IEEE Access
Multidisciplinary : Rapid Review : Open Access Journal

Received November 26, 2020, accepted December 8, 2020, date of publication December 17, 2020, date of current version January 5, 2021.

Digital Object Identifier 10.1109/ACCESS.2020.3045469

Exploring Machine Learning Algorithms to Find the Best Features for Predicting Modes of Childbirth

MUHAMMAD NAZRUL ISLAM[1], **(Member, IEEE), TAHASIN MAHMUD**[1], **NAFIZ IMTIAZ KHAN**[1], **SUMAIYA NUHA MUSTAFINA**[1], **(Student Member, IEEE), AND A. K. M. NAJMUL ISLAM**[2]

[1]Department of Computer Science and Engineering, Military Institute of Science and Technology (MIST), Dhaka 1216, Bangladesh
[2]LUT School of Engineering Science, LUT University, 53850 Lappeenranta, Finland

Corresponding author: Muhammad Nazrul Islam (nazrul@cse.mist.ac.bd)

ABSTRACT The mode of delivery is a crucial determinant for ensuring the safety of both mother and child. The current practice for predicting the mode of delivery is generally the opinion of the physician in charge, but choosing the wrong method of delivery can cause different short-term and long-term health issues for both mother and baby. The purpose of this study was twofold: first, to reveal the possible features for determining the mode of childbirth, and second, to explore machine learning algorithms by considering the best possible features for predicting the mode of childbirth (vaginal birth, cesarean birth, emergency cesarean, vacuum extraction, or forceps delivery). An empirical study was conducted, which included a literature review, interviews, and a structured survey to explore the relevant features for predicting the mode of childbirth, while five different machine learning algorithms were explored to identify the most significant algorithm for prediction based on 6157 birth records and a minimum set of features. The research revealed 32 features that were suitable for predicting modes of childbirth and categorized the features into different groups based on their importance. Various models were developed, with stacking classification (SC) producing the highest f1 score (97.9%) and random forest (RF) performing almost as well (f1-score = 97.3%), followed by k-nearest neighbors (KNN; f1-score = 95.8%), decision tree (DT; f1-score = 93.2%), and support vector machine (SVM; f1-score = 88.6%) techniques, considering all (n = 32) features.

INDEX TERMS Machine learning, prediction, vaginal childbirth, cesarean childbirth, data mining, childbirth, modes of delivery.

I. INTRODUCTION

At the end of pregnancy, one or more babies leaves the womb through normal delivery or by cesarean section. Overall, the most common mode of childbirth is natural delivery, while others are cesarean section, emergency cesarean section, vacuum extraction, and forceps delivery [1]. All these types of births have pros and cons, and the chosen mode of delivery may not suit with the characteristics of the mother. Choosing the wrong mode of delivery may pose different risks, such as fetal termination, excessive bleeding, breathing problems for the baby, and the like [2].

The associate editor coordinating the review of this manuscript and approving it for publication was Wentao Fan.

Although natural delivery is the most common procedure, this can cause complications for symptomatic mothers, such as mothers who are older than 35 years, suffering from diseases such as diabetes or preeclampsia, or carrying more than one fetus [3]–[6]. Cesarean section, also known as c-section, or cesarean delivery, is often necessary when a normal delivery might place the baby or mother at risk [7]. It can be a lifesaving procedure for both mother and infant if, for example, a baby is in an awkward position in the womb or labor is not progressing as it should, even though cesarean section might not be the best alternative to normal delivery.

Apart from these concerns, the rates of cesarean section around the world are rapidly. According to the World Health Organization (WHO), the number of babies born through cesarean section almost doubled between

2000 and 2015—from 12% to 21% of all births [8]. The overall maternal mortality rate was 6 to 22 deaths per 100 000 live births, with approximately one third to one half of maternal deaths following cesarean delivery [9]–[11]. This situation is worse in developing and underdeveloped countries; for example, in Bangladesh, the avoidable cesarean section rate increased by 51% during 2016–2018 [12] and 77% of all cesarean sections in 2018 were medically unnecessary. In Bangladesh during 2017, the maternal mortality rate was 173 deaths per 100 000 live births, which was higher than for most developed countries; in the same year the maternal mortality rate in the United States was 17 deaths per 100 000 live births [13]. Furthermore, according to the Institute of Public Health Nutrition's *National Low Birth Weight Survey Bangladesh, 2015*, the cesarean section rate was 35.5% which exceeded the WHO's recommended range of 10–15% [14], [15]. Compared to normal delivery, maternal mortality and morbidity increase to approximately twice the rate for cesarean delivery. Cesarean section may cause various health issues for mothers, such as blood loss, injury to an organ, infection, complications in future pregnancies, and similar [16]. Sometimes, the surgical cesarean section procedure is urgently needed due to immediate concerns for the health of mother and/or baby [17], [18] and is known as emergency cesarean section. Despite its similarity to classic cesarean section, emergency cesarean section carries a higher risk of surgical injury and infection [19].

Usually, the choice of the mode of delivery is that of the medical professional in charge. A maternal healthcare application to assist doctors in predicting a mode of delivery that is compatible with the characteristics of the mother would be helpful in reducing childbirth complications. Few studies have considered the prediction of pregnancy outcomes and modes of delivery, and even fewer clinical decision-support systems have been developed in accordance with such research; for example, Pereira and colleagues [20] conducted a study to design data mining classification models to predict types of delivery using obstetric risk factors in real time. Usman and coworkers [21] conducted a study based on ultrasound technology and developed an Android application called ''Intrapartum,'' which allows medical professionals to estimate the likelihood of a normal delivery. Other studies are reviewed in the Section II.

Further research and investigation are required to explore the features that are crucial for predicting modes of delivery and to identify supporting features that can increase the accuracy of such predictions. Similarly, studies should determine which algorithms or techniques can provide greater accuracy for considering a specific group of features, since analyzing all the features of a pregnancy is rarely feasible or possible during childbirth.

The objectives of this research were therefore: (a) to explore and prioritize the features that are most essential for predicting modes of childbirth and (b) to explore machine learning techniques and develop several frameworks, based on different machine learning algorithms, for predicting modes of delivery with reasonable accuracy using a minimum number of features. To achieve these objectives, a literature review, interviews, and a mini survey were carried out to explore and prioritize the required features. Several prediction models were proposed based on decision tree (DT), k-nearest neighbors (KNN), random forest (RF), support vector machine (SVM) and stacking classification (SC) techniques to determine effective modes of delivery, and the performance of these models was evaluated in terms of precision, recall, and f1 scores.

Later sections of this article are organized as follows. The related research is briefly introduced in Section II. The methodology of the study is presented in Section III. The procedure followed to explore the important features for predicting modes of delivery are explained in Section IV. Section V briefly discusses the performance of five different prediction models, and Section VI compares the performance of the various machine learning models. The discussion and conclusion are presented in Section VII. Finally, the contributors to this study are acknowledged in section VIII.

