



Applied Artificial Intelligence

An International Journal

ISSN: (Print) (Online) Journal homepage: <https://www.tandfonline.com/loi/uaai20>

Analysis of Birth Data using Ensemble Modeling Techniques

Sohaib Latif, Xian Wen Fang, Kaleem Arshid, Abdullah Almuhaimeed, Azhar Imran & Mansoor Alghamdi

To cite this article: Sohaib Latif, Xian Wen Fang, Kaleem Arshid, Abdullah Almuhaimeed, Azhar Imran & Mansoor Alghamdi (2023) Analysis of Birth Data using Ensemble Modeling Techniques, Applied Artificial Intelligence, 37:1, 2158273, DOI: [10.1080/08839514.2022.2158273](https://doi.org/10.1080/08839514.2022.2158273)

To link to this article: <https://doi.org/10.1080/08839514.2022.2158273>



© 2023 The Author(s). Published with license by Taylor & Francis Group, LLC.



Published online: 28 Feb 2023.



Submit your article to this journal [↗](#)



Article views: 606



View related articles [↗](#)



View Crossmark data [↗](#)

Analysis of Birth Data using Ensemble Modeling Techniques

Sohaib Latif^a, Xian Wen Fang ^a, Kaleem Arshid^b, Abdullah Almuhaimeed^c, Azhar Imran^d, and Mansoor Alghamdi^e

^aSchool of Mathematics and Big Data, Anhui University of Science and Technology, Huainan, Anhui, China; ^bBeijing Key Laboratory of Trusted Computing, Faculty of Information Technology, Beijing University of Technology, Beijing, China; ^cDigital Health Institute, King Abdulaziz City for Science and Technology, Riyadh, Saudi Arabia; ^dFaculty of Computing and Artificial Intelligence, Air University, Islamabad, Pakistan; ^eDepartment of Computer Science, Applied College, University of Tabuk, Tabuk, Saudi Arabia

ABSTRACT

Machine learning and data mining are being used in different fields like data analysis, prediction, image processing, etc., and particularly in healthcare. Over the past decade, several types of research have been carried out focusing on machine learning and data mining application to generate intuitions from historical data and make predictions about the results. Machine learning algorithms play a vital role in improving healthcare systems due to continuous research in machine learning applications. Several researchers have used algorithms of machine learning to develop systems for decision support, analyze clinical aspects, use historical data to extract useful information, make future predictions and categorize diseases, etc. to help physicians make better decisions. In this study, we used an ensemble modeling voting technique for the classification of the birth dataset. Ensemble models combine individual machine learning algorithms to improve the accuracy by predicting from the combined output of the base classifiers. Gradient boosting classifier (GBC), random forest (RF), bagging classifier (BC), and extra trees classifier (ETC) were used as base learners for making a voting ensemble model for the classification of the birth dataset. The results produced have shown that the voting classifier of support vector machine (SVM), random forest (RF), extra trees classifier, and bagging classifier has given the best results with the proportion of 94.78%, gradient boosting classifier has 84.39% accuracy, the random forest has 94.26% accuracy, extra trees classifier have 94.02% accuracy and bagging classifier has 93.65% accuracy. The accuracy achieved by ensemble modeling is far higher than the machine learning algorithms. Ensemble models increase the accuracy of machine learning algorithms by reducing variance and classification errors. The development of such a system will not only help health organizations to take effective measures to improve the maternal health assessment process but will also open the doors for interdisciplinary research in two different fields in the region.

ARTICLE HISTORY

Received 1 October 2022
Revised 23 November 2022
Accepted 7 December 2022

CONTACT Azhar Imran  azhar.imran@mail.au.edu.pk  Faculty of Computing and Artificial Intelligence, Air University, Islamabad 42000, Pakistan

© 2023 The Author(s). Published with license by Taylor & Francis Group, LLC.
This is an Open Access article distributed under the terms of the Creative Commons Attribution License (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Introduction

Machine learning is a domain that provides different computer algorithms to solve some specific problem and keep on improving with the passage of time and the quality of the dataset. These algorithms build a base model and inject a dataset into it called a training dataset. From the training dataset, we obtain some results after algorithmic processing and then compare the generated results with the actual results of the dataset. The portion of the dataset which is used to compare the results of the algorithm to calculate accuracy is called the testing dataset. After this, we get an accuracy percentage, and we keep on repeating this process until we get a better accuracy percentage (Goodfellow, Bengio, and Courville 2016). There are numerous applications of machine learning, some of which are being used in different domains and particularly in healthcare. Healthcare is a vital part of our lives and we can use machine learning to save thousands of human lives. It helps doctors in different types of operational and decision-making activities. These algorithms facilitate physicians to choose the effective and better operational method, specifically, in the worst medical situations where it is difficult to predict the results when analyzing patients (Milovic and Milovic 2012). Many studies have been conducted to classify the birth data to help doctors to predict the birth type, i.e., normal delivery or cesarean section (C-Section).

