Intelligent Automation & Soft Computing

Tech Science Press

DOI: 10.32604/iasc.2023.034029

ARTICLE

Ensemble Modeling for the Classification of Birth Data

Fiaz Majeed[1], Abdul Razzaq Ahmad Shakir[1], Maqbool Ahmad[2], Shahzada Khurram[3], Muhammad Qaiser Saleem[4], Muhammad Shafiq[5,*], Jin-Ghoo Choi[5], Habib Hamam[6,7,8,9,10] and Osama E. Sheta[11]

[1]Department of Information Technology, University of Gujrat, Gujrat, 50700, Pakistan

[2]School of Digital Convergence Business, University of Central Punjab, Rawalpindi, 46000, Pakistan

[3]Faculty of Computing, The Islamia University of Bahawalpur, Bahawalpur, 63100, Pakistan

[4]College of Computer Science and Information Technology, Al Baha University, Al Baha, 1988, Saudi Arabia

[5]School of Computer Science and Engineering, Yeungnam University, Gyeongsan, 38541, Republic of Korea

[6]Faculty of Engineering, Uni de Moncton, Moncton, NB E1A3E9, Canada

[7]International Institute of Technology and Management, Commune d'Akanda, Libreville, 1989, Gabon

[8]Department of Electrical and Electronic Engineering Science, School of Electrical Engineering, University of Johannesburg, Johannesburg, 2006, South Africa

[9]Spectrum of Knowledge Production & Skills Development, Sfax, 3027, Tunisia

[10]College of Computer Science and Engineering, University of Ha'il, Ha'il, 55476, Saudi Arabia

[11]College of Science, Zagazig University, Zagazig, 44511, Egypt

*Corresponding Author: Muhammad Shafiq. Email: shafiq@ynu.ac.kr

Received: 04 July 2022 Accepted: 19 May 2023 Published: 06 September 2024

ABSTRACT

Machine learning (ML) and data mining are used in various fields such as data analysis, prediction, image processing and especially in healthcare. Researchers in the past decade have focused on applying ML and data mining to generate conclusions from historical data in order to improve healthcare systems by making predictions about the results. Using ML algorithms, researchers have developed applications for decision support, analyzed clinical aspects, extracted informative information from historical data, predicted the outcomes and categorized diseases which help physicians make better decisions. It is observed that there is a huge difference between women depending on the region and their social lives. Due to these differences, scholars have been encouraged to conduct studies at a local level in order to better understand those factors that affect maternal health and the expected child. In this study, the ensemble modeling technique is applied to classify birth outcomes based on either cesarean section (C-Section) or normal delivery. A voting ensemble model for the classification of a birth dataset was made by using a Random Forest (RF), Gradient Boosting Classifier, Extra Trees Classifier and Bagging Classifier as base learners. It is observed that the voting ensemble modal of proposed classifiers provides the best accuracy, i.e., 94.78%, as compared to the individual classifiers. ML algorithms are more accurate due to ensemble models, which reduce variance and classification errors. It is reported that when a suitable classification model has been developed for birth classification, decision support systems can be created to enable clinicians to gain in-depth insights into the patterns in the datasets. Developing such a system will not only allow health organizations to improve maternal health assessment processes, but also open doors for interdisciplinary research in two different fields in the region.

KEYWORDS
Birth data classification; ensemble model; machine learning

1 Introduction

There is a growing need for machine learning (ML) and data mining in a number of fields, including data analysis, prediction, image processing, etc., and predominantly in healthcare. ML is the process by which we use various computer algorithms to solve a specific problem that keeps improving with time and quality of the data set, resulting in more accurate predictions. Using these algorithms, we create a base model, input a data set called a training data set and then process that dataset to get some results. These results are compared to the actual results of the dataset. Testing data set refers to the portion of the dataset used to compare algorithm results. This process is repeated until we get a higher accuracy percentage [1]. A variety of ML algorithms are used in routine life to solve different types of problems to improve our quality of life. Several applications of ML are being used in different domains, including healthcare. The healthcare industry is vital to our lives. With the help of artificial intelligence (AI) and ML, we can save thousands of lives. Medical professionals can use AI and ML to make better decisions in different types of operations. Additionally, it allows them to handle a wide range of medical situations that require more accurate predictions. ML algorithms can be applied not only in medical state of affairs in emergency, but also in general primary care for medical patients. By means of these algorithms, doctors are able to select the most efficient operational method, specifically in the worst medical conditions while it is hard to forecast the results by analyzing the patient data [2]. Studies have been done to categorize the birth data to support the type of birth and doctors' forecast, such as normal delivery or cesarean section.

A cesarean section (C-Section) is a surgical method therein the fetus is carryforward by cutting the abdomen of the expectant mother. This is a life-saving technique when during pregnancy certain complications arise for both mother and baby. This is a key surgery with instant risks to both the pregnant mother and her child [3]. A cesarean section might be inevitable for an expecting mother in some cases, such as if there is more than one baby in the uterus, the baby is traversed, or if she has already had a cesarean delivery. The doctors recommend cesarean section delivery if normal delivery is unsafe for the expecting mother or child [4]. The use of cesarean sections in low-income countries and developed countries has intensely increased in recent decades. Research shows that the growing number of cesarean sections puts both perinatal and maternal lives along at stack and with their heaths on severe risks [5]. The health services of the United Kingdom (UK) estimate that the chances of death from cesarean section are 3–4 times bigger than that from a normal birth [6]. In 2012 alone, the recorded figure for cesarean section worldwide was 23 million. This is a veritably high rate of cesarean sections across the world [7]. Apart from that noted number, the home deliveries and deliveries in other private clinics are very high. So, these increasing numbers make this a burning issue. Moreover, it should be noted that if a woman goes with cesarean section multiple times, every time the situation became more complex and dangerous than the last time. Medical field concluded that the caesarean section disturbs the perinatal and maternal health. It is, therefore, necessary to analyze the related or physical factors contributing to the situation that requires a cesarean section.

Women face various types of experiences and medical implications during pregnancy, depending on their medical conditions and social background. The dynamic nature of living styles, the social life,

the region and the medical implications have a collective impact on the residents of society. Researchers are encouraged by these different types of parameters to conduct studies at a regional rather than global level in order to study the medical factors that contribute to maternal health and, therefore, to the health of the expecting child. Gathering information related to medical and social factors related to pregnant women would be useful at this stage. This information could be used later to predict whether the expecting child will be delivered by cesarean section or by normal delivery. This information can also be used in many other ways in addition to generating predictions about the type of birth. Using this information, we can identify the influential factors that can be helpful to physicians for handling the problems that are contributing towards the need for a cesarean section.