II. LITERATURE REVIEW

This section briefly discusses studies concerning the prediction of modes of delivery, complications during childbirth, premature births, and suchlike. Pereira and colleagues [20] conducted a study to predict types of delivery (normal, cesarean, forceps, and vacuum) by identifying obstetric risk factors through data mining. In their study, different data mining techniques were applied, such as DT, generalized linear models (GLMs), support vector machines (SVMs), and naive Bayes (NB). Among these models, the most satisfactory results for statistical metrics, with the best accuracy and specificity, were achieved using DT. Some studies focused on the values of ultrasound measurements relative to the outcomes of deliveries; for example, Khazardoost and team [22] compared Bishop's scores and translabial ultrasound measurements to determine the suitability of induction of labor. Cervical length and fetal head–pubis symphysis distance were measured using translabial ultrasound. The predictive value of the Bishop's score, cervical length, and fetal head–pubis symphysis distance were determined using multivariate analysis. The results showed that translabial measurements were a more suitable method for monitoring labor progress than the Bishop's score. Ramanathan and colleagues [23] examined the likely outcomes of labor, including the risk of prolonged pregnancy and the need for cesarean section due to failed induction. The study concluded that measurement of cervical length at 37 weeks could determine the risk of an emergency cesarean section. Some studies have been carried out to predict risks during pregnancy, such as the probability of premature birth, vaginal delivery after cesarean section, and the suchlike. For example, Lipschuetz and coworkers [24] conducted a study to develop a personalized tool for predicting vaginal births after cesarean deliveries using different machine learning methods: gradient boosting, RF, balanced RF, and AdaBoost ensembles. Similarly, Tessmer-Tuck and team [25] developed a model to predict vaginal births after cesarean sections using multivariate analysis. The multivariate model using the features evaluated through stepwise regression had an area under the curve of 72.3%, while the multivariate model developed using features reported by the Grobman prediction model [26] had an area under the curve of 75.7%. In another study, Brandão and colleagues [27] aimed to evaluate the voluntary termination of pregnancy and identify the consequent risks for patients, using DT, SVMs, and GLMs to perform classification. The study achieved a sensitivity of approximately 93% and an accuracy of 68% for SVMs. Similarly, Lakshmi and coworkers [28] proposed a prediction model based on a C4.5 classification tree to determine the importance of different pregnancy attributes or features for predicting risk levels and complications during pregnancy. In another study, Goodwin and Maher [29] developed a decision support system to accurately identify mothers who were at risk of preterm birth and the attributes responsible for it. The performance of the different techniques used by the studies reviewed above is presented in Table 1.

In summary, the literature review identified certain issues. First, some studies predicted modes of delivery and the risks associated with childbirth using several machine learning techniques. Second, different sets of features were used in different studies, and different algorithms were used for the same purpose. Third, in some cases, the accuracy differed despite the number of features being the same. Fourth, most of the research focused on predicting modes of delivery, rather than focusing on the features, although it was clear that the number of features played an important role. Fifth, despite the studies carried out in this field, the researchers did not reach decisive conclusions about the best features for predicting delivery outcomes. More studies therefore need to be conducted to enable modes of delivery to be predicted in real time, with acceptable accuracy, using only necessary and minimum features. Such studies would allow mothers to give birth to their children using the safest possible childbirth procedures; thus, this work focused on predicting modes of delivery, using an optimum number of features, to assist medical professionals and pregnant mothers by reducing severe risks and complications during childbirth.

III. RESEARCH METHOD

The research was conducted in two phases. The objective of the first phase was to explore and prioritize the features necessary for predicting modes of delivery. This phase included a literature review, structured interviews, and a mini survey. The results of these methods were synthesized and analyzed to identify all possible features and their importance/priorities in predicting modes of childbirth.

The purpose of the second phase was to develop several machine learning models to effectively predict modes of delivery using an optimal number of features. In this phase, different predictive machine learning models were developed using five supervised learning algorithms: DT, SVM, KNN, RF, and SC. For applying the machine learning algorithms, an open access data set was used, which contained 6157 birth records for 2014 from four public hospitals located in three different autonomous communities of Spain [30]. The features found in Phase I were used to develop and test the models. All possible subsets of features based on their priority/importance were separately tested, and the effectiveness of each algorithm for each of the subsets was explored. The outcome of this phase revealed which algorithm best predicted the performance for a particular set of features.

IV. EXPLORING THE FEATURES

The features required to predict modes of childbirth were extracted through the literature review, structured interviews, and survey. This section briefly introduces the study procedure and the findings of each study, then presents the identified set of required features for predicting modes of childbirth.

A. LITERATURE REVIEW

1) STUDY PROCEDURE

A systematic literature review [31] was carried out to review relevant articles. Articles were searched in major scholarly databases, including Google Scholar, ScienceDirect, IEEE Xplore, ACM Digital Library, Scopus, and SpringerLink, using search strings like "machine learning and modes of delivery", "predicting mode and likelihood of cesarean/normal delivery", "risk factors associated with childbirth methods", "predicting mode of delivery using machine learning", "predicting mode of delivery using data mining", "predicting cesarean delivery", "predicting mode of delivery", "predicting normal delivery", "cesarean delivery and machine learning", and "predicting type of birth using machine learning". These search strings were entered into all the above-mentioned scholarly databases as well as the Google search engine. According to the inclusion and exclusion criteria, 13 articles, from an initial set of 324 articles, were selected for review and feature extraction. The inclusion criteria for the literature review were: papers using machine learning algorithms to predict pregnancy; articles published in conference proceedings, magazines, journals, or books; and articles written in English. Duplicate articles and short papers were excluded. Data relating to the study objectives and the selection of the features were extracted from the selected articles.

2) STUDY FINDINGS

The extracted features were synthesized and are presented in Table 2. The results showed that 8 out of 13 articles mentioned possible features for predicting modes of delivery or adverse outcomes of pregnancy. In 5 of these 13 articles, the methodologies used in the studies were mentioned, but details of all the features were not. The number of features listed in each article varied from 11 to 161. In each article, some distinct features were identified: Pereria and colleagues [20] mentioned allergies and gestation, while Yagel and coworkers [24] used previous abortions, inter-pregnancy interval,

TABLE 1. Works on the prediction of modes of delivery and associated risks.

Scope	Reference	Algorithms/ Methods	Result
Predicting mode of delivery or predicting outcomes of labor	[20]	DT	Accuracy = 72%–83.91 % Sensitivity = 66.6%-88.28% Specificity =75.99%–80.05%
		GLM	Accuracy = 58.57%–77.41% Sensitivity = 88.28%–91.1% Specificity =29.14%–67.55%
		SVM	Accuracy = 62.06% Sensitivity = 85.57% Specificity =27.92%
		NB	Accuracy = 74.69% Sensitivity = 84.29% Specificity =62.98%
	[22]	Multivariate analysis—Bishop's score	Area under the curve (AUC) = 61.8% Sensitivity = 84% Specificity = 70%
		Multivariate analysis—cervical length	(AUC) = 62.8% Sensitivity = 90% Specificity = 65%
		Multivariate analysis—fetal head–pubis symphysis distance	(AUC) = 65.6% Sensitivity = 88% Specificity = 70%
Predicting risks during childbirth, predicting premature birth, or predicting cesarean delivery	[24]	Gradient Boosting	(AUC) = 79.3%
		RF	(AUC) = 75.6%
		Balanced random forest	(AUC) = 78.2%
		AdaBoost ensemble	(AUC) = 78.4%
	[25] [26]	Multivariate logistic regression–stepwise regression	(AUC) = 78.4%
		Multivariate logistic regression–Grobman model	(AUC) = 75.7%
	[27]	DT	Accuracy = 44.6%–55.5% Sensitivity = 52.4%-92.6% Specificity =9%–52.4%
		SVM	Accuracy = 57.4%–68% Sensitivity = 60.8%–92.9% Specificity =9.3%–82.2%
		GLM	Accuracy = 43.7%–60.5% Sensitivity = 59.4% - 92.9% Specificity = 9.2%–61.4%
	[28]	C4.5 Classifier	Accuracy = 66%–71.3%
	[29]	Neural Net	(AUC) = 64%–68%
		Stepwise logistic regression	(AUC) = 66%
		Classification And Regression Trees	(AUC) = 65%
		PPV Rule	(AUC) = 67%
		FactMiner	(AUC) = 72.5% - 75.7%

inter-pregnancy interval with cesarean delivery, gestational stage at previous deliveries, gestational stage in last delivery and the likes. Some features, however, were common across many of these articles, such as age, body mass index (BMI), parity, and gestational stage [24], [32], while some very uncommon features were found in other articles, such as failed pregnancy termination, mother/sister with preeclampsia, mother/sister with gestational diabetes, Apgar score in previous deliveries, femur length, and humerus length. Moreover, different terms were used for a specific feature by different authors (e.g., cervical dilation at admission in [33] was referred to as dilation in [20]).

B. STRUCTURED INTERVIEWS

1) PARTICIPANTS PROFILES AND STUDY PROCEDURE

Using structured interviews, seven participants were interviewed to gather their opinions about possible features for predicting the mode of delivery or childbirth. All the participants were Bangladeshi medical doctors (five female and two male) with specializations in gynecology, anatomy, and

TABLE 2. Findings from literature review.