Cesarean section is a surgical technique in which an expecting child is delivered by cutting the expecting mother's abdomen. It is a life-saving method for the mother and baby when certain complications appear during pregnancy. This is a major operation that involves immediate risks for both the expecting child and the mother (Gregory et al. 2012). There are some cases when a cesarean section becomes unavoidable for an expecting woman, these cases include more than one baby in the Uterus, traverse position of baby, already having cesarean section deliveries, substantial infant, and so forth. If normal delivery put the health of the expecting mother or child at risk, then the physicians suggest cesarean section delivery (Pasche Guignard 2015). In the last decades, the use of cesarean section in middle-income countries and developed countries is dramatically increasing worldwide. Studies show that the increasing number of cesarean sections can put both perinatal and maternal health at risk and disturbs perinatal and maternal health (Lumbiganon et al. 2010). According to British health services, the risk of death by cesarean section is three to four times higher than during a normal birth. In 2012 alone, the recorded figure for cesarean section was 23 million worldwide. This is a very high rate of cesarean sections across the world (Molina et al. 2015). Apart from that noted number, the

home deliveries and deliveries in other private clinics are very high. So, this alarming increase in numbers makes this a burning issue. Moreover, it should be noted that if a woman goes with C-section multiple times, every time the situation became more complex and dangerous than the last time. Medical field concluded that the cesarean section disturbs the perinatal and maternal health. Therefore, studies should be conducted to find out the physical or related factors that contribute to the situation requiring a cesarean section.

Algorithms of machine learning are used to produce valuable information from the dataset and learn hidden patterns and understand the structure of data to make predictions and automate the decision-making process (*An Introduction to Machine Learning* n.d.; Goodfellow, Bengio, and Courville 2016). Machine learning creates a relationship between the predictor variable and the independent variable to find hidden patterns. If the data is already labeled with the independent variables to find the hidden patterns of data, the process is called supervised learning. To classify the birth data, too many algorithms of supervised machine learning have been developed. The purpose of all these algorithms for classification is almost the same, i.e., training the algorithms with the training dataset and then evaluating their performance with the testing dataset. There are different categories of algorithms of machine learning which are used for classification. Statistical learning methods for classification (James et al. 2013), decision trees, and classification based on perceptron, Support Vectors, and Ensemble Learning methods are some well-known methods for performing classification. Ensemble modeling is a concept of machine learning in which different classifiers are combined to improve accuracy and reduce the variance in decision-making processes. Literature shows that machine learning algorithms using ensemble modeling techniques can make a strong classifier to generate better predictions and this technique has the lowest possibility of classification errors.

In the current study, we will use the birth dataset collected by (Abbas et al. 2018) from the regional level so that we can apply the machine learning ensemble modeling on the women of a particular (targeted) region. In the conducted study, the above mentioned dataset will be used to train the ensemble modeling technique of machine learning to discover hidden patterns and understand the structure of data to produce the best classification ensemble model and predict the birth outcomes. This dataset was aligned with different cleaning and reduction techniques.

The main objective of conducting this study is the following:

- (a) To enhance the accuracy of machine learning algorithms using Ensemble modeling techniques.
- (b) To reduce variance and classification error for birth data classification
- (c) To generate predictions about the birth type, i.e., cesarean section or normal delivery.

A detailed review of the background study is discussed in [section 2](#). The architecture of the methodology used in the manuscript and the description of the data are discussed in [section 3](#). [Section 4](#) presents the research and related discussions supported by different evaluation methods. The conclusion (summary) and future recommendations are discussed in [section 5](#).

Literature Review

Various databases, including Wiley, Science Direct, Google Scholar, IEEE explorer, Re-searchGate, and other data sources, were used to locate studies related to determining the mode of birth. Among the most popular search terms were “machine learning in maternity care,” “Data Mining in Healthcare to Predict Cesarean Delivery,” “c-sections using machine learning,” and “artificial intelligence in maternity care.” In addition, searching various publications, such as conference papers, available research articles, and journals, yielded the most relevant reports.

