93357	0.88587
			5	0.78558	0.81366	0.93649	0.93308	0.88813
			10	0.80129	0.84773	0.93438	0.93487	0.89486
		30	3	0.79744	0.82046	0.93243	0.93422	0.88070
			5	0.78889	0.82439	0.93097	0.93520	0.88247
			10	0.79050	0.82054	0.93519	0.93227	0.88021
		40	3	0.77859	0.80875	0.93178	0.93162	0.88442
			5	0.78996	0.81659	0.93649	0.93503	0.88422
			10	0.80587	0.82145	0.93178	0.93422	0.88168
		50	3	0.78195	0.82339	0.93097	0.92415	0.89212
			5	0.79240	0.83610	0.93649	0.93065	0.88169
			10	0.79191	0.82147	0.93162	0.93649	0.88507
	0.1	20	3	0.78681	0.80095	0.93438	0.93114	0.88928
			5	0.79156	0.80878	0.93422	0.93406	0.88530
			10	0.79015	0.82046	0.93194	0.93324	0.88665
		30	3	0.77273	0.81949	0.93146	0.92837	0.88608
			5	0.78911	0.80488	0.93243	0.93211	0.89018
			10	0.80627	0.80971	0.93178	0.93210	0.88401
		40	3	0.78448	0.81265	0.93097	0.93357	0.88382
			5	0.78775	0.80683	0.93633	0.93438	0.88500
			10	0.79650	0.81652	0.93714	0.93422	0.89329

		50	3	0.77861	0.82047	0.93243	0.93260	0.88490
			5	0.78648	0.81073	0.93130	0.93243	0.88940
			10	0.79306	0.82824	0.93081	0.93487	0.89190
	1	20	3	0.78555	0.80584	0.93341	0.93146	0.88704
			5	0.79326	0.81854	0.93536	0.93227	0.88393
			10	0.79799	0.83024	0.93259	0.93568	0.88448
		30	3	0.77964	0.81557	0.93130	0.93292	0.88675
			5	0.79223	0.81561	0.93373	0.93308	0.88432
			10	0.79169	0.80191	0.93600	0.93145	0.88851
		40	3	0.76881	0.81562	0.92821	0.93406	0.88509
			5	0.79311	0.81659	0.93422	0.93373	0.88091
			10	0.80847	0.81849	0.93584	0.93324	0.88791
		50	3	0.77670	0.82536	0.93000	0.93016	0.88208
			5	0.79057	0.83024	0.92886	0.93406	0.88120
			10	0.80540	0.81261	0.93584	0.93357	0.89100
	0.01	20	3	0.84823	0.86829	0.95241	0.95176	0.98336
			5	0.84801	0.86732	0.95176	0.95225	0.98390
			10	0.85256	0.86734	0.95111	0.95192	0.98470
		30	3	0.85258	0.86439	0.95306	0.95192	0.98355
			5	0.85217	0.87024	0.95371	0.95274	0.98379
			10	0.85085	0.86735	0.95127	0.95322	0.98513
		40	3	0.84699	0.86244	0.95290	0.95241	0.98406
			5	0.84913	0.86341	0.95241	0.95176	0.98427
			10	0.84955	0.86637	0.95290	0.95241	0.98356
		50	3	0.85121	0.86244	0.95274	0.95225	0.98344
			5	0.84985	0.86634	0.95274	0.95306	0.98398
			10	0.84976	0.86446	0.95176	0.95144	0.98457
Random Forest		20	3	0.85103	0.86340	0.95225	0.95128	0.98326
			5	0.84481	0.86732	0.95306	0.95290	0.98418
			10	0.85112	0.86735	0.95257	0.95257	0.98470
		30	3	0.85380	0.86049	0.95160	0.95111	0.98403
			5	0.84878	0.86439	0.95111	0.95176	0.98483