Reference	Purpose of the Study	Total Features	Example features
Pereira et al. [20]	To predict the type of delivery in real time	26	Age, allergies, whether or not delivery was programmed, motive, vigilance, gestation
Campillo-Artero el al. [30]	To predict the indicators of emergency cesarean delivery	161	Age, height, weight, obstetric risk, previous cesarean section number of previous cesarean sections, BMI, gestational stage
Khazardoost et al. [22]	To determine the pregnancy outcomes	x	Not all the features used in this study were explicitly stated
Birara and Gebrehiwot [33]	To explore features for successful vaginal birth after one cesarean section	20	Maternal age, marital status, address, education, parity, antenatal care (ANC) booking, gestational stage
Jennewein et al. [34]	To compare the outcomes of vaginal intended breech deliveries between children of different Birth-weight groups	x	Not all the features used in this study were explicitly stated
Lipschuetz et al. [24]	To predict vaginal births after cesarean deliveries	34	Maternal age, gravidity, parity, previous abortions, previous ectopic pregnancies, number of live children .
Lakshmi et al. [28]	To find the importance of parameters selected for predicting pregnancy outcomes	12	Age, pregnancy parity, history of gestational diabetes, history of preeclampsia, mother/sister with preeclampsia, mother/sister with gestational diabetes
Brandão et al. [27]	To predict risks associated with pregnancy	11	Age, number of previous voluntary terminations of pregnancy
Tessmer-Tuk et al. [25]	To predict vaginal births after cesarean sections	15	Maternal age,BMI, prior vaginal delivery, prior vaginal birth after cesarean section (VBAC), indications of prior cesarean section, ethnicity
Vankan et al. [35]	To discover the effects of unplanned cesarean sections	x	Not all the features used in this study were explicitly stated
Goodwin and Maher [29]	To predict preterm birth risk features	x	Not all the features used in this study were explicitly stated
Guann et al. [36]	To predict emergency cesarean sections	22	Maternal age, height, weight before pregnancy, BMI before pregnancy, weight at delivery
Ramanath et al. [23]	To predict pregnancy outcomes	x	Not all the features used in this study were explicitly stated
Li et al. [37]	To predict the success of vaginal births after previous cesarean deliveries	21	Maternal age, education level, pre-pregnancy BMI, gestational age, parity, history of abortion, history of vaginal delivery

surgery. The doctors had an age range of 45–60 years and each had at least 15 years of experience. The interview sessions were conducted one-on-one, and each interview lasted for 10–15 minutes. During the interviews, participants were first briefed about the purpose of the interview and ethical issues for conducting such research, and participants signed a consent form. Second, each participant was asked about their biographical profile (age, experience, expertise, and service location) and the features that they generally looked for when making a decision to determine the mode of delivery for an expectant mother. Interviews were audio recorded with the consent of the doctors and transcribed for analysis.

2) STUDY FINDINGS

The analysis of the interview data showed that 17 features were primarily considered by the medical doctors to predict modes of delivery. Of these, the most frequently stated features were height, weight, BMI, antenatal care, and gestational stage, which were mentioned by 90–100% of the participants. The results showed that each participant listed 11 to 12 features, indicating the approximate number of features doctors generally considered when predicting modes of delivery. The resulting data from the structured interviews are presented in Table 3.

C. STRUCTURED SURVEY

1) PARTICIPANTS PROFILES AND STUDY PROCEDURE

The features obtained from the literature review and structured interviews were combined, and a set of 111 features was compiled to underpin the structured survey. Participants were invited by e-mail to participate in the study. More than 100 potential participants (doctors) were invited to complete the structured survey, and 21 responded. The participants' average age was 50.71 years (Std. 4.16) and their average experience was 21 years. Thirteen of the doctors were female and eight were male. Fourteen specialized in gynecology, three in medicine, two in anatomy, and two in surgery. Based on the extracted features from the literature survey and the structured interviews, 111 questions were prepared using Google forms. Each question related to one of the identified features. Participants were requested to state whether they considered each specific feature when predicting or determining the mode of childbirth, and if so, the importance/priority of this specific feature rated on a scale of one to five, with one meaning unimportant and five representing extremely important.

2) STUDY FINDINGS

The survey results showed that none of the participants mentioned 5 of the features, including failed pregnancy termination, incomplete pregnancy termination, marital status, blood type, and fetus gender, while the other 106 features were stated by 2 to 100 percent of participants. Ten percent of the features were supported by all participants. The weighted average score, on a scale of 1 to 5, was calculated for all 111 features, based on the average score obtained for each particular feature and the number of participants who agreed on the importance of that feature. The weighted average of

TABLE 3. The summarized responses from interview.

Serial No.	Features	Doctors							Frequency
		D1	D2	D3	D4	D5	D6	D7	
1.	Age	x	✓	x	x	✓	✓	x	3
2.	Height	✓	✓	✓	✓	✓	✓	✓	7
3.	Weight	✓	✓	✓	✓	✓	✓	✓	7
4.	BMI	✓	✓	✓	✓	✓	✓	✓	7
5.	Started antenatal care	✓	✓	✓	✓	✓	✓	✓	7
6.	Weight increased during pregnancy	✓	✓	✓	✓	✓	✓	✓	7
7.	Previous cesarean	x	✓	✓	✓	✓	x	✓	5
8.	Number of previous cesareans	✓	✓	✓	✓	x	✓	✓	6
9.	Comorbidity	✓	✓	✓	✓	✓	✓	✓	7
10.	Obstetric risk	✓	✓	✓	✓	✓	✓	✓	5
11.	Previous preterm pregnancies	✓	x	x	x	✓	x	x	2
12.	Miscarriages	✓	✓	✓	✓	✓	✓	✓	7
13.	Parity	✓	✓	✓	✓	✓	✓	✓	7
14.	Gestational age	✓	✓	✓	✓	✓	✓	✓	7
15.	Cardiotocography	✓	✓	✓	✓	✓	✓	✓	7
16.	Amniotic liquid	✓	x	✓	✓	✓	✓	✓	6
17.	Amniocentesis	✓	x	✓	✓	✓	x	✓	5

the priority/importance rating showed that 11 features were rated above 4.0, 11 features were rated between 2.5 and 4.0, 10 features were rated between 1.5 and 2.5, and 79 features were rated below 1.5.

D. REVEALED FEATURES WITH THEIR PRIORITIES

To determine the optimal set of features that should be considered for predicting childbirth methods, the weighted average score from the structured survey was used to select the initial set of features, based on 111 features obtained from the literature survey and structured interviews. Thirty-two features had a weighted average score above 1.5, whereas 79 features had a weighted average score below 1.5. By fixing a threshold value of 1.5, 32 features were selected as the initial set of features, and the other 79 features were considered as much less important. To obtain the optimum subset of features for prediction, the univariate feature selection (UFE) method was applied to the initial set of 32 features to determine the priority of the features based on their correlation with the target feature (type of birth). UFE inspects each feature individually to discover its relationship with the dependent variable [38]. In this study, a chi-squared test [39] was performed to obtain the ranked list of features, classifying the features into high-, medium-, and low-importance groups. There are many other state-of-the-art feature selection methods, such as mutual information-based feature selection [40] and LDA-based dimensionality reduction [41], which could also have been used to categorize the features. The classification of the revealed features into different groups of importance is shown in Table 6.

In this study, we excluded the least important features from consideration; hence, by taking features in different combinations from the three highest importance groups (high, medium, and low), seven different classes were formed, which are presented in Table 4. To form classes numbered one to seven, different sets of features from the high-, medium-, and low-priority groups were chosen. Classes one, two, and three were formed by taking all the features from the high-, medium-, and low-importance groups, respectively; however, to form classes four to seven, multiple feature groups were considered (e.g., class four was formed by taking all the features from the high- and medium-importance groups). High- and low-importance features were selected for class five, while medium- and low-importance features were included in class six. Last, by combining all the features, class seven was formed.