In many different areas of computing, machine learning methods are becoming more and more common. Academics are increasingly interested in using machine learning techniques to investigate prognosis and diagnosis. A few different kinds of research that are relevant to the current topic are discussed in this section.

The number of researchers involved in the classification of birth datasets using algorithms of machine learning for predicting outcomes and diagnosis is constantly growing. This section is about some research studies that are conducted by the researchers related to the current subject under observation. This section will discuss the literature review of related research. To successfully complete the research, the related research papers needed to be reviewed. More papers are studied for this research listed in [Table 1](#).

This study evaluated the use of maternal factors in creating a machine-learning model that provides the most accurate classification for unplanned C-sections and used ensemble classifiers to classify locally collected birth data. Various ensemble classifiers are used to determine

which method is best for this dataset.

The research objectives for this study are as follows:

- Construct a decision system for making precise delivery method predictions.
- To examine the outcomes of various machine learning algorithms frequently used in research papers to predict the mode of birth.
- To see how various machine learning approaches perform on locally collected birth data.

Methodology

Research has engrossed in problems to increase the accuracy of predicting birth type as either normal or C-Section. In the current study, we used an ensemble modeling voting technique for the classification of the birth dataset. Ensemble models combine individual machine learning algorithms to improve the accuracy by predicting from the combined output of the base classifiers. Gradient boosting classifier, random forest, bagging classifier, and extra trees classifier were used as base learners for making a voting ensemble model for the classification of

the birth dataset. It is observed that there is a huge difference among the women because of the region and social life.

The major focus of our research work is on:

- (1) To enhance the accuracy of machine learning algorithms using Ensemble modeling techniques.

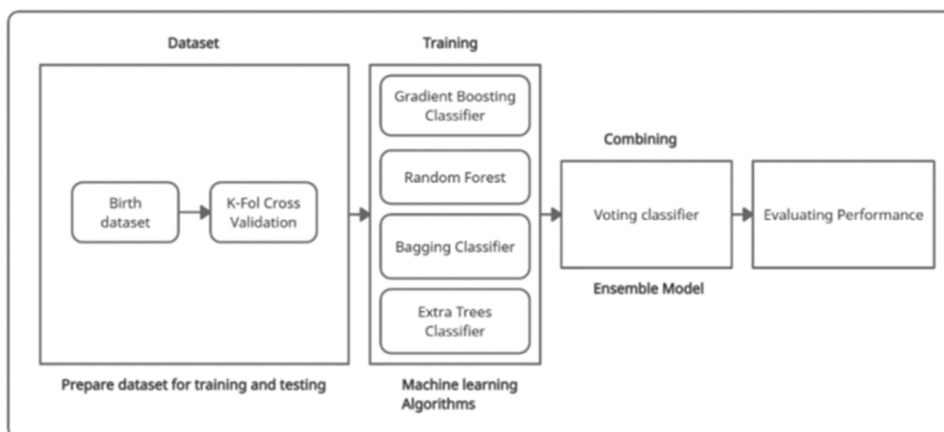


Figure 1. Architecture of methodology.

