Predicting the Appropriate Mode of Childbirth using Machine Learning Algorithm

Md. Kowsher[1]
Department of Applied Mathematics
Noakhali Science & Technology University
Noakhali 3814, Bangladesh

Nusrat Jahan Prottasha[2]
Md. Abdur-Rakib[5], Md. Shameem Alam[6]
Department of CSE
Daffodil International University
Dhaka 1207, Bangladesh

Anik Tahabilder[3]
School of Engineering + Technology
Western Carolina University
Cullowhee, NC 28723, USA

Kaiser Habib[4]
Department of RET
University of Dhaka
Dhaka 1206
Bangladesh

***Abstract*—A woman's satisfaction with childbirth may have immediate and long-term effects on her health as well as on the relationship with her newborn child. The mode of baby delivery is genuinely vital to a delivery patient and her infant child. It might be a crucial factor for ensuring the safety of both the mother and the child. During the baby delivery, decision-making within a short time becomes very challenging for the physician. Besides, humans may make wrong decisions selecting the appropriate delivery mode of childbirth. A wrong decision increases the mother's life risk and can also be harmful to the newborn baby's health. Computer-aided decision-making can be an excellent solution to this problem. Considering this scope, we have built a supervised machine learning-based decision-making model to predict the most suitable childbirth mode that will reduce this risk. This work has applied 32 supervised classifier algorithms and 11 training methods on the real childbirth dataset from the Tarail Upazilla Health complex, Kishorganj, Bangladesh. We have also analyzed the result and compared them using various statistical parameters to determine the best-performed model. The quadratic discriminant analysis has shown the highest accuracy of 0.979992 with the F1 score of 0.979962. Using this model to decide the appropriate labor mode may significantly reduce maternal and infant health risks.**

***Keywords*—Childbirth; labour mode; supervised machine learning; maternal death; infant**

I. Introduction

In baby delivery, we want to make sure that the mother and the child are safe. For this safety, the method of baby delivery is very significant. Usually, the corresponding physician chooses the mode of delivery from two options, includes (i) vaginal birth or (ii) Cesarean area (c-section) birth. So, the patient herself cannot contribute to the decision-making procedure. When a child contains a low-risk pregnancy and is in the head-down position, and the patient is at least 37 weeks pregnant, gynecologists suggest attempting a vaginal birth. In this case, a newborn baby usually gets essential gut bacteria from the mother. Besides, it can help press liquid out of a baby's lungs, decreasing the risk of the baby's breathing problem. This way of birthing also helps for breastfeeding and reduces the baby's risk of asthma and obesity. Additionally, parents will be able to avoid the cost and potential risk of surgery. That is why normal birth is most suitable for both the baby's and the mother's health.

On the other hand, there are cases like twins, or the mother has diabetes, high blood weight, HIV, or active herpes, or the baby is not in a head-down position, which complies the patient to have a c-section delivery. However, it increases the risk of asthma from the early childhood of the baby. In other instances, including delivering a comparatively larger baby for the maternal pelvis, or if the baby is not in a head-down position, the c-section delivery becomes an essential mode of childbirth.

However, many times, physicians are more biased for a c-section delivery than a vaginal delivery. The number count of c-sections is increasing day by day, and it got doubled during 1980. Another record says, as of 2015, the cesarean section rate is exceeding by 35.5% than WHO's recommendation [1]. Lately, during the period 2017-2018, the c-section rate in Bangladesh has increased by 51%, and in 2018, 77% of those c-sections were unnecessary [2]. Save the Children, a popular magazine, has recently documented a 51% increment in extraneous c-section delivery in Bangladesh [3]. In addition to that, maternal mortality is also shown and has been a big problem for most South Asian countries. Compared to the developed countries, the maternal mortality rate in Bangladesh is extremely high. In 2017, the maternal mortality rate in the USA was 0.000017 percent [4]. In the same year, this rate was 0.000113 percent in Bangladesh, which is a few times more than in other developed countries. WHO has reported maternal mortality of 194 per 1000 in Bangladesh. This high maternal death rate can be significantly reduced by selecting the birth mode appropriately. "National Low Birth Weight Survey Bangladesh, 2015" has reported that Maternal mortality and cesarean delivery rates have doubled compared to regular deliveries [5]. It gives us the scope to develop a decision-making model to choose the appropriate mode of childbirth.

Both of the processes of childbirth have advantages and disadvantages based on the particular patient's situation. However, the gynecologist decides the birth mode considering the mother's biological factors, including counting, age, ANC, para, partograph, AMTSL, blood circulation, birth weight, BP, PNC-1 presentation, cervix(OS), membrane, and so on. This research proposed a scientific method to decide childbirth mode considering the mother's present situation and earlier records.