TABLE 4. Class formation.

Class No.	Feature group/groups considered			Total No. of features
	High	Medium	Low	
1.	✓	x	x	11
2.	x	✓	x	11
3.	x	x	✓	10
4.	✓	✓	x	22
5.	✓	x	✓	21
6.	x	✓	✓	21
7.	✓	✓	✓	32

V. ANALYZING ML ALGORITHMS

One of the main objectives of analyzing different machine learning algorithms was to find the optimum subset of features necessary for predicting appropriate modes of delivery. To do this, we carried out thorough data collection, data synthesis, and development of prediction models. These phases are briefly described as follows:

1) Data Collection: To develop efficient prediction models, a data set[1] consisting of diverse sets of features relating to pregnant women was used, as collected by Campillo-Artero and colleagues [30]. The data set contained 6157 singleton birth records, with no exclusions, which occurred in 2014 at four public hospitals located in three different autonomous Spanish communities. A total of 161 features were included in the dataset—142 categorical and 19 numerical.
2) Data Synthesis: In the second phase, the collected data were synthesized. First, irrelevant features were removed from the data set, and 32 important features out of 153 features were selected for further analysis based on a feature exploration study, which was briefly discussed in the previous section. Second, duplicate instances were removed from the data set (18 out of 6157 instances). Third, null values were handled by applying likewise replacement techniques, where a particular numerical observation was replaced with a mean value and categorical observation was replaced with the most frequent value. Fourth, an adaptive synthetic (ADASYN) sampling approach was applied to the data to make the class distribution of the target feature uniform, since the data set had a class imbalance problem (e.g., for the type of birth feature, most of the class consisted of natural births).
3) Development of prediction models: In this phase, different machine learning algorithms were chosen based on recent work relating to the prediction of childbirth methods [20], [23] [27], [30], which included DT, KNN, RF, and SVM, and these four different algorithms were then combined to develop a SC model. The models were developed using scikit-learn [42], which is a Python module integrated into a wide range of machine learning algorithms. To evaluate the performance of each prediction model, a random train/test split was applied to the data set, with 80% of the data considered as training data and 20% of the data considered as testing data. To build each of the models, the default hyperparameters of the scikit-learn library (the parameter values of which are used to control the learning process) were employed. The performance of each prediction model was measured in terms of its precision, recall, and f1 scores, which are shown in Appendix A. Each of the evaluation parameters was obtained by performing macro averaging on the actual and model-predicted class labels, which calculated parameters for each class label and found their unweighted means. Since the resampling method had been utilized to balance the classes, accuracy was not considered a performance metric for evaluating the performance of the classifiers as the literature had shown that accuracy was not an appropriate metric to use in such a case. The results are briefly analyzed in the following subsections.

[1] https://figshare.com/articles/dataset/Predictive_modeling_of_emergency_cesarean_delivery/5814399

A. DECISION TREE

DT is a non-parametric supervised learning method used for classification and regression [43]; however, for this classification problem, DT was applied to predefined classes of features, and the f1 scores for the DTs for all classes are plotted in Figure 3 along with the other four algorithms. Table 5 indicates that class five, which consisted of 21 features, had precision, recall, and f1 scores of 92.3%, 92.4%, and 92.3%, respectively. It can be seen that satisfactory values for the precision, recall, and f1 scores were achieved for classes four to seven. From Figure 3 and Table 5, it is clear that class seven, consisting of all 32 features, had the highest f1 score (93.2%) and, when applying the DT algorithm, the precision and recall were also 93.2%. Combining all 11 high-importance features gave precision, recall, and f1 scores of 82.9%, 82.6%, and 82.8%, respectively. Lastly, it was evident that the high-importance feature class performed better than the medium-importance feature class, whereas the medium-importance feature class performed better than the low-importance feature class.

B. RANDOM FOREST

RF is an ensemble learning method for classification, regression, and other tasks that operate by constructing a multitude of decision trees at training time and outputting the class that is the mode of the classes (classification) or mean prediction (regression) of the individual trees [44]. The results for running the RF algorithm on predefined classes of features are shown in Figure 3 and Table 5. Among the high, medium, and low-importance feature classes, the high-importance class had the highest (83.9%) f1 score and the medium-importance class performed better than the low-importance class. Nonetheless, the f1 scores and other performance measures increased from class four onwards except for class six, whereas the best performance measure was achieved for class seven, consisting of all 32 features with precision, recall, and f1 scores of 97.3%.

C. K-NEAREST NEIGHBORS

KNN is a non-parametric, lazy learning algorithm that uses a database in which the data points are separated into several classes to predict the classification of a new sample point [45]. The findings observed using the KNN algorithm to build prediction models based on predefined classes of features are shown in Figure 3 and Table 5. It is clear that the high-importance class had a higher f1 score than the medium-importance class, which in turn had a higher f1 score than the low-importance class. The performance of the other classes increased gradually from class four to class seven; however, the best performance when applying KNN was observed for class seven, with precision, recall, and f1 scores of 95.9%, 96%, and 95.8%, respectively.

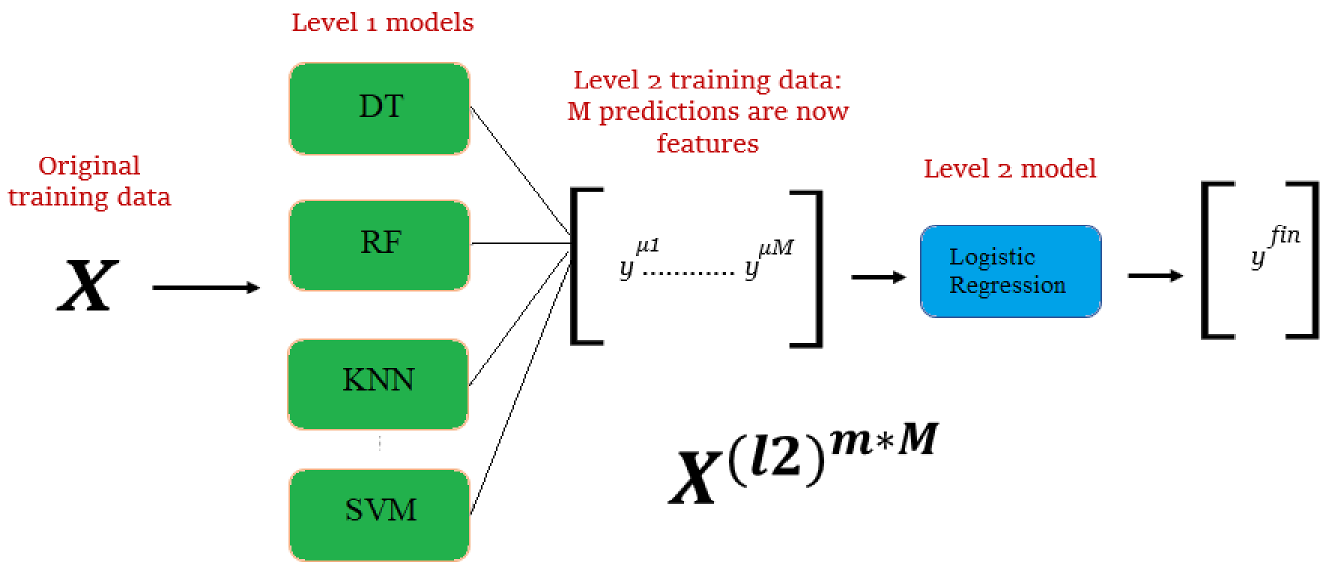

FIGURE 1. Architecture of the proposed stacking classifier.