ML algorithms can be used to generate valuable information from the dataset and to learn hidden patterns and understand the structure of the data to make predictions and automate the decision-making process [1]. We can find hidden patterns by establishing a relationship between the predictor and independent variables through the ML methods. This relationship can be used to explain which factors affect the predictor variable and it can also be used to predict the outcomes of the dependent variable by using independent variables. When the data is already labeled with the independent variables, this process is known as supervised ML. In order to classify birth data, too many supervised ML algorithms [8] have been developed. In general, the purpose of all these algorithms for classification is to train the algorithms with the training dataset and then to measure their performance with the testing dataset. There are different categories of ML algorithms for classification. Some well-known methods for performing classification are Statistical learning methods for classification [9], decision trees [10] and classification based on perceptron [11], support vectors and ensemble learning methods [12]. Ensemble modeling is a concept of ML in which different classifiers are combined to reduce variance in decision-making processes and to improve accuracy. Literature shows that the ML algorithms using ensemble modeling techniques can make a strong classifier to generate better predictions and this technique has lowest possibility of classification errors [13].

In the current study, we used the birth dataset collected by [14] from the regional level so that we can apply the ML ensemble modeling on the women of a particular (targeted) region. The collected birth dataset is used to train the ensemble modeling technique of ML for discovering hidden patterns and understanding the structure of data to produce the best classification ensemble model and predict the birth outcomes. Different cleaning and reduction techniques were used to align this dataset.

For a doctor, it is likely impossible to capture all the required information by simply looking at historical data. Mistakes made by a doctor during the prediction of birth type can result in adverse outcomes, even death. In operations and decision-making, ML algorithms help doctors by reducing the risk of human error. However, using the algorithms individually can change the outcomes and may produce errors in classification. Therefore, a strong classification model is needed to improve the accuracy and reduce the variance in decision-making processes.

This paper has following contributions:

- We have enhanced the accuracy of ML algorithms by using ensemble modeling techniques, to reduce variance and classification errors for birth data and to predict the type of birth, such as cesarean section.
- Our study increases the performance of ML algorithms for birth dataset classification as well as to predict whether a birth type will be normal or a cesarean section. In order to resolve this issue, an ensemble modeling approach is used, which combines different ML algorithms to decrease variance and increase accuracy for predicting birth type in a specified region.

There is a detailed review of the literature in Section 2. The methodology used in this study and the description of dataset are in Section 3. Section 4 presents the research and related discussion supported by different evaluation methods. The summary of proposed work and future recommendations are discussed in Section 5.

2 Related Work

In the field of ML, researchers are constantly developing algorithms for predicting outcomes and diagnosing birth data. Using cardiotocographic readings of maternal contractions and fetal heart rate (FHR), the authors in [15] examined the temporal and complex relationships between the maternal contractions and FHR to classify vaginally (Normal) and cesarean section delivery types. Using these recordings, different algorithms of ML, i.e., decision trees, support vector machines (SVM) and AdaBoost algorithm [16] were trained for classifying cesarean section and normal delivery types. Among these classifiers, AdaBoost had the best accuracy. The authors of another study [17] used real test data to calculate the accuracy of algorithms used in ML for predicting the birth type from the birth dataset. Using signal processing, they extracted 13 features from the dataset of 552 raw FHR recordings from cardiotocography [18]. The results show that the ensemble modeling technique consisting of support vector machine, random forest (RF) and Fishers linear discriminant analysis produced the best results as compared to the individual performance of these ML classifiers.

In [19], the authors evaluated the performance of bagging and boosting classifiers for the classification of a regional birth dataset. They collected birth data from two government hospitals to conduct a study on a regional level. The bagging and boosting algorithms were trained using a testing dataset for the comprehensive comparison of boosting and bagging algorithms. In our study, BagFDA and Adabag algorithms performed best. The accuracy of BagFDA was high from all the boosting and bagging algorithms included in our study. In [20], the authors evaluated the performance of various supervised algorithms of ML and proposed a system to predict the decisions during health care operations. The system was tested on cesarean section (obstetric surgery most often performed) to save mothers and babies. In this case study, the RF and k-nearest neighbor (KNN) algorithms are better at classifying cesarean and normal delivery types.

In another study [19], the authors presented a study using clustering techniques to combine the relevant data into clusters. They also conducted a study on a regional level and they used a dataset collected by [14]. They used two different techniques for clustering. In the first technique, they used K-medoids and K-means algorithms for making clusters of the birth dataset using distance metrics. In the second technique, they transformed data using different techniques of transformation like scale, range and Yeo-Johnson and then used K-means and K-medoids to make clusters of data. They concluded that transformed data produces better results than raw data. K-means produced 67.58% accuracy, K-medoids and Rank K-medoids produced 69.58% and 62.64% accuracy, respectively. In [14], the authors collected data on birth type, i.e., cesarean section and normal deliveries, from public hospitals in their area of residence. There are not sufficient health facilities and equipment for pregnant women in these hospitals. A bivariate analysis identified the dominant factors, which are associated with pregnancy disorders, leading to a cesarean section. They have also created various models for the classification of birth using ML. Using various evaluation criteria, they concluded that the best algorithm is RF as compared to neural networks (NN), linear discriminant analysis (LDA), SVM and other classifiers used in this study.

The authors in [21] conducted a study to identify the factors that are associated with perinatal and neonatal mortality. After identifying the influential factors, they found that the neonatal mortality

rate from the 1000 live births was 31.4 and the perinatal mortality rate from 1000 pregnancies was 49.7. Robu et al. [22] conducted a study using several algorithms of ML to classify the birth data. They recorded 2325 reports of birth data from Gynecology Clinique and Bega Obstetrics to train the ML algorithms for the classification purpose. The purpose was to look over the relationship or associations among umbilical cord, neonatal cry, blood glucose, Apgar score and maternal body mass index prior to pregnancy. Logit Boost algorithms were used to develop a special application based on Weka application programming interface (API) to classify the birth outcomes. Another study [23] was conducted to examine whether the feature selection has any impact on the Naïve Bayes algorithm's performance feature selection effect on the performance of the Naive Bayes algorithm for patterns of fetal and FHR. Mutual Information, Information Gain, Correlation-based and ReliefF methods were used to select features. As a result, ReliefF was better as compared to the performance of other methods for the classification of fatal. They reported that feature selection methods have no significant effect on FHR classification.