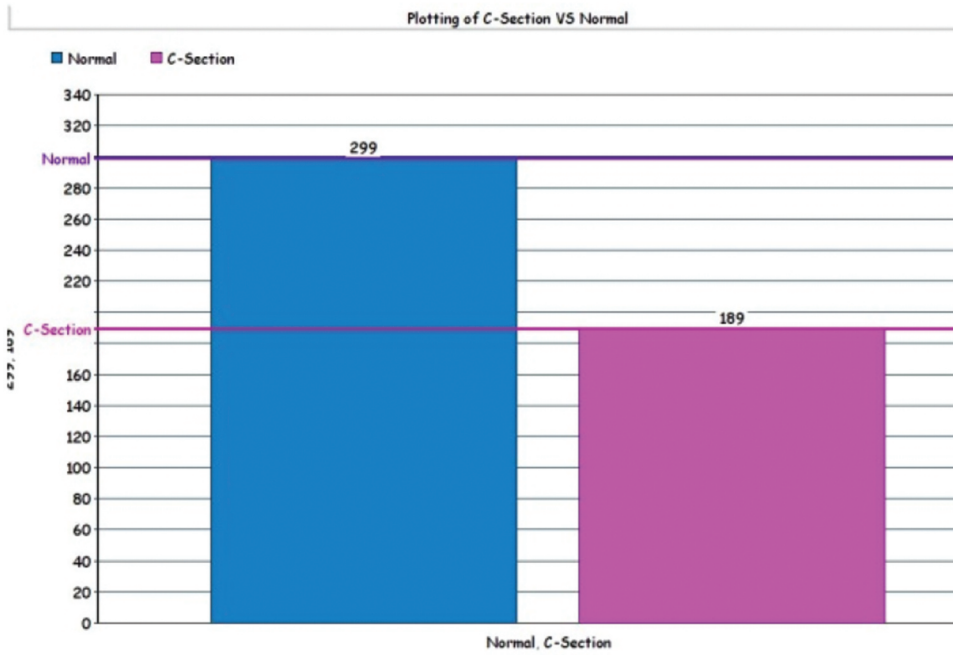


Figure 2. C-section and normal deliveries distribution.

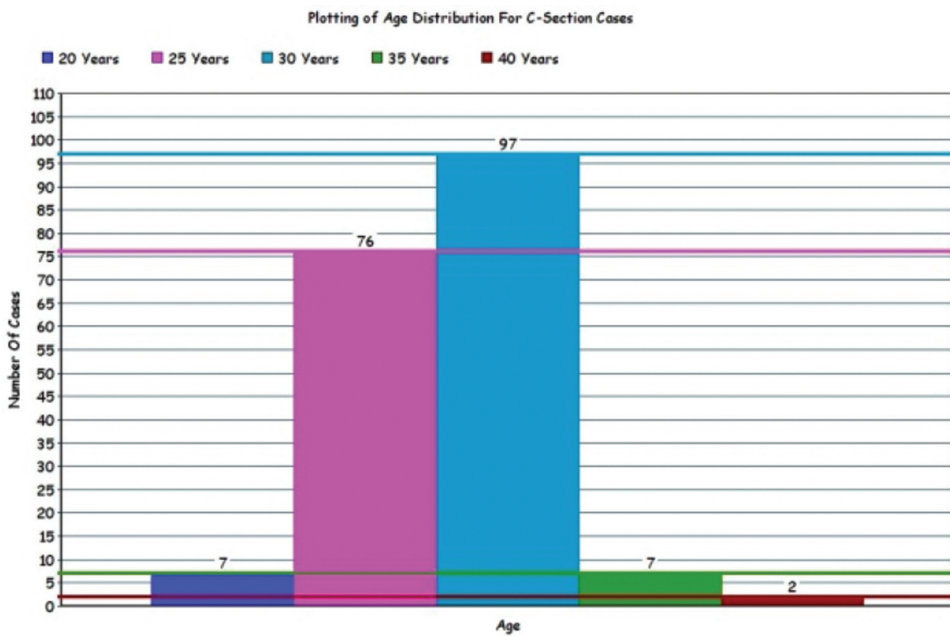


Figure 3. Age distribution for C-section cases.

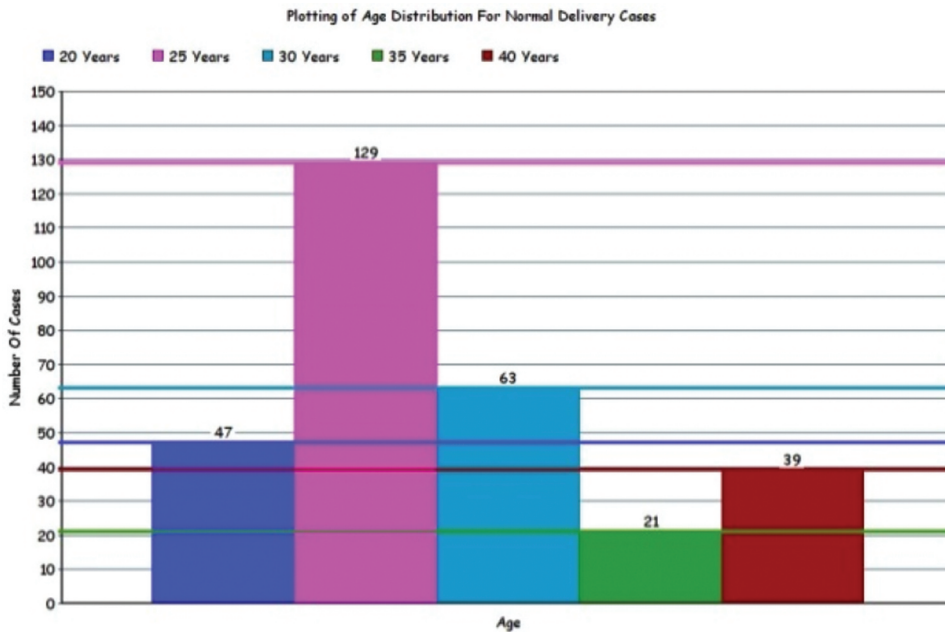


Figure 4. Age distribution for normal delivery cases.