D. SUPPORT VECTOR MACHINE

SVM implementation is unique compared to other machine learning algorithms with the capability to handle multiple continuous and categorical variables [46]. SVM was considered in the analysis, and graphs depicting the performance of the algorithm are shown in Figure 3 and Table 5. In relation to SVM, it can be seen that performance had a positive relationship with the importance level (i.e., the high-, medium- and low-importance classes had high, medium, and low performance scores respectively). The performance of SVM increased linearly from class four to class seven, whereas class seven had the best precision, recall, and f1 scores; however, SVM had relatively low performance compared to the other algorithms for all seven classes of features. The highest f1 score (88.6%) was observed for class seven, which consisted of 32 features, whereas an 85.1% f1 score was achieved for class six, with 21 features, and an 83.4% f1 score for class five, with the same number of features. It can therefore be said that class seven had the best result for SVM.

E. STACKING CLASSIFIER

Stacking is an ensemble learning technique in which the best predictions of multiple classifiers are combined to generate a new training set for a meta-classifier. The individual base classifiers are trained based on the complete training set, and the meta-classifier is fitted based on the output of individual classification models. The base classifier can be considered a level 1 model and the meta-classifier a level 2 model/models. The architecture of the proposed stacking classifier is shown in Figure 1. The initial training data (X) had m observations (m = 6157) and n features (n = 2). M different models (M = 4) were trained on X, and each of the model's prediction outcomes (y) was combined to generate data set X^{l2} for the level 2 model. In this study, a strong SC was proposed in which DT, KNN, RF, and SVM were the base classifiers and logistic regression was the meta-classifier. The machine learning models developed with SC showed the best performance of the five algorithms. For class one with 11 important features, satisfactory results were observed with an F1 score of 84.8%. The highest (97.9%) f1 score was observed when considering all 32 features, while for feature classes four, five, and six, almost identical performance (f1 score = 95 97%) was observed.

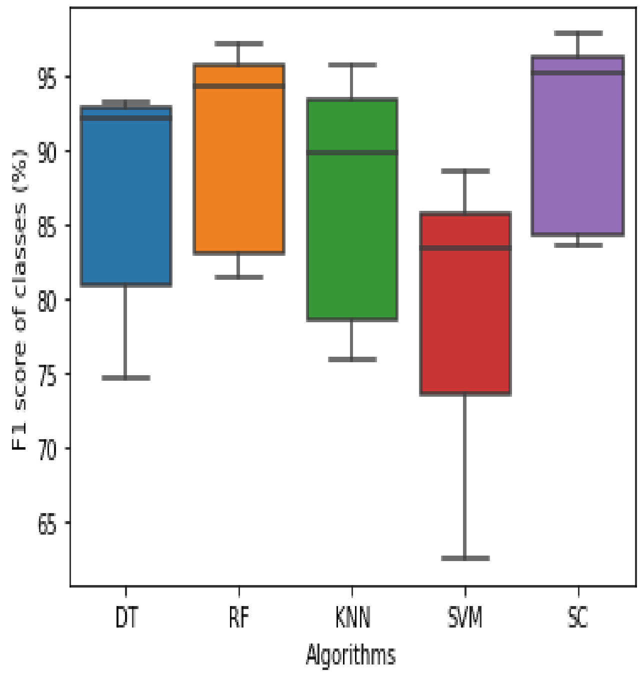

FIGURE 2. F1 scores for the proposed algorithms.

VI. COMPARISON BETWEEN ALGORITHMS

The f1 scores obtained by running different algorithms on specified classes of features were combined and are shown

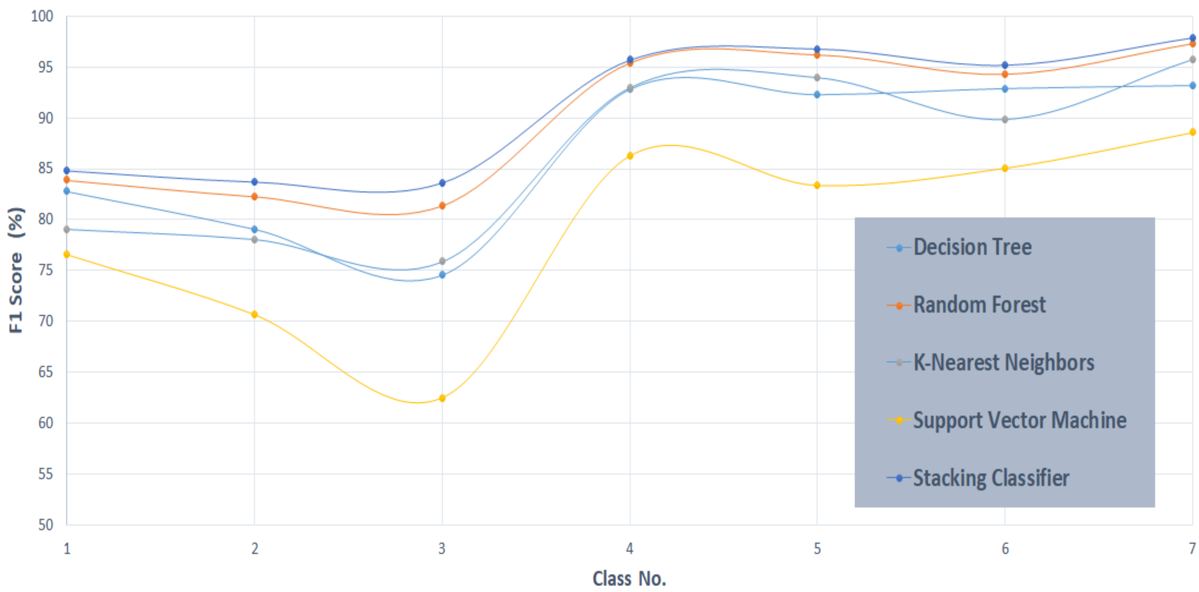

FIGURE 3. Performance (f1 score) of the selected algorithms on predefined classes of features.

in Figure 3. The dispersion in f1 scores for all the proposed models, using different feature classes, are shown in a boxplot in Figure 2. The boxes on the figures show the interquartile ranges with the median values shown as horizontal lines in the middle. The upper and lower whiskers represent 1.5 times the interquartile range exceeding the first and third quartiles, respectively. It is evident that SVM had greater dispersion than the other algorithms. From the performance graph of the different algorithms, the importance of the different feature sets can be seen.

A. HIGHEST F1 SCORES FOR DIFFERENT ALGORITHMS

In our study, five machine learning algorithms were considered for predicting appropriate childbirth methods. Applying each of the algorithms led to different results for the same class of features, although the best performance for each algorithm was observed when considering all 32 features. When applying DT, the highest f1 score (93.2%) was achieved for class seven, consisting of 32 features. Similarly, for SVM, the highest (88.6%) f1 score was observed for class seven; however, the RF algorithm gave the highest (97.3%) f1 score for the same feature class. For KNN, the highest f1 score (95.8%), and finally, for SC, the highest f1 score (97.9%) were observed for feature class seven. It can therefore be said that, in our study, the highest f1 score was obtained by applying SC and considering all the features of our study.

B. HIGHEST F1 SCOREs FOR HIGH-IMPORTANCE FEATURES

In this study, 11 features were considered to be the most important (class one) features based on the feature exploration study. F1 scores of 82.8%, 83.9%, 79.1%, 76.6%, and 84.8% were achieved by applying DT, RF, KNN, SVM, and SC, respectively. The highest precision (83.6%) and f1 (84.8%) scores were achieved by applying SC to the high-importance features.

C. HIGHEST F1 SCORES FOR MEDIUM-IMPORTANCE FEATURES

Features that were more important than the low-importance features and less important than high-importance features were considered to be medium-importance features (class two) in the feature exploration feature section, consisting of 11 features. When applying DT, RF, KNN, SVM, and SC to the medium-importance features, f1 scores of 79.1%, 82.3%, 78.1%, 70.7%, and 83.7% were achieved respectively. However, the highest f1-score (83.7%) was achieved by applying the SC algorithm to the medium-importance features.

D. HIGHEST F1 SCORE FOR LOW-IMPORTANCE FEATURES

In this study, 10 features were considered to be low-importance features (class three). F1-score of 74.6%, 81.4%, 75.9%, 62.5%, and 83.6% were achieved by applying DT, RF, KNN, SVM, and SC algorithms, respectively. Also, for class three, the highest f1 score (83.6%) was achieved by applying the SC algorithm, while RF showed similar results (f1 score = 81.4%).