Authors in another research [24] modeled signal pairs from cardiotocography that are FHR and uterine pressure (UP) for input and output systems. They also used power spectral density, calculated from the autoregressive model, for modeling the baseline of FHR to the linear fit and variability of FHR, independent of UP. SVM classifier was trained using normal and pathological cases from the perinatal database and feature set of this model. Only 7.5% false positive was detected with this method. Reference [25] examined the association between cesarean section and maternal age for understanding the increased rate of Caesarean section. They used multiple logistic regressions for this purpose. The results showed that pregnant women with an age of more than 44 years have medical complications, which increases the probability of cesarean section. Cesarean section rates in women aged 22 to 29 are lower than expected in older women.

A similar statistical analysis in [26] was used for evaluating the association between maternal age and deliveries with cesarean section. They made three groups of expecting women based on their ages. The expecting women below 35 years were categorized in one group, women between 35 to 39 years were categorized in another group and above 40 years in another group. The results indicate that there are more chances of having cesarean section delivery in the expecting women aged more than 40 and are prone to placenta abrupt and placenta Persia. The expecting women between the ages 35 to 39 were likely to have miscarriages and chromosomes defects. In addition, the authors of reference [27] have examined the association between fetal weight, age and cesarean section. This relationship was studied by using different methods like T-test, multiple logistic regression and Chi-square. They found that if the expecting baby weighed 3,600 grams or more, then there are more chances to have a cesarean section in the expecting women over the age of 35. The authors in [28] predicted preterm births by monitoring the Uterine Electrical Signals (Electro-hysterography). The dataset consisting of 38 preterm and 262 term records was used to conduct the study for classification. In comparison to the existing researches of time, the authors observed improvements.

The authors of [29] defined the classes by combining a latent class analysis-based model with the RF. This study concluded that this method is promising and likely to be useful in developing a decision support system instead of relying more on information from pH only. The previous section presented some research and their outcomes regarding the dominant social, physical and medical factors that influence maternal health and cause complicated situations leading towards the occurrence of C-Sections. Additionally, the application of ML methods is progressive in prognosis and diagnosis analysis. Several types of research have witnessed such applications towards the diagnosis and classification of factors associated with maternal health or birth outcome.

Previous studies used individual machine learning algorithms to predict birth outcomes. Research has shown that individual machine learning algorithms can generate variance and classification errors when making predictions. Therefore, we will use machine learning ensemble modeling techniques to reduce variance and classification errors. We will use ensemble modeling techniques to improve the performance of ML algorithms to classify birth dataset and predict whether the birth type is normal or C-Section.

3 Materials and Methods

The architecture of the methodology shown in Fig. 1 has the following modules:

- Preparing Dataset for training and testing (Birth dataset collected by [14])
- ML Algorithms (Gradient Boosting Classifier, Bagging Classifier, RF, Extra Trees Classifier)
- ML Ensemble Model
- Testing Performance of Ensemble Model

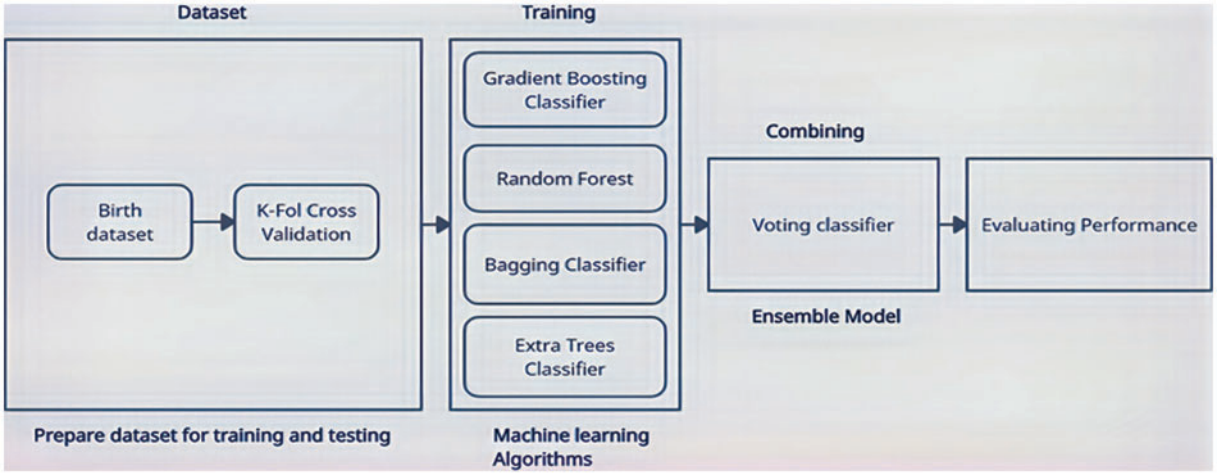

Figure 1: Architecture of the proposed methodology

Fig. 1 depicts a framework for building strong classification models by using ML ensembles. The dataset is divided into a training dataset (for the training of ensemble modeling of ML algorithms) and a testing dataset (for testing the accuracy of ensemble modeling of ML algorithms) using the K-Fold cross-validation technique. Following this, different algorithms of ML (Bagging Classifier, RF, Gradient Boosting Classifier and Extra Trees Classifier) are trained using a training dataset. The dataset was divided into 10 folds. We started from k−1 as test dataset and remaining folds as training dataset. All the algorithms were trained with training dataset and then we saved the result of each algorithms by testing on test dataset. After that, we used k−2 as test dataset and remaining folds as training dataset and tested each algorithm. This process was repeated for each fold. An ensemble model is then constructed using these algorithms. To make an ensemble model from the selected ML algorithms, the ensemble modeling Voting technique is used. Ensemble models and individual models are then tested on the birth dataset to determine which algorithms are most accurate.