The following points denote the main contributions of this research paper:

- We have proposed a computerized method of decision-making for selecting the appropriate mode of childbirth.
- Since this process is computerized and machine learning-based, it will be less error-prone.
- We have used 32 different classifier algorithms to make the decision more accurate and reliable.
- This model can analyze and use such big data for decision-making that it is merely impossible for a human being to analyze.

The rest of the sections are organized as follows: Section II describes the related work. Section III describes the methods and materials used, Section IV depicts the experimental procedure and the model. Section V examines and evaluates the results, and Section VI describes the conclusion and proposes the future direction of this research.

II. Related Work

A lot of research is being done in the machine learning domain for biomedical decision-making. Mboya IB et al. have proposed a machine learning-based method that can predict perinatal death using supervised machine learning algorithms [6]. ML-based model is also being used for predicting a lot of factors of childbirth. For example, Abraham, Abin, et al. described a new technique for gathering various information from EHRs in order to predict singleton preterm birth by applying various machine learning models [7]. Recently, Islam, Muhammad Nazrul, et al. has presented research regarding childbirth mode with two-fold findings: first, the potential highlights for deciding the method of labor, and second, machine learning algorithms for anticipating the suitable way of labor (vaginal birth, crisis cesarean, cesarean birth) [8]. Kowsher, M. et al. has reported good accuracy in applying machine learning-based recommendation system to predict the most appropriate childbirth mode [9]. Also, Khan, Nafiz Imtiaz, et al. have a similar test to anticipate whether the cesarean area is essential with the assistance of information mining and subsequently expand the mother and infant's security during and after labor by staying away from a pointless cesarean segment [10]. Besides, Fu, Yuanqing, et al. had described a model to recognize early life hazard factors for youth overweight/stoutness among preterm babies and decided to take care of practices that could alter the distinguished danger factors [11]. Other researchers also have applied machine learning models to classify various biomedical factors and hence to conclude the adverse effect of c-section delivery. For instance, Siddiqui, Mohammad Khubeb, et al. described that machine learning classifiers could use EEG information and identify seizures alongside uncovering applicable reasonable examples without trading off execution [12]. In addition, Soh, Yan Xi, et al. had explained the relationships among sociodemographic and medicine factors, the concern of parturition, psychosocial wellbeing, and childbirth self-efficacy employing a structural equation modeling approach [3]. c-section delivery may have a postbirth adverse health effect on a mother. Chen, Yanfang, et al. showed the relationship between conveyance and post-traumatic stress problems that yielded conflicting outcomes. This examination is expected to research the relationship between conveyance and post-traumatic stress in an associate of Chinese ladies with a high pace of cesarean conveyance [14].

Zhang, Yiye, et al. propose a machine learning structure for PPD hazard expectation utilizing information extricated from electronic wellbeing records (EHRs) [15]. Later on, Lipschuetz, Michal, et al. presented to decide the customized forecast of vaginal birth after cesarean conveyance utilizing 30 an AI calculation that might help patient-doctor dynamic and 31 increment paces of preliminaries of work [16]. Also, Serçekuş, Pınar, Okan Vardar, and Sevgi Özkan proposed to recognize and think about the dread of labor and related to variables among pregnant ladies and their accomplices [17]. Onchonga, David et al. described a new investigating ladies' experience from maternity specialists drove incorporated pre-birth preparing and its effect on the dread of labor [18]. After that, Liu, Ligue, et al. proposed an expectation model of undeveloped improvement by using machine learning algorithms dependent on authentic case information. In this way, specialists can make more exact ideas on the quantity of patient subsequent meet-ups and give choice help to moderately unpracticed specialists in clinical practice [19]. On the other hand, Lindblad Wollmann, Charlotte, et al. described the predicting vaginal birth in ladies with one earlier cesarean and no vaginal conveyances utilizing machine learning strategies [20].

Unlike their works, we showed and analyzed various methodology of supervised classifiers based on a real dataset of childbirth to figure out the best model to predict the suitable mode of delivery.

III. Methodology

To build our proposed model, we went through four significant phases: Dataset formation, data preprocessing, training the models, the performance analysis of the model. We have collected the data from the Tarail Upazilla Health Complex, a specialized clinic for maternal care located in Tarail, Kishorganj, Bangladesh. First, we determined the features that influence our targeted feature, the mode of childbirth. We kept the most significant of them, and some other features having less impact were deleted as they don't contribute much to the targeted variable. Then the data was split into two sections, i.e., training set and test set, which are later used for training and testing correspondingly. After collecting raw data, we had preprocessing to make it suitable for the machine learning model. Data pre-