E. MINIMUM NUMBER OF FEATURES WITH THE HIGHEST F1 SCORES

For all the algorithms, class five, which consisted of 21 features (including high- and low-importance features) displayed almost the same accuracy as class seven, which consisted of all 32 features; hence, we concluded that class five was the optimum set of features compared to the other classes of features.

TABLE 5. Results obtained from the selected algorithms.

Class No.	Total Features	Groups Considered			Algorithm	Performance (%)		
		High	Medium	Low		Precision	Recall	F1-score
1	11	✓	x	x	DT	82.9	82.6	82.8
					RF	83	83.8	83.9
					KNN	79.5	79	79.1
					SVM	77.2	76.9	76.6
					SC	83.6	84.9	84.8
2	11	x	✓	x	DT	79.1	79.3	79.1
					RF	82.2	82.4	82.3
					KNN	78	78.4	78.1
					SVM	70.7	70.5	70.7
					SC	83.9	83.6	83.7
3	10	x	x	✓	DT	74.4	74.8	74.6
					RF	81.4	81.7	81.4
					KNN	75.7	76.3	75.9
					SVM	63.5	64	62.5
					SC	83.6	83.6	83.6
4	22	✓	✓	x	DT	92.7	92.8	92.8
					RF	95.4	95.5	95.4
					KNN	93	93.1	93
					SVM	86.5	86.4	86.3
					SC	95.7	95.8	95.7
5	21	✓	x	✓	DT	92.3	92.4	92.3
					RF	96.2	96.3	96.2
					KNN	94	94.1	94
					SVM	83.6	83.4	83.4
					SC	96.8	96.8	96.8
6	21	x	✓	✓	DT	92.9	92.9	92.9
					RF	94.3	94.3	94.3
					KNN	90	90.1	89.9
					SVM	85.5	86.1	85.1
					SC	95.2	95.2	95.2
7	32	✓	✓	✓	DT	93.2	93.2	93.2
					RF	97.3	97.3	97.3
					KNN	95.9	96	95.8
					SVM	88.6	88.6	88.6
					SC	97.9	97.9	97.9

VII. DISCUSSION AND CONCLUSION

A. PRINCIPAL RESULTS

This research followed three clearly-defined stages. First, a set of features for predicting childbirth methods was collected through a literature review and by interviewing concerned doctors. Recent works relating to this topic were reviewed and seven doctors were interviewed. Second, a structured survey was carried out with the participation of 21 doctors to obtain a weighted average score for each collected feature. An initial set of 32 features was selected based on their weighted average scores, and other features were classified as much less important. The initial set of features was then divided into high-, medium-, and low-importance groups based on a chi-squared test. Apart from the least important group, seven different classes were formed by considering all possible combinations of the groups of features. Third, the performance of different machine learning algorithms was analyzed according to the predefined classes to determine the best subset of features for each algorithm. This study showed that the highest (97.9%) f1 score was obtained using an SC that considered 32 features, formed by taking all the features from the high-, medium-, and low-importance groups. The best prediction model in this research had the highest performance of all the reviewed papers.

TABLE 6. Classifying the revealed features in different importance groups.

Importance Level	Features
High	Previous cesarean? (y/n), complications, Robson group, assisted reproductive technology (art) mode, previous preterm pregnancies, amniocentesis, pre-induction, induction, episiotomy, oxytocin, fetal intrapartum pH
Medium	Parity, obstetric risk, comorbidity, number of previous cesarean sections, weight increased during pregnancy, start week of antenatal care, art, previous term pregnancies, amniotic liquid, number of miscarriages in the past, anesthesia
Low	Gestational stage, height, weight, BMI, age, cardiotocography, maternal education, substance abuse, smoking, alcohol
Very Low	Allergies, motive, vigilance, gestation (singular or multiple pregnancy), blood pressure, marital status, blood type, RFC, streptococcus, rhesus (rh) factor, fetus weight, dilatation/cervical dilatation, consistency of pregnancy discharge, extinction, fetal position, Bishop's score, Hodge plan, antenatal care (ANC) booking, presence of comorbid medical illness, indication of past cesarean section, inter-delivery interval, successful vaginal birth after one cesarean section, live newborn, past spontaneous vaginal delivery, membrane status at admission, presence of meconium, duration of labor, gravidity, previous abortions, previous ectopic pregnancies, inter-pregnancy interval, inter-pregnancy interval with cesarean delivery, previous delivery weights, head circumferences in previous deliveries, gestational stage for previous deliveries, gestational age in last delivery, maternal age at previous deliveries, modes of previous deliveries, vaginal birth after cesarean in the past, 5-minute Apgar score in previous deliveries, of diagnosis of dysfunctional labor in previous deliveries, Group B Strep (GBS) status, gestation diabetes status, onset of labor, first cervical examination—effacement, first cervical examination—head station, history of preeclampsia, mother/sister with preeclampsia, mother/sister with gestational diabetes, present trimester of pregnancy, number of previous voluntary interruptions of pregnancy, professional status, achieved pregnancy termination(PT), failed PT, incomplete PT, if patient attended the consultant, contraceptive method, ethnicity, Apgar scores, fetal monitoring assessment, prior uterine layer closure, characteristics of amniotic fluid, umbilical cord status, hypertensive disorders of pregnancy, low lying placenta, placental abruption, oligohydramnios, hydramnion, macrosomia, history of vaginal delivery, recurrence of previous cesarean indications, augmentation, analgesic administration, rupture of membranes, bi-parietal diameter, abdominal circumference, femur length, humerus length, fetus gender

B. LIMITATIONS

In this study, the numbers of participants in the mini survey and interviews were inadequate; future work should focus interviews and surveys with larger numbers of participants. For building prediction models, traditional machine learning algorithms were used, which may not offer the best performance for all kinds of diverse data sets. In the future, DNN-based learning models could be constructed for better prediction of childbirth methods. An application could be developed in the future to support medical professionals in predicting childbirth methods and making appropriate decisions.

C. COMPARISON WITH PRIOR WORK

Some research related to pregnancy has been carried out in the past, as briefly described in the literature review section. Most previous works focused on determining pregnancy outcomes based on certain factors; for example, Usman and colleagues [21] and Khazardoost and coworkers [22] predicted the outcomes of deliveries using ultrasound measurements. Similarly, Pereira and team [20] predicted the mode of delivery using obstetric risk factors. Different risks associated with pregnancy were predicted by Brandão and colleagues [27] and Goodwin and Maher [29]. Some studies were conducted to predict vaginal deliveries after cesarean sections; for example, Lipschuetz and coworkers [24] and Tessmer-Tuck and team [25]. Campillo-Artero [30] predicted the indicators of emergency cesarean sections and identified the factors that placed mothers at risk of emergency cesarean section. Our work was more generalized than the other studies for predicting modes of childbirth. In this study, to predict modes of delivery, necessary features were first identified through a feature exploration study, and the features were then classified into different importance groups based on a feature selection method. Five different machine learning algorithms were applied to different subsets of these explored features with the aim of identifying the effective features for determining modes of delivery.

D. CONCLUSION

Choosing the most suitable modes of childbirth is vital for the safety of both mothers and infants, but the best sets of features to consider when taking such decisions remained to be explored. This study therefore examined all possible features and classified them into different categories by conducting an extensive empirical study, then applied a feature selection method of machine learning. Later, the results of applying five machine-learning algorithms to combinations of these categories (classes) were used to determine the most appropriate algorithm for predicting the best childbirth model with the minimum number of features. The performance for different classes of features showed the efficacy of the approach followed in this study.

APPENDIX A
ALGORITHM RESULTS
See Table 5.

APPENDIX B
FEATURE GROUPS
See Table 6.

ACKNOWLEDGMENT
The authors would like to thank the participants of the structured survey whose online participation made the study possible. Their efforts are gratefully acknowledged.