	0.1		10	0.85418	0.86636	0.95225	0.95144	0.98532
		40	3	0.85164	0.86342	0.95274	0.95176	0.98410
			5	0.85135	0.87317	0.95274	0.95176	0.98367
			10	0.85168	0.86732	0.95355	0.95144	0.98509
		50	3	0.85457	0.86927	0.95225	0.95192	0.98372
			5	0.85200	0.86829	0.95225	0.95144	0.98440
			10	0.85099	0.86149	0.95306	0.95273	0.98449
	1	20	3	0.85209	0.86535	0.95290	0.95063	0.98317
			5	0.85134	0.86049	0.95176	0.95306	0.98429
			10	0.85817	0.86637	0.95192	0.95176	0.98407
		30	3	0.85736	0.86244	0.95241	0.95322	0.98425
			5	0.85018	0.85951	0.95111	0.95322	0.98470
			10	0.84864	0.86440	0.95273	0.95225	0.98481
		40	3	0.85253	0.86732	0.95387	0.95290	0.98394
			5	0.84557	0.86049	0.95404	0.95192	0.98426
			10	0.85124	0.86244	0.95338	0.95192	0.98508
		50	3	0.84705	0.86145	0.95241	0.95079	0.98451
			5	0.85227	0.86146	0.95209	0.95322	0.98407
			10	0.85321	0.87125	0.95290	0.95192	0.98517

Five classifiers were applied by applying stratified K fold cross validation. Performances of all the algorithms with different performance parameters like precision, recall and F1 score and ROC-AUC is shortlisted on the basis of ROC-AUC score of all the models with different parameters and shown in Table 7.

Table 7. Comparative results of Regression Methods

Model Used	Precision	Recall	Accuracy	F1 Score	ROC-AUC
Decision Tree	0.79651	0.81268	0.93422	0.93276	0.89232
Random Forest	0.85808	0.86542	0.95306	0.95355	0.98513
K Nearest Neighbour	0.65088	0.88195	0.90255	0.90125	0.96037
Support Vector Machine	0.80274	0.89464	0.94591	0.94543	0.97358
AdaBoost	0.85418	0.86636	0.95225	0.95144	0.98532

Thus, the final observation are as follows

Highest ROC-AUC score is observed as 0.98532 when ensemble machine learning technique Adaboost is considered to train the data with Random Forest base estimator along with the combination of 0.1 as

learning rate, 30 as number of estimators and 10 folds and second highest is obtained for Random Forest as 0.98513. Thus, it can be observed that ensemble machine learning techniques give better performance than others.