3.1 Data Description

The data was collected from two different locations including the Combined Military Hospital Muzaffarabad and abbas institute of medical sciences (AIMS) in Muzaffarabad, Pakistan. The factors are related to the physical condition of the pregnant woman, her social life, her pre-pregnancy and her post-pregnancy characteristics. 488 observations were collected for the classification task. A total of 24 variables related to medical and health-related factors of pregnant women are collected by completing questionnaires in the presence of an Ob-gyn [14]. The study only examined variables related to health and birth outcomes. Women of different age groups, i.e., from 17 years to 45 years were under treatment during the data collection period. A higher number of C-Section cases were reported in earlier age subjects and women with ages exceeding 40 years. All women with a previous C-Section have been delivered surgically in their current pregnancy. Women with diabetes are inclined more towards C-Section birth as compared to the natural one. High blood pressure is also in positive correlation with C-Sections. Women with higher blood pressure are more likely to have C-Sections. The rate of natural deliveries has been reported in middle-aged groups, i.e., 25–30 years of age. Women with no previous surgeries, less usage of medicine and rich in iron tend to deliver naturally, as reported in data. The variables selected in the dataset are described in Table 1.

Table 1: Specification of the dataset

Historical factors	During pregnancy	Others
Age	Maternal age	Heartbeat
Surgeries (Not C-Section)	Bleeding	Inherited diseases
Previous C-Section	Blood pressure max	Breathing issues
Miscarriages	Blood pressure min	Fatigue
Abortions	Hemoglobin	Diabetes
Previous live births	Use of folic acid	BP problem
Menstruation	Medication	Hypertension
Menstrual cycle length	Headache	Iron Deficient

The distribution of normal deliveries and C-Sections is illustrated below in Fig. 2. Fig. 3 represents the frequency distribution of age according to C-Section. Fig. 4 illustrates the frequency distribution of age according to Normal Deliveries. Details of patients range from 17–40 are revealed in Fig. 5.

3.2 Ensemble Models Used for Classification

Ensemble models were initially designed to reduce variance in automated decision-making systems, thereby increasing accuracy. However, ensemble models have been effectively applied to several real life ML problems, including feature selection, confidence estimation, missing features, incremental learning, error correction, class imbalanced data, and learning concept drift for non-stationary distributions, etc. [30]. Three ensemble modeling techniques are discussed as follows. Firstly, the voting method in which ML uses the voting ensemble technique to solve the classification problem. Predictions of the outcome variable are made using all the base models in the voting ensemble technique. These predictions are taken as a vote by all the classifiers and the prediction from the majority of base models is used as an outcome of the dependent variable [31]. Secondly the averaging method, same as the voting technique, all the base models are used to make predictions of outcome

variables for all the data points in averaging. We then average all the predicted outcomes from all the base algorithms used in the ensemble and use this average as the final output of the ensemble model. This technique is used for regression analysis and finding the probability of models for classification problems [32]. Finally, the weighted average technique of ensemble modeling is the same as averaging technique. However, we assign different votes to each base model used in ensemble modeling which defines the importance of each model for predicting the outcomes [32].

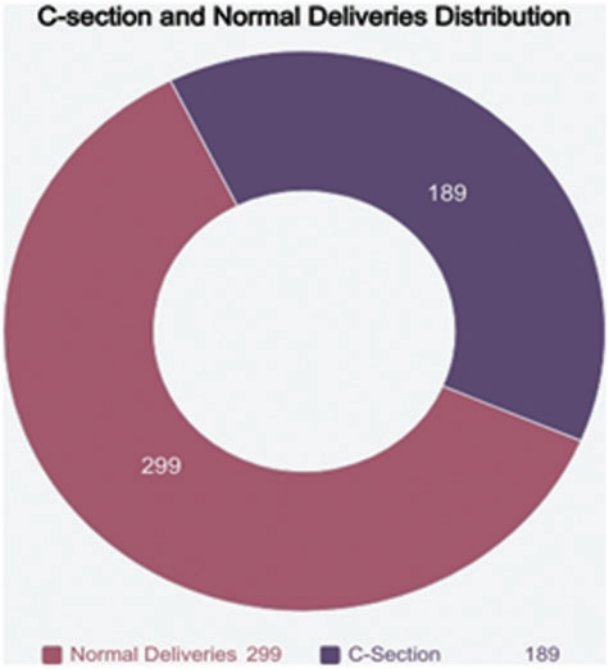

Figure 2: C-Section and normal deliveries distribution

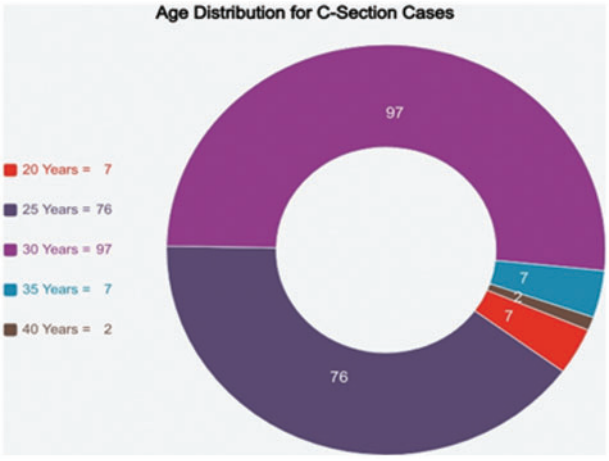

Figure 3: Age distribution for C-Section case

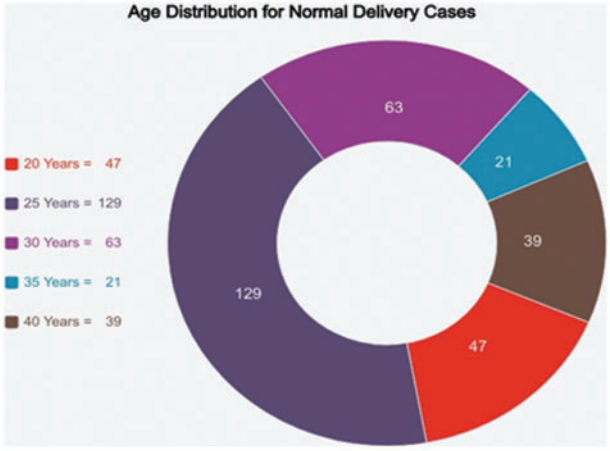

Figure 4: Age distribution for normal delivery case

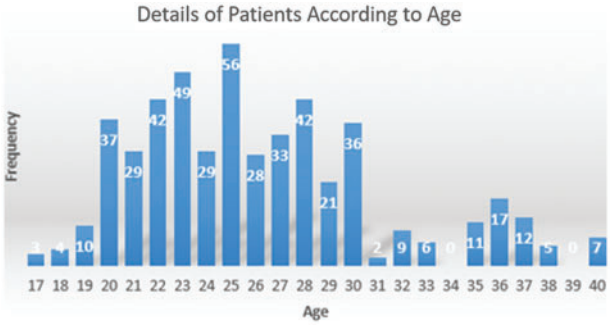

Figure 5: Details of patients according to age

In this study, ML algorithms are used to perform classification using a voting ensemble. Ensemble methods are those in which the numbers of weak learners (ML algorithms) are combined to produce a strong learner. This strong learner can give a better performance as compared to individual ML algorithms.