REFERENCES

[1] M. Arora. *6 Different Types of Delivery Methods You Must Know*. Accessed: Jul. 11, 2020. [Online]. Available: https://parenting.firstcry.com/articles/different-childbirth-methods-you-must-know

[2] I. Hendler, M. Kirshenbaum, M. Barg, S. Kees, S. Mazaki-Tovi, O. Moran, A. Kalter, and E. Schiff, "Choosing between bad, worse and worst: What is the preferred mode of delivery for failure of the second stage of labor?" *J. Maternal-Fetal Neonatal Med.*, vol. 30, no. 15, pp. 1861–1864, Aug. 2017.

[3] K. Gregory, S. Jackson, L. Korst, and M. Fridman, "Cesarean versus vaginal delivery: Whose risks? Whose benefits?" *Amer. J. Perinatol.*, vol. 29, no. 1, pp. 07–18, Jan. 2012.

[4] S. Buckley, *Gentle Birth, Gentle Mothering: A Doctor's Guide to Natural Childbirth and Gentle Early Parenting Choices*. San Francisco, CA, USA: Celestial Arts, 2013.

[5] I. M. Gaskin, *Ina May's Guide to Childbirth*. New York, NY, USA: Bantam, 2003.

[6] S. McCutcheon, *Natural Childbirth the Bradley Way: Revised Edition*. Baltimore, MD, USA: Penguin, 2017.

[7] E. L. Ryding, "Investigation of 33 women who demanded a cesarean section for personal reasons," *Acta Obstetricia et Gynecol. Scandinavica*, vol. 72, no. 4, pp. 280–285, Jan. 1993.

[8] T. Boerma, C. Ronsmans, D. Y. Melesse, A. J. D. Barros, F. C. Barros, L. Juan, A.-B. Moller, L. Say, A. R. Hosseinpoor, M. Yi, D. de Lyra Rabello Neto, and M. Temmerman, "Global epidemiology of use of and disparities in caesarean sections," *Lancet*, vol. 392, no. 10155, pp. 1341–1348, Oct. 2018.

[9] V. M. Allen, C. M. O'Connell, R. M. Liston, and T. F. Baskett, "Maternal morbidity associated with cesarean delivery without labor compared with spontaneous onset of labor at term," *Obstetrics Gynecology*, vol. 102, no. 3, pp. 477–482, Sep. 2003.

[10] S. Liu, R. M. Liston, K. S. Joseph, M. Heaman, R. Sauve, and M. S. Kramer, "Maternal mortality and severe morbidity associated with low-risk planned cesarean delivery versus planned vaginal delivery at term," *Cmaj*, vol. 176, no. 4, pp. 455–460, Feb. 2007.

[11] R. M. Silver, M. B. Landon, D. J. Rouse, K. J. Leveno, C. Y. Spong, E. A. Thom, A. H. Moawad, S. N. Caritis, M. Harper, R. J. Wapner, Y. Sorokin, M. Miodovnik, M. Carpenter, A. M. Peaceman, M. J. O'Sullivan, B. Sibai, O. Langer, J. M. Thorp, S. M. Ramin, and B. M. Mercer, "Maternal morbidity associated with multiple repeat cesarean deliveries," *Obstetrics Gynecol.*, vol. 107, no. 6, pp. 1226–1232, Jun. 2006.

[12] *BANGLADESH: 51 Per Cent Increase In 'Unnecessary' C-Sections In Two Years | Save The Children International*. Accessed: Jul. 11, 2020. [Online]. Available: https://www.savethechildren.net/news/bangladesh-51-cent-increase-%E2%80%9Cunnecessary%E2%80%9D-c-sections-two-years

[13] *Maternal Mortality Ratio (Modeled Estimate, Per 100, 000 Live Births)—Bangladesh*. Accessed: Jul. 11, 2020. [Online]. Available: https://data.worldbank.org

[14] A. P. Betrán, M. R. Torloni, J.-J. Zhang, A. M. Gülmezoglu, H. A. Aleem, F. Althabe, T. Bergholt, L. de Bernis, G. Carroli, and C. Deneux-Tharaux, "Who statement on caesarean section rates," *BJOG An, Int. J. Obstetrics Gynaecol.*, vol. 123, no. 5, pp. 667–670, 2016.

[15] *Number of Caesarean Births on Rise*. Accessed: Jul. 11, 2020. [Online]. Available: https://www.newagebd.net/article/24660/number-of-caesarean-births-on-rise

[16] S. M. Koroukian, "Relative risk of postpartum complications in the Ohio Medicaid population: Vaginal versus cesarean delivery," *Med. Care Res. Rev.*, vol. 61, no. 2, pp. 203–224, Jun. 2004.

[17] T. D. Collard, H. Diallo, A. Habinsky, C. Hentschell, and T. M. Vezeau, "Elective cesarean section: Why women choose it and what nurses need to know," *Nursing Women's Health*, vol. 12, no. 6, pp. 480–488, Dec. 2008.

[18] P. L. Van Horn, "What you should know about Cesarean section," *Female Patient*, vol. 25, no. 3, p. 47, 2000.

[19] R. Cnattingius, B. Höglund, and H. Kieler, "Emergency cesarean delivery in induction of labor: An evaluation of risk factors," *Acta Obstetricia et Gynecologica Scandinavica*, vol. 84, no. 5, pp. 456–462, May 2005.

[20] S. Pereira, F. Portela, M. F. Santos, J. Machado, and A. Abelha, "Predicting type of delivery by identification of obstetric risk factors through data mining," *Procedia Comput. Sci.*, vol. 64, pp. 601–609, 2015.

[21] S. Usman, B. H. Kahrs, C. Wilhelm-Benartzi, W. A. Hassan, H. Barton, K. A. Salvesen, T. M. Eggebø, and C. Lees, "Prediction of mode of delivery using the first ultrasound-based 'intrapartum app,'" *Amer. J. Obstetrics Gynecol.*, vol. 221, no. 2, pp. 163–166, 2019.

[22] S. Khazardoost, F. Ghotbizadeh Vahdani, S. Latifi, S. Borna, M. Tahani, M. A. Rezaei, and M. Shafaat, "Pre-induction translabial ultrasound measurements in predicting mode of delivery compared to bishop score: A cross-sectional study," *BMC Pregnancy Childbirth*, vol. 16, no. 1, p. 330, Dec. 2016.

[23] G. Ramanathan, C. Yu, E. Osei, and K. H. Nicolaides, "Ultrasound examination at 37 weeks' gestation in the prediction of pregnancy outcome: The value of cervical assessment," *Ultrasound Obstetrics Gynecol.*, vol. 22, no. 6, pp. 598–603, Dec. 2003.

[24] M. Lipschuetz, J. Guedalia, A. Rottenstreich, M. N. Persky, S. M. Cohen, D. Kabiri, G. Levin, S. Yagel, R. Unger, and Y. Sompolinsky, "Prediction of vaginal birth after cesarean deliveries using machine learning," *Amer. J. Obstetrics Gynecol.*, vol. 222, no. 6, pp. 613.e1–613.e12, Jun. 2020.

[25] J. A. Tessmer-Tuck, S. A. El-Nashar, A. R. Racek, C. M. Lohse, A. O. Famuyide, and M. J. Wick, "Predicting vaginal birth after cesarean section: A cohort study," *Gynecol. Obstetric Invest.*, vol. 77, no. 2, pp. 121–126, 2014.

[26] W. A. Grobman, Y. Lai, M. B. Landon, C. Y. Spong, K. J. Leveno, D. J. Rouse, M. W. Varner, A. H. Moawad, S. N. Caritis, M. Harper, and R. J. Wapner, "Development of a nomogram for prediction of vaginal birth after cesarean delivery," *Obstetrics Gynecol.*, vol. 109, no. 4, pp. 806–812, 2007.

[27] J. M. Machado, A. Abelha, M. Santos, F. Portela, E. Pereira, and A. Brandão, "Predicting the risk associated to pregnancy using data mining," in *Proc. Int. Conf. Agents Artif. Intell.*, vol. 2. Setúbal, Portugal: SCITEPRESS, Jan. 2015, pp. 594–601.