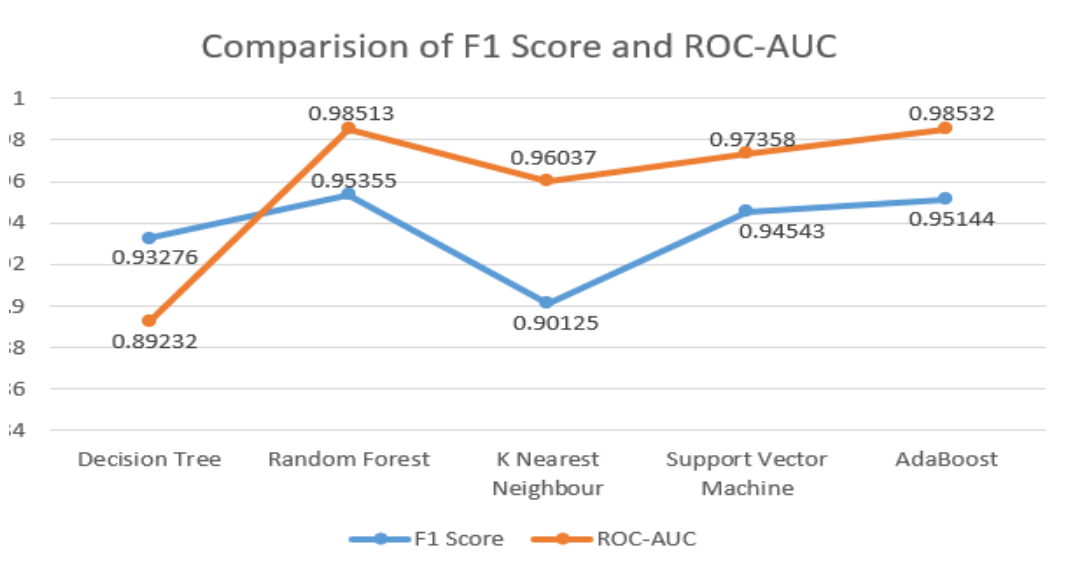


Fig 7. Graph comparing F1 scores and ROC-AUC score of all the model

5. Conclusion

Selecting suitable modes of childbirth is important for the safety of mother as well as new born child. But the best set of features for predicting can be explored more. This study examined all possible features for classification of type of birth and different models were applied. In this report, we compared a variety of traditional classifiers like Decision tree, KNN, SVM with the ensemble classifiers like AdaBoost and Random Forest which are the novel techniques used in prediction of childbirth mode. It has helped the doctors as well as the user to identify key factors that influence caesarean section. In this research, performances of all the algorithms with different performance parameters like precision, recall, F1 score and ROC-AUC are shortlisted on the basis of ROC-AUC score of all the models with different parameters. The research concluded with the AdaBoost model having the highest ROC-AUC score i.e. 0.98532 when it is considered to train the data with Random Forest base estimator along with the combination of 0.1 as learning rate, 30 as number of estimators and 10 folds. The second highest score was obtained for Random Forest i.e. 0.98513. Thus, it is observed that ensemble machine learning models show a better performance than the traditional machine learning models.

Future enhancements

In future, the models having the best accuracy can be used to make an online web application that would assist the doctors in the developing countries such as Bangladesh, to make predictions on the mode of childbirth on the basis of specific features associated with a woman. It would also assist expecting mothers to understand their body and take smart action regarding their mode of delivery. Advance machine

learning algorithms and Deep Learning can be used to increase the model performance. In future this can be deployed in the form of web application or mobile application for easy use.

References

- [1] [1]. Md. Sakib Bin Alam, Muhammed J. A. Patwary, Maruf Hassan. "Birth Mode Prediction Using Bagging Ensemble Classifier: A Case Study of Bangladesh", 2021 International Conference on Information and Communication Technology for Sustainable Development (ICICT4SD), 2021.
- [2] <https://reproductive-health-journal.biomedcentral.com/articles/10.1186/s12978-020-01021-7>
- [3] M.N. Islam, T. Mahmud, N. I. Khan, S. N. Mustafina And A.K.M.N. Islam "Exploring Machine Learning Algorithms to Find the Best Features for Predicting Modes of Childbirth" 2016 International Conference on Computing Communication Control and automation (ICCUBEA), 2016.
- [4] <https://www.marchofdimes.org/pregnancy/c-section-medical-reasons.aspx>
- [5] C. Campillo-Artero, M. Serra-Burriel, and A. Calvo-Pérez, "Predictive modeling of emergency cesarean delivery," *PloS one*, vol. 13, no. 1, p. e0191248, 2018.
- [6] Abbas, Syed & Riaz, Rabia & Kazmi, Syed & Rizvi, Dr. Sanam & Kwon, Se. (2018). Cause Analysis of Cesarean Sections and Application of Machine Learning Methods for Classification of Birth Data. *IEEE Access*. PP. 1-1. 10.1109/ACCESS.2018.2879115.
- [7] Muhammad Nazrul Islam, Tahasin Mahmud, Nafiz Imtiaz Khan, Sumaiya Nuha Mustafina, A. K. M. Najmul Islam. "Exploring Machine Learning Algorithms to Find the Best Features for Predicting Modes of Childbirth", *IEEE Access*, 2021
- [8] Wie, J.H.; Lee, S.J.; Choi, S.K.; Jo, Y.S.; Hwang, H.S.; Park, M.H.; Kim, Y.H.; Shin, J.E.; Kil, K.C.; Kim, S.M.; Choi, B.S.; Hong, H.; Seol, H.-J.; Won, H.-S.; Ko, H.S.; Na, S. Prediction of Emergency Cesarean Section Using Machine Learning Methods: Development and External Validation of a Nationwide Multicenter Dataset in Republic of Korea. *Life* **2022**, *12*, 604.
- [9] Jamjoom, Mona M. "Data Mining in Healthcare to Predict Cesarean Delivery Operations using a Real Dataset." *first International Conference on Computing and Emerging Sciences ICCE*. Vol. 2020. 2020.
- [10] <https://towardsdatascience.com/decision-tree-in-machine-learning-e380942a4c96>
- [11] <https://machinelearningmastery.com/bagging-and-random-forest-ensemble-algorithms-for-machine-learning/>
- [12] <https://medium.com/analytics-vidhya/k-nearest-neighbors-algorithm-7952234c69a4>
- [13] <https://deeptai.org/machine-learning-glossary-and-terms/f-score>
- [14] www.raw.githubusercontent.com
- [15] <https://www.freecodecamp.org/news/svm-machine-learning-tutorial-what-is-the-support-vector-machine-algorithm-explained-with-code-examples/>
- [16] <https://towardsdatascience.com/basic-ensemble-learning-random-forest-adaboost-gradient-boosting-step-by-step-explained-95d49d1e2725>