- (2) To reduce variance and classification error in birth data classification
- (3) To generate predictions about the birth type, i.e., cesarean section or normal delivery.

The architecture of the methodology is shown in [Figure 1](#).

The proposed solution in [Figure 1](#) is a framework for making a strong classification model using machine learning ensemble modeling. In the first step, the dataset is split into a training dataset (for the training of ensemble modeling technique of machine learning algorithms) and a testing dataset (for testing the accuracy of ensemble modeling technique of machine learning algorithms) using the KFold cross-validation technique. In the next step, different algorithms of machine learning (Bagging Classifier, Random Forest (RF), Gradient Boosting Classifier, Extra Trees Classifier) are trained using a training dataset. These algorithms are selected to make an ensemble model. After selecting the machine learning algorithms, the ensemble modeling Voting technique is used to make an ensemble model of selected machine learning algorithms. The Ensemble model is then tested to check which algorithm has the highest accuracy on the birth dataset. The variables selected in the dataset are described in [Table 2](#). The C-section and normal deliveries distribution over the age is illustrated in the [Figure 2, 3 and 4](#).

**Table 1.** Literature review in the field of birth data.

Authors	Journal	Purpose	Methodology (1)	Result Predictions	Limitations & Future directions
(Abbas et al. 2020)	IEEE Access	To study and classify the birth dataset to either C-Section or Normal delivery	Different machine-learning algorithms for bagging and boosting ensembles were trained on regional birth data	BagFDA classifier provided the best accuracy of 94.53%	Study the factors related to the safety of expecting mother and child
(Abbas et al. 2020)		To perform clustering analysis on transformed and raw birth dataset	K-means and K-medoids clustering algorithms were used to make clusters of the birth dataset	Clustering algorithms perform best on transformed data rather than the raw dataset	---
(Fergus, Selvaraj, and Chalmers 2018)	Computers in biology and medicine	To evaluate the performance of different machine learning classifiers with real test data for the classification of birth data using cardiocotography	Different machine learning algorithms	The ensemble modeling technique produced the best results as compared to other machine-learning techniques	Create an extended dataset by extracting features from UC signals and remove the feature engineering stage for deep learning and stacked autoencoders
(Amin and Ali 2018)	Wavy AI Research Foundation	To propose a system for predicting decision-making during healthcare operations	Evaluated the performance of various supervised machine-learning algorithms	The k nearest neighbor classifiers (K-NN) and Random Forest (RF) are better at classifying cesarean and normal delivery types	Increase the reliability of decision-making by adding core attributes and instances
(Abbas et al. 2018)	IEEE Access	To identify the dominant medical factors which lead to the need for a cesarean section	Machine Learning Algorithms	Random Forest is the best classifier	Strengthening the women's healthcare system and childcare field
(Menai, Mohder, and Al-Mutairi 2013)	J. Med. Bioeng	To investigate the influence of feature selection on the naive Bayes algorithm's performance for fetal states and the patterns of fetal heart rate (FHR)	Tour methods of feature selection, namely: Correlation-based, Relief, Information Gain, and Mutual Information	The performance of ReliefF feature selection for the classification of fetal state was better	Collaborate with obstetric physicians to assess the computational results and investigate other feature selection methods
(Fergus et al. 2013)	PloS one	To predict the preterm births	The Electro-hysterography (Uterine Electrical Signals)	Observed improvements as compared to the existing studies of time	Evaluate different parameter adjustment settings and use advanced algorithms

Results and Discussion

This study used an ensemble modeling voting technique to classify the birth dataset. Ensemble models combine unique machine learning algorithms with improving accuracy by predicting from the combined output of the base classifiers. Gradient boosting classifier (GBC), random forest (RF), bagging classifier (BC), and extra trees classifier (ETC) were used as base learners for making a voting ensemble model for the classification of the birth dataset. The accuracy achieved by ensemble modeling is far higher than the machine learning algorithms. The current study incorporates the ensemble modeling technique for classifying birth outcomes, i.e., either cesarean section or normal delivery. Ensemble models increase the accuracy of machine learning algorithms by reducing variance and classification errors.