[28] B. N. Lakshmi, T. S. Indumathi, and N. Ravi, "A Study on C. 5 decision tree classification algorithm for risk predictions during pregnancy," *Procedia Technol.*, vol. 24, pp. 1542–1549, Jan. 2016.

[29] L. Goodwin and S. Maher, "Data mining for preterm birth prediction," in *Proc. ACM Symp. Appl. Comput. (SAC)*, vol. 1, 2000, pp. 46–51.

[30] C. Campillo-Artero, M. Serra-Burriel, and A. Calvo-Pérez, "Predictive modeling of emergency cesarean delivery," *PLoS ONE*, vol. 13, no. 1, Jan. 2018, Art. no. e0191248.

[31] B. Kitchenham, O. Pearl Brereton, D. Budgen, M. Turner, J. Bailey, and S. Linkman, "Systematic literature reviews in software engineering—A systematic literature review," *Inf. Softw. Technol.*, vol. 51, no. 1, pp. 7–15, Jan. 2009.

[32] Y.-X. Li, Z. Bai, D.-J. Long, H.-B. Wang, Y.-F. Wu, K. H. Reilly, S.-R. Huang, and Y.-J. Ji, "Predicting the success of vaginal birth after caesarean delivery: A retrospective cohort study in China," *BMJ Open*, vol. 9, no. 5, May 2019, Art. no. e027807.

[33] M. Birara and Y. Gebrehiwot, "Factors associated with success of vaginal birth after one caesarean section (VBAC) at three teaching hospitals in Addis Ababa, Ethiopia: A case control study," *BMC Pregnancy Childbirth*, vol. 13, no. 1, p. 31, Dec. 2013.

[34] L. Jennewein, U. Kielland-Kaisen, B. Paul, C. J. Möllmann, A.-S. Klemt, S. Schulze, N. Bock, W. Schaarschmidt, D. Brüggmann, and F. Louwen, ''Maternal and neonatal outcome after vaginal breech delivery at term of children weighing more or less than 3.8 kg: A FRABAT prospective cohort study,'' *PLoS ONE*, vol. 13, no. 8, Aug. 2018, Art. no. e0202760.
[35] E. Vankan, E. Schoorel, S. van Kuijk, J. Nijhuis, R. Hermens, and H. Scheepers, ''The effect of the use of a decision aid with individual risk estimation on the mode of delivery after a caesarean section: A prospective cohort study,'' *PLoS ONE*, vol. 14, no. 9, Sep. 2019, Art. no. e0222499.
[36] P. Guan, F. Tang, G. Sun, and W. Ren, ''Prediction of emergency cesarean section by measurable maternal and fetal characteristics,'' *J. Investigative Med.*, vol. 68, no. 3, pp. 799–806, Mar. 2020.
[37] W.-Y. Li, T. Liabsuetrakul, and B. Stray-Pedersen, ''Effect of mode of delivery on perceived risks of maternal health outcomes among expectant parents: A cohort study in Beijing, China,'' *BMC Pregnancy Childbirth*, vol. 14, no. 1, p. 12, Dec. 2014.
[38] A. Jović, K. Brkić, and N. Bogunović, ''A review of feature selection methods with applications,'' in *Proc. 38th Int. Conv. Inf. Commun. Technol., Electron. Microelectron. (MIPRO)*, May 2015, pp. 1200–1205.
[39] L. Ali, C. Zhu, M. Zhou, and Y. Liu, ''Early diagnosis of Parkinson's disease from multiple voice recordings by simultaneous sample and feature selection,'' *Expert Syst. Appl.*, vol. 137, pp. 22–28, Dec. 2019.
[40] F. S. Ahmad, L. Ali, H. A. Khattak, T. Hameed, I. Wajahat, S. Kadry, and S. A. C. Bukhari, ''A hybrid machine learning framework to predict mortality in paralytic ileus patients using electronic health records (EHRs),'' *J. Ambient Intell. Humanized Comput.*, pp. 1–11, Aug. 2020.
[41] L. Ali, I. Wajahat, N. Amiri Golilarz, F. Keshtkar, and S. A. C. Bukhari, ''LDA–GA–SVM: Improved hepatocellular carcinoma prediction through dimensionality reduction and genetically optimized support vector machine,'' *Neural Comput. Appl.*, pp. 1–10, Jul. 2020.
[42] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, and E. Duchesnay, ''Scikit-learn: Machine learning in Python,'' *J. Mach. Learn. Res.*, vol. 12, pp. 2825–2830, Oct. 2011.
[43] R. Quinlan, *C4.5: Programs for Machine Learning*. San Mateo, CA, USA: Morgan Kaufmann, 1993.
[44] L. Breiman, ''Random forests,'' *Mach. Learn.*, vol. 45, no. 1, pp. 5–32, 2001.
[45] D. W. Aha, D. Kibler, and M. K. Albert, ''Instance-based learning algorithms,'' *Mach. Learn.*, vol. 6, no. 1, pp. 37–66, Jan. 1991.
[46] O. L. Mangasarian and D. R. Musicant, ''Lagrangian support vector machines,'' *J. Mach. Learn. Res.*, vol. 1, pp. 161–177, Sep. 2001.

MUHAMMAD NAZRUL ISLAM (Member, IEEE) received the B.Sc. degree in computer science and information technology from the Islamic University of Technology, Bangladesh, in 2002, the M.Sc. degree in computer engineering from the Politecnico di Milano, Italy, in 2007, and the Ph.D. degree in information systems from Åbo Akademi University, Finland, in 2014. He is currently an Associate Professor with the Department of Computer Science and Engineering, Military Institute of Science and Technology (MIST), Mirpur Cantonment, Dhaka, Bangladesh. Before joining MIST, he has worked as a Visiting Teaching Fellow with Uppsala University, Sweden, and a Postdoctoral Research Fellow with Åbo Akademi University. He was also a Lecturer and an Assistant Professor with the Department of Computer Science and Engineering, Khulna University of Engineering and Technology (KUET), Bangladesh, from 2003 to 2012. He is the author of more than 80 peer-reviewed publications in international journals and conferences. His research interests include, but not limited to, human–computer interaction (HCI), humanitarian technology, health informatics, military information systems, information systems usability, and computer semiotics. He is a member of The Institution of Engineers, Bangladesh (IEB).

TAHASIN MAHMUD is currently pursuing the B.Sc. degree in computer science and engineering with the Military Institute of Science and Technology (MIST), Dhaka, Bangladesh. His research interests include machine learning, artificial intelligence, computer programming, and computer security.

NAFIZ IMTIAZ KHAN is currently pursuing the B.Sc. degree with the Department of Computer Science and Engineering, Military Institute of Science and Technology (MIST). His research interests include machine learning, artificial intelligence, data science, image processing, edge computing, bioinformatics, and human–computer interaction (HCI).

SUMAIYA NUHA MUSTAFINA (Student Member, IEEE) is currently pursuing the B.Sc. degree with the Department of Computer Science and Engineering, Military Institute of Science and Technology (MIST). Her research interests include machine learning, artificial intelligence, bioinformatics, and data security.

A. K. M. NAJMUL ISLAM received the M.Sc. (Eng.) from the Tampere University of Technology, Finland, and the Ph.D. degree in information systems from the University of Turku, Finland. He is currently an Adjunct Professor with Tampere University, Finland. He is also an Associate Professor with LUT University, Finland. He also works as a University Research Fellow with the Department of Future Technologies, University of Turku. He has more than 90 publications. His research interest includes human centered computing. His research has been published in top outlets, such as IEEE Access, *European Journal of Information Systems*, *Information Systems Journal*, *Journal of Strategic Information Systems*, *Technological Forecasting and Social Change*, *Computers in Human Behavior*, *Internet Research*, *Computers and Education*, *Journal of Medical Internet Research*, *Information Technology and People*, *Telematics and Informatics*, *Journal of Retailing and Consumer Research*, *Communications of the AIS*, *Journal of Information Systems Education*, *AIS Transactions on Human-Computer Interaction*, and *Behaviour and Information Technology*.