To classify birth data, we used the following ML algorithms to make different ensembles.

3.2.1 Random Forest

A RF is a meta-estimator that fits multiple decision trees to different subsamples of the dataset and uses the mean to improve the precision of predictions and control overfitting. The size of the subsample is controlled by the max_samples parameter when bootstrap = True (default). Otherwise, the dataset is used to build each tree [33].

The model for classification using RF is RF Classifier. For using RF in python we have to import "RF Classifier" from "sklearn" library. For the classification purpose, we use the "RandomForest-Classifier()" function using different parameters like "max_depth" and "random_state".

3.2.2 Extra Trees Classifier

This class implements a meta-estimator that includes several random decision trees (a.k.a. additional trees) to different subsamples of the dataset and uses the mean to improve the accuracy of predictions and control for over-fitting [33].

The model for classification using Extra Trees is the Extra_Trees_Classifier. For using the Extra Trees Classifier in python we have to import "Extra Trees Classifier" from "sklearn.ensemble" library. For the classification purpose, we use the "ExtraTreesClassifier()" function using different parameters like "n_estimators" and "random_state".

3.2.3 Bagging Classifier

A Bagging Classifier is an ensemble meta-estimator that fits a base classifier to a random subset of the original dataset and then aggregates their predictions to form the final prediction. Such meta-estimators can often be used to reduce the variance of black-box estimators (such as decision trees) by introducing randomization into their construction and then making them an ensemble [33].

The classification model is Bagging Classifier. For using Bagging Classifier in python we need to import "BaggingClassifier" from "sklearn.ensemble" library. For the classification purpose, we use "BaggingClassifier()" function using different parameters like "base_estimator", "n_estimators" and "random_state". As a base estimator for the Bagging Classifier, we used a decision tree classifier.

3.2.4 Gradient Boosting Classifier

Gradient Boosting Classifier creates an additive stepwise model. It allows enhancing arbitrary differentiable loss functions. At each step, the n_classes regression tree fits the negative gradient of the binomial or polynomial bias loss function. For using Gradient Boosting Classifier in python we have to import "GradientBoostingClassifier" from "sklearn.ensemble" library. For the classification purpose, we use "GradientBoostingClassifier ()" function [33].

3.2.5 Voting Classifier Algorithm

The Voting Classifier is a flexible voting/majority rule classifier. It is an ensemble model that uses various ML algorithms as the base learner and uses the voting rule to predict the outcome. In this study, we used previously discussed ML models, i.e., Gradient Boosting Classifier, Random forest, Extra Trees Classification and Bagging Classification as a base model to make an ensemble model based on the voting rule [33].

The classification model is Voting Classifier. We need to import "VotingClassifier" from "sklearn.ensemble" in order to use Voting Classifier in Python. For the classification purpose, we use "VotingClassifier()" function which takes the base models list as a parameter to predict the outcomes.

We used cross-validation to train the Voting Classifier. ML models are evaluated using cross-validation. It is a resampling technique of limited datasets that divides the dataset into k numbers of folds. Each fold is used to test and train a ML model. As a first step, the first fold is used to test the ML model, while other folds are used as a training dataset. After that, the second fold is used as a testing dataset and the rest as a training dataset and so forth. The pseudo code for the ensemble model is given in Algorithm 1.

Algorithm 1: Ensemble model

Input: Dataset, Base Learning Algorithms, Meta Learning Algorithm
Output: Average the predicted output of all base learners to classify the dataset
Process:

- Apply K-fold cross validation on dataset
- Train base algorithms with training dataset
- Use testing dataset to classify the training example
- Train the meta-learner
- Repeat this process for all learners

4 Results and Discussions

Following this section, we will present detailed results obtained after combining different ML algorithms into an ensemble. The classification analysis is performed using R studio and Python language, which provides the ability to apply different classifiers such as Extra Trees Classifier, Gradient Boosting Classifier, RF Classifier, Bagging Classifier and Voting Classifier.

4.1 Algorithms

The detailed methodology is already described in the previous section. Using K-Fold cross-validation, the birth dataset collected by [14] is used to train ML algorithms including Gradient Boosting Classifier, RF, Extra Trees Classifier and Bagging Classifier. We used 10 as K-value for cross validation. 20 estimators were used for all the individual ensemble models of machine learning with random state as 0. After training these algorithms, an Ensemble Modeling Voting Classifier is used to create the ensemble model of these algorithms and then the accuracy of this ensemble model is determined by mean cross-validation scores. The key to selecting algorithms for ensemble models is that the individual ML models or base models must conform to the particular accuracy criteria. The accuracy criterion for the current study is 80%, so we used those algorithms that have accuracy greater than 80%. The most important factor to consider when selecting the base models for the ensemble model is that most of the algorithms must have almost the same accuracy. In the current study, RF, Bagging Classifier, Gradient Boosting Classifier and Extra Trees Classifier have mean accuracy greater than 80% and RF, Bagging Classifier and Extra Trees Classifier have maximum and almost the same mean accuracy. These algorithms have been used for making a voting ensemble model.

4.2 Accuracy of Algorithms

RF is the first algorithm used for making an ensemble. The results produced by RF have a mean accuracy of 94.26 percent. Another algorithm for making an ensemble is a Gradient Boosting Classifier. The results of the Gradient Boosting Classifier yield an accuracy of 84.39%. The Extra Trees Classifier has provided results with a mean accuracy of 94.02%. The Bagging Classifier using a decision tree as a base learner yields results with a mean accuracy of 93.65%. Using Gradient Boosting Classifier, RF, Extra Trees Classifier and Bagging Classifier as base learners in Voting Ensemble Model have given the highest mean accuracy of 94.78% in this study. The following Fig. 6 shows the mean accuracy of all the described algorithms.

Mean Accuracy of RF, Bagging Classifier and Voting Classifier is almost the same, but the Voting Classifier has the highest mean accuracy.

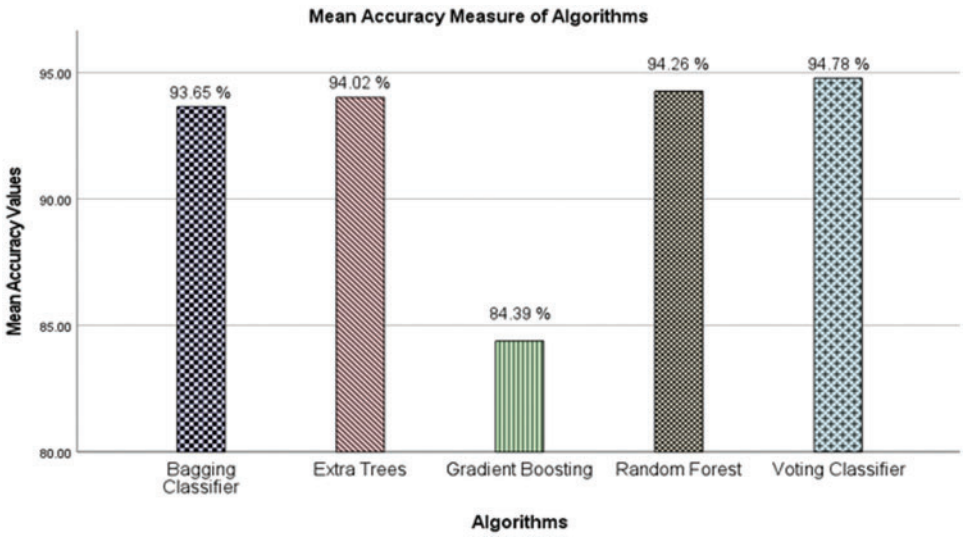

Figure 6: Mean accuracy measure of algorithms

4.3 Kappa Value Interpretation

Inter classifier variation can be evaluated in any condition in which two or more independent classification methods are assessing the same thing, by utilizing Kappa statistics. The evaluation is based on the dissimilarity between how much agreement is present ("observed" agreement) and how much agreement would be expected to be present by chance alone ("expected" agreement). The interpretations of Kappa statistic values are provided in Table 2. The value of Kappa exceeding 0.81 refers to the substantial agreement between observed and expected occurrences. Table 3 shows which classifier falls into a substantial agreement. RF and Voting Classifier have marginally higher mean Kappa measure as compared to the rest and Voting Classifier has the highest mean Kappa measure. The Voting Classifier has the highest mean Kappa value (0.8913) from all the individual ML algorithms.

Table 2: Kappa value interpretation

Sr #	Kappa value	Agreement
1	<0	Less than chance agreement
2	0.01–0.20	Slight agreement
3	0.21–0.40	Fair agreement
4	0.41–0.60	Moderate agreement
5	0.61–0.80	Substantial agreement
6	0.81–0.99	Almost perfect agreement

Table 3: Kappa measures of algorithms

Algorithms	Mean Kappa	Mean Kappa %age
RF	0.8769	87.69%
Gradient Boosting Classifier	0.6628	66.28%
Extra Trees Classifier	0.8728	87.28%
Bagging Classifier	0.8633	86.33%
Voting Classifier	0.8913	89.13%

4.4 Balanced Accuracy, Precision, and Recall

Here are some more statistics for the models which are used to calculate accuracy, precision, recall and balanced accuracy for models. Key terms used in these statistics are as follows:

- TP (True positive) = case was positive (C-Section) and predicted positive (C-Section).
- TN (True Negative) = case was negative (Normal Delivery) and predicted negative (Normal Delivery).
- FP (False Positive) = case was negative (Normal Delivery) and predicted positive (C-Section).
- FN (False Negative) = case was positive (C-Section) and predicted positive (Normal Delivery).

Balance Accuracy avoids inflated performance estimates on imbalanced datasets. If all the classes in the dataset have equal entries, then the balanced accuracy is equal to the accuracy of algorithms. The balanced accuracy in binary and multiclass classification problems is to deal with imbalanced datasets. It can be the average recall acquired in each class. The balanced accuracy can be written as follows:

$$Balanced\ Accuracy = \frac{1}{2}\left(\frac{TP}{TP+FN} + \frac{TN}{TN+FP}\right) \tag{1}$$

Precision and recall are important models of evaluation metrics. Precision states the percentage of relevant results, while recall denotes to the percentage of relevant results correctly classified by the algorithm in total.

We can evaluate the precision and recall as follows:

$$Precision = \frac{TP}{(TP+FP)} \tag{2}$$

$$Recall = \frac{TP}{(TP+FN)} \tag{3}$$

The results for balanced accuracy of C-Section and normal deliveries, precision and recall for Gradient Boosting Classifier, RF, Extra Tress Classifier, Bagging Classifier and Voting Ensemble Classifier are shown in the following Table 4.

Fig. 7 illustrates the summarized results of Accuracy, balanced accuracy, Precision and Recall for the RF, Gradient Boosting Classifier, Extra Trees Classifier, Bagging Classifier and Voting Classifier. We observe that the Voting Classifier consisting of Gradient Boosting Classifier, RF, Extra Trees Classifier and Bagging Classifier had the best accuracy with 94.78%, which is higher than the individual accuracy of ML algorithms, namely Gradient Boosting Classifier has 84.39% accuracy, RF

has 94.26% accuracy, Extra Trees Classifier has 94.02% accuracy and Bagging Classifier has 93.65% accuracy.

Table 4: Balanced accuracy, precision and recall of algorithms

Algorithms	Balanced accuracy	Precision	Recall
RF	0.938	0.944	0.942
Gradient Boosting Classifier	0.830	0.85	0.843
Extra Trees Classifier	0.937	0.943	0.94
Bagging Classifier	0.931	0.939	0.936
Voting Classifier	0.946	0.948	0.947

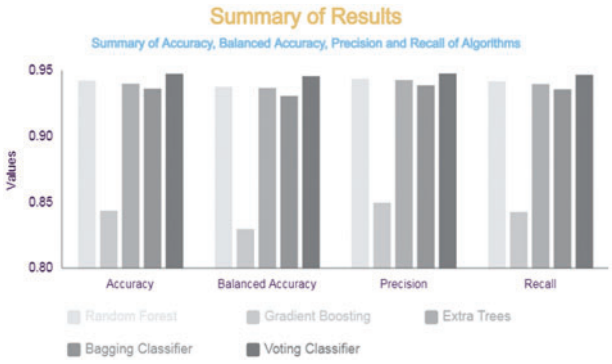

Figure 7: Summary of results

5 Conclusions

The ensemble ML aims to reduce variance and thus improve the accuracy of different ML algorithms by combining them. The purpose of this study is to apply different ML algorithms to show the high accuracy of birth date, and then combine these algorithms to improve the accuracy of birth results by the voting technique of the ensemble model, i.e., normal delivery or C-Section. However, ensemble models have since been successfully applied to a variety of ML problems including feature selection, missing features, incremental learning, confidence estimation, error correction, class imbalanced data and learning concept drift from non-stationary distributions. Our study uses cross-validation (CV) to train gradient boosting, extra trees, bagging and a Voting Classifier algorithm on a birth dataset. Each of these algorithms was used for classification purposes. The results show that the accuracy voting ensemble of proposed algorithms is higher than the previous studies. This study aimed to make an ensemble of different ML algorithms and to check their accuracy in classification. In the next phase, ML algorithms and ensemble model voting techniques are to be used to perform a comprehensive comparison. As it is believed that when a suitable classification model has been developed for birth classification, decision support systems can be created to enable clinicians to gain in-depth insights into the patterns in the datasets. Based on this study, an automated solution using ensemble modeling can be provided against the best prediction model in order to assist a physician in taking precautionary measures to ensure the health of the mother and the fetus. Social factors, which are included in current data but not included in this study, can also be looked at to understand their

contribution to C-Sections. By incorporating more medical, physical, and social factors, the data set can also be extended to other regions of Pakistan.

Acknowledgement: The author acknowledges Natural Sciences and Engineering Research Council of Canada (NSERC) and New Brunswick Innovation Foundation (NBIF) for the financial support of the global project.

Funding Statement: The author thanks Natural Sciences and Engineering Research Council of Canada (NSERC) and New Brunswick Innovation Foundation (NBIF) for the financial support of the global project. These granting agencies did not contribute in the design of the study and collection, analysis, and interpretation of data.

Author Contributions: The authors confirm contribution to the paper as follows: study conception and design: Fiaz Majeed, Abdul Razzaq Ahmad Shakir, Maqbool Ahmad, Shahzada Khurram, Muhammad Qaiser Saleem; data collection and creation: Muhammad Shafiq, Jin-Ghoo Choi, Habib Hamam, Osama E. Sheta; data curation, analysis and interpretation of results: Muhammad Shafiq, Fiaz Majeed, Abdul Razzaq Ahmad Shakir, Maqbool Ahmad, Shahzada Khurram, Muhammad Qaiser Saleem; draft investigation and revision: Muhammad Shafiq, Jin-Ghoo Choi, Habib Hamam, Osama E. Sheta. All authors reviewed the results and approved the final version of the manuscript.

Availability of Data and Materials: The data underlying this article will be shared (after the patent of the underlying research is filled) upon reasonable request to the corresponding author.

Ethics Approval: Not applicable.

Conflicts of Interest: The authors declare that they have no conflicts of interest to report regarding the present study.

References

[1] D. Sarkar, R. Bali, and T. Sharma, *Practical Machine Learning with Python*, 1st ed., Berkely: Apress, 2018.
[2] S. G. Kanakaraddi, K. C. Gull, J. Bali, A. K. Chikaraddi, and S. Giraddi, "Disease prediction using data mining and machine learning techniques," in *Advanced Prognostic Predictive Modelling in Healthcare Data Analytics*, 1st ed., Singapore: Springer, 2021, vol. 1, pp. 71–92.
[3] K. D. Gregory, S. Jackson, L. Korst, and M. Fridman, "Cesarean versus vaginal delivery: Whose risks? whose benefits?," *Am. J. Perinatol.*, vol. 29, no. 1, pp. 7–18, 2012. doi: 10.1055/s-0031-1285829.
[4] F. P. Guignard, "A gendered bun in the oven. The gender-reveal party as a new ritualization during pregnancy," *Stud. Relig./Sci. Relig.*, vol. 44, no. 4, pp. 479–500, 2015. doi: 10.1177/0008429815599802.
[5] P. Lumbiganon *et al.*, "Method of delivery and pregnancy outcomes in Asia: The WHO global survey on maternal and perinatal health 2007–08," *The Lancet*, vol. 375, no. 9713, pp. 490–499, 2010. doi: 10.1016/S0140-6736(09)61870-5.
[6] A. M. Romano, "Research summaries for normal birth," *J. Perinat. Educ.*, vol. 17, no. 1, pp. 48–52, 2008. doi: 10.1624/105812408X266304.
[7] G. Molina *et al.*, "Relationship between cesarean delivery rate and maternal and neonatal mortality," *J. Med. Assoc. Netw.*, vol. 314, no. 21, pp. 2263–2270, 2015. doi: 10.1001/jama.2015.15553.
[8] A. Singh, N. Thakur, and A. Sharma, "A review of supervised machine learning algorithms," in *Proc. of Int. Conf. Comput. Sustain. Global Develop.*, New Delhi, India, 2016, pp. 1310–1315.
[9] G. James, D. Witten, T. Hastie, and R. Tibshirani, "An introduction to statistical learning," New York: Springer, vol. 112, p. 18, 2013.

[10] V. A. Hernández, R. Monroy, M. A. Medina-Pérez, O. Loyola-González, and F. Herrera, "A practical tutorial for decision tree induction: Evaluation measures for candidate splits and opportunities," *ACM Comput. Surv.*, vol. 54, no. 1, pp. 1–38, 2021. doi: 10.1145/3429739.
[11] S. I. Gallant, "Perceptron-based learning algorithms," *IEEE Trans. Neural Netw.*, vol. 2, no. 2, pp. 179–191, 1990. doi: 10.1109/72.80230.
[12] O. Sagi and L. Rokach, "Ensemble learning: A survey," *Wiley Interdiscip. Rev.: Data Min. Knowl. Discov.*, vol. 8, no. 4, pp. 386, 2018, Art. no. e1249. doi: 10.1002/widm.1249.
[13] C. Zhang and Y. Ma, "Ensemble machine learning: Methods and applications," in *Springer Science and Business Media*, 1st ed., New York, USA: Springer, 2012.
[14] S. A. Abbas, R. Riaz, S. Z. H. Kazmi, S. S. Rizvi, and S. J. Kwon, "Cause analysis of caesarian sections and application of machine learning methods for classification of birth data," *IEEE Access*, vol. 6, pp. 67555–67561, 2018. doi: 10.1109/ACCESS.2018.2879115.
[15] S. Saleem, S. S. Naqvi, T. Manzoor, A. Saeed, and J. Mirza, "A strategy for classification of vaginal vs. cesarean section delivery: Bivariate empirical mode decomposition of cardiotocographic recordings," *Front. Physiol.*, vol. 10, Art. no. 246, 2019. doi: 10.3389/fphys.2019.00246.
[16] T. Hastie, S. Rosset, J. Zhu, and H. Zou, "Multi-class adaboost," *Stat. Interface*, vol. 2, no. 3, pp. 349–360, 2009. doi: 10.4310/SII.2009.v2.n3.a8.
[17] P. Fergus, M. Selvaraj, and C. Chalmers, "Machine learning ensemble modelling to classify caesarean section and vaginal delivery types using Cardiotocography traces," *Comput. Biol. Med.*, vol. 93, no. 3, pp. 7–16, 2018. doi: 10.1016/j.compbiomed.2017.12.002.
[18] R. M. Grivell, Z. Alfirevic, G. M. Gyte, and D. Devane, "Antenatal cardiotocography for fetal assessment," *Cochrane Database Syst. Rev.*, vol. 1, no. 9, pp. 1–26, 2015. doi: 10.1002/14651858.CD007863.pub4.
[19] S. A. Abbas, A. Aslam, A. U. Rehman, W. A. Abbasi, S. Arif and S. Z. H. Kazmi, "K-means and K-medoids: Cluster analysis on birth data collected in city Muzaffarabad, Kashmir," *IEEE Access*, vol. 8, pp. 151847–151855, 2020. doi: 10.1109/ACCESS.2020.3014021.
[20] M. Amin and A. Ali, "Performance evaluation of supervised machine learning classifiers for predicting healthcare operational decisions," *Wavy AI Res. Foundation: Lahore, Pak.*, vol. 90, pp. 1–7, 2018.
[21] J. B. Warren, W. E. Lambert, R. Fu, J. M. Anderson, and A. B. Edelman, "Global neonatal and perinatal mortality: A review and case study for the Loreto Province of Peru," *Res. Rep. Neonatol.*, vol. 2, pp. 103–113, 2012. doi: 10.2147/RRN.S33704.
[22] R. Robu and S. Holban, "The analysis and classification of birth data," *Acta Polytechnica Hungarica*, vol. 12, no. 4, pp. 77–96, 2015.
[23] M. E. B. Menai, F. J. Mohder, and F. Al-mutairi, "Influence of feature selection on Naïve Bayes classifier for recognizing patterns in cardiotocograms," *J. Med. Bioeng.*, vol. 2, no. 1, pp. 66–70, 2013. doi: 10.12720/jomb.2.1.66-70.
[24] P. A. Warrick, E. F. Hamilton, D. Precup, and R. E. Kearney, "Classification of normal and hypoxic fetuses from systems modeling of intrapartum cardiotocography," *IEEE Trans. Biomed. Eng.*, vol. 57, no. 4, pp. 771–779, 2010. doi: 10.1109/TBME.2009.2035818.
[25] M. Dulitzki, D. Soriano, E. Schiff, A. Chetrit, S. Mashiach and D. S. Seidman, "Effect of very advanced maternal age on pregnancy outcome and rate of cesarean delivery," *Obstet. Gynecol.*, vol. 92, no. 6, pp. 935–939, 1998. doi: 10.1016/s0029-7844(98)00335-4.
[26] J. Cleary-Goldman *et al.*, "Impact of maternal age on obstetric outcome," *Obstet. Gynecol.*, vol. 105, no. 5, pp. 983–990, 2005. doi: 10.1097/01.AOG.0000158118.75532.51.
[27] K. Fuma, Y. Maseki, A. Tezuka, M. Kuribayashi, H. Tsuda and M. Furuhashi, "Factors associated with intrapartum cesarean section in women aged 40 years or older: A single-center experience in Japan," *J. Mater.-Fetal Neonatal Med.*, vol. 34, no. 2, pp. 216–222, 2021. doi: 10.1080/14767058.2019.1602601.
[28] P. Fergus *et al.*, "Prediction of preterm deliveries from EHG signals using machine learning," *PLoS One*, vol. 8, no. 10, 2013, Art. no. e77154. doi: 10.1371/journal.pone.0077154.

[29] J. Spilka, G. Georgoulas, P. Karvelis, V. Chudáček, C. D. Stylios and L. Lhotska, "Discriminating normal from "abnormal pregnancy cases using an automated fhr evaluation method," in *Proc. Hellenic Conf. Artif. Intell.*, QC, Canada, 2014, pp. 521–531.
[30] R. Polikar, "Ensemble learning," in *Ensemble Machine Learning*. Switzerland: Springer, pp. 1–34, 2012.
[31] I. Gandhi and M. Pandey, "Hybrid ensemble of classifiers using voting," in *Proc. Int. Conf. Green Comput. Internet Things*, Greater Noida, India, 2015, pp. 399–404.
[32] G. Ke *et al.*, "LightGBM: A highly efficient gradient boosting decision tree," *Adv. Neural Inf. Process Syst.*, vol. 30, no. 1, pp. 3146–3154, 2017.
[33] G. Lemaître, F. Nogueira, and C. K. Aridas, "Imbalanced-learn: A python toolbox to tackle the curse of imbalanced datasets in machine learning," *J. Mach. Learn. Res.*, vol. 18, no. 1, pp. 559–563, 2017.