This section describes the detailed results attained by using different machine learning algorithms and after making an ensemble of these machine learning algorithms. R studio and python language for classification analysis are used which helps to apply different classifiers like extra trees classifier, gradient boosting classifier, random forest, bagging classifier, and voting classifier.

The results produced show that random forest has given mean accuracy of 94.26%. The gradient boosting classifier gives the results with a mean accuracy of 84.39%, extra trees classifier gives results with a mean accuracy of 94.02%. A bagging classifier using a decision tree as a base learner yields the results with a mean accuracy of 93.65%. Using Support Vector Machine, Random Forest, Extra Trees Classifier, and Bagging Classifier as base learners have given the highest mean accuracy of 94.78% in this study.

After taking the mean accuracy measure, the voting classifier has the highest mean accuracy. Following [Table 3](#) shows the overall data of normal vs C-section deliveries whereas [Table 4](#) shows the accuracy measures of all the algorithms.

Inter-classifier variation can be measured in any situation in which two or more independent classification methods are evaluating the same thing, by utilizing Kappa statistics. The calculation is based on the difference between how much agreement is present (“observed” agreement) compared to how much agreement would be expected to be present by chance alone (“expected” agreement). The interpretations of Kappa statistic values are provided in [Table 5](#). [Table 6](#) shows which classifier falls into the substantial agreement. Random Forest and Voting Classifier have marginally higher mean Kappa measure as compared to the rest and Voting Classifier has the highest mean Kappa measure.

Table 2. Specification of the dataset.

Historical Factors	During Pregnancy	Others
(1) Age	(1) Maternal Age	(1) Heartbeat
(2) Surgeries (Not C Section).	(3) Bleeding	(4) Inherited diseases
(5) Previous C Section	(6) Blood Pressure Max	(7) Breathing issues
(8) Miscarriages	(9) Blood Pressure Min	(10) Fatigue
(11) Abortions	(12) Hemoglobin	(13) Diabetes
(14) Previous Live Births	(15) Use of Folic Acid	(16) BP problem
(17) Menstruation	(18) Medication	(19) Hypertension
(20) Menstrual Cycle length	(21) Headache	(22) Iron Deficient

Table 3. Data distribution normal VS C-section.

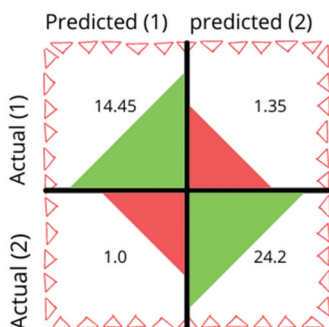
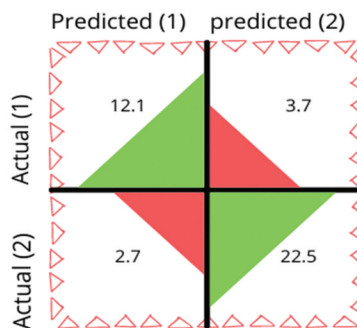
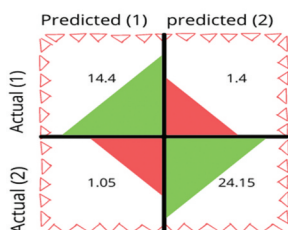
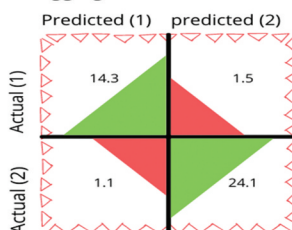
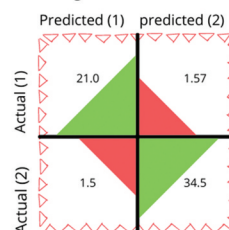
Total data	Normal Deliveries	C-Section
488	299	189

Table 4. Detail of patients w.r.t age.

Age	Frequency
17	3
18	4
19	10
20	37
21	29
22	42
23	49
24	29
25	56
26	28
27	33
28	42
29	21
30	36
31	2
32	9
33	6
34	0
35	11
36	17
37	12
38	5
39	0
40	7
Total	488

The comparison of all classifiers is based on Kappa measures. The voting classifier has the highest mean kappa value of 0.8913 from all the individual machine learning algorithms.

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.

Random Forest Confusion Matrix**Gradient Boosting Confusion Matrix****Figure 5.** Confusion matrix of random forest and gradient boosting classifier.**Extra Trees Confusion Matrix****Bagging Confusion Matrix****Voting Confusion Matrix****Figure 6.** Confusion matrix of extra trees, bagging, and voting classifier.

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

The following [Figure 5 and 6](#) present a confusion matrix based on the mean values of TP, TN, FP, and FN for all the proposed algorithms from k-fold cross-validation.

It is defined as the average of recall obtained in each class. The equation for balanced accuracy is as follows:

$$\text{Balanced accuracy} = \frac{1}{2} \left(\frac{TP}{TP + FP} + \frac{TP}{TP + FN} \right)$$

Table 5. Accuracy measures of algorithms.

Algorithms	Mean Accuracy	Mean Accuracy %age
(23) Random Forest	(24) 0.9426	(25) 94.26%
(26) Gradient Boosting Classifier	(27) 0.8439	(28) 84.39%
(29) Extra Trees Classifier	(30) 0.9402	(31) 94.02%
(32) Bagging Classifier	(33) 0.9365	(34) 93.65%
(35) Voting Classifier	(36) 0.9478	(37) 94.78%

Table 6. Kappa value interpretation.

Sr. #	Kappa Value	Agreement
(38) 1	(39) <0	(40) Less than chance agreement
(41) 2	(42) 0.01–0.20	(43) Slight agreement
(44) 3	(45) 0.21–0.40	(46) Fair agreement
(47) 4	(48) 0.41–0.60	(49) Moderate agreement
(50) 5	(51) 0.61–0.80	(52) Substantial agreement
(53) 6	(54) 0.81–0.99	(55) Almost perfect agreement

Table 7. Kappa measures of algorithms.

Algorithms	Mean Kappa	Mean Kappa %age
(56) Random Forest	(57) 0.8769	(58) 87.69%
(59) Gradient Boosting Classifier	(60) 0.6628	(61) 66.28%
(62) Extra Trees Classifier	(63) 0.8728	(64) 87.28%
(65) Bagging Classifier	(66) 0.8633	(67) 86.33%
(68) Voting Classifier	(69) 0.8913	(70) 89.13%

Table 8. Balanced accuracy, precision, and recall of algorithms.

Algorithms	Balanced Accuracy	Precision	Recall
Random Forest	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

Precision and recall are two extremely important model evaluation metrics. While precision refers to the percentage of your results that is relevant, recall refers to the percentage of total relevant results correctly classified by your algorithm.

The equation for Precision and Recall is as follows:

$$Precision = \frac{TP}{TP + FP}$$

$$Recall = \frac{TP}{TP + FN}$$

The results of mean kappa and mean kappa percentage is shown in [Table 7](#) while the results for balanced accuracy of C-section and normal deliveries, precision, and recall for the model's gradient boosting classifier, random forest, extra trees classifier, bagging classifier, and voting ensemble classifier are shown in [Table 8](#).

Conclusion and Recommendations

Ensemble modeling using machine learning techniques is developed to reduce the variance and increase the accuracy of different machine learning algorithms by combining them. The main goal of this study is to apply different machine learning techniques to increase the accuracy of birth outcomes, i.e., either normal delivery or C-Section. In this study, gradient boosting classifier (GBC), random forest (RF), extra trees classifier (ETC), bagging classifier (BC), and voting classifier (VC) algorithms are trained using cross-validation (CV) technique on death cases dataset. The results produced have shown that the voting classifier of support vector machine (SVM), random forest (RF), extra trees classifier, and bagging classifier has given the best results with the proportion of 94.78%, gradient boosting classifier has 84.39% accuracy, the random forest has 94.26% accuracy, extra trees classifier have 94.02% accuracy and bagging classifier has 93.65% accuracy. The accuracy achieved by ensemble modeling is far higher than the machine learning algorithms.

This study can be enhanced in various ways. First, the current study is intended to make an ensemble of different machine learning algorithms and check their credibility for classification purposes. In the next phases, a comprehensive comparison is intended using machine learning algorithms and ensemble model voting techniques. The dataset can also be extended to other regions of Pakistan by incorporating more medical, physical, and also involving social factors.

Conflict of interest

No potential conflict of interest was reported by the authors.

ORCID

Xian Wen Fang  <http://orcid.org/0000-0001-8531-7215>

References

- Abbas, S. A., A. Aslam, A. U. Rehman, W. A. Abbasi, S. Arif, and S. Z. H. Kazmi. 2020. K-means and K-medoids: Cluster analysis on birth data collected in City Muzaffarabad, Kashmir. *IEEE Access* 8:151847–55. doi:10.1109/ACCESS.2020.3014021.
- Abbas, S. A., A. U. Rehman, F. Majeed, A. Majid, M. S. A. Malik, Z. H. Kazmi, and S. Zafar. 2020. Performance analysis of classification algorithms on birth dataset. *IEEE Access* 8:102146–54. doi:10.1109/ACCESS.2020.2999899.
- Abbas, S. A., R. Riaz, S. Z. H. Kazmi, S. S. Rizvi, and S. J. Kwon. 2018. Cause analysis of caesarian sections and application of machine learning methods for classification of birth data. *IEEE Access* 6:67555–61. doi:10.1109/ACCESS.2018.2879115.
- Amin, M., and A. Ali. 2018. Performance evaluation of supervised machine learning classifiers for predicting healthcare operational decisions. In *Wavy AI research foundation: Lahore, Pakistan* 90.
- Fergus, P., P. Cheung, A. Hussain, D. Al-Jumeily, C. Dobbins, and S. Iram. 2013. Prediction of preterm deliveries from EHG signals using machine learning. *Plos One* 8 (10):e77154. doi:10.1371/journal.pone.0077154.
- Fergus, P., M. Selvaraj, and C. Chalmers. 2018. Machine learning ensemble modelling to classify caesarean section and vaginal delivery types using Cardiotocography traces. *Computers in Biology and Medicine* 93:7–16. doi:10.1016/j.combiomed.2017.12.002.
- Goodfellow, I., Y. Bengio, and A. Courville. 2016. Machine learning basics. *Deep Learning* 1:98–164.
- Gregory, K. D., S. Jackson, L. Korst, and M. Fridman. 2012. Cesarean versus vaginal delivery: Whose risks? Whose benefits? *American Journal of Perinatology* 29 (01):07–18. doi:10.1055/s-0031-1285829.
- Hastie, T., R. Tibshirani, and J. Friedman. 2009. Overview of supervised learning. In *The elements of statistical learning: Data mining, inference, and prediction*, 9–41. *An Introduction to Machine Learning*. (n.d.). DigitalOcean. Retrieved February 20, 2021, from <https://www.digitalocean.com/community/tutorials/an-introduction-to-machine-learning>
- James, G., D. Witten, T. Hastie, and R. Tibshirani. 2013. *An introduction to statistical learning*, vol. 112, 18. New York: Springer.
- Lumbiganon, P., M. Laopaiboon, A. M. Gülmezoglu, J. P. Souza, S. Taneepanichskul, P. Ruyan, D. E. Attygalle, N. Shrestha, R. Mori, and N. D. Hinh. 2010. Method of delivery and pregnancy outcomes in Asia: The WHO global survey on maternal and perinatal health 2007–08. *The Lancet* 375 (9713):490–99. doi:10.1016/S0140-6736(09)61870-5.
- Menai, M. E. B., F. J. Mohder, and F. Al-Mutairi. 2013. Influence of feature selection on naïve Bayes classifier for recognizing patterns in cardiotocograms. *Journal of Medical and Bioengineering* 2 (1):66–70. doi:10.12720/jomb.2.1.66-70.
- Milovic, B., and M. Milovic. 2012. Prediction and decision making in health care using data mining. *International Journal of Public Health Science (IJPHS)* 1 (2):69–78. doi:10.11591/ijphs.v1i2.1380.
- Molina, G., T. G. Weiser, S. R. Lipsitz, M. M. Esquivel, T. Uribe-Leitz, T. Azad, N. Shah, K. Semrau, W. R. Berry, and A. A. Gawande. 2015. Relationship between cesarean delivery rate and maternal and neonatal mortality. *Jama* 314 (21):2263–70. doi:10.1001/jama.2015.15553.
- Pasche Guignard, F. 2015. A gendered bun in the oven. The gender-reveal party as a new ritualization during pregnancy. *Studies in Religion/Sciences Religieuses* 44 (4):479–500. doi:10.1177/0008429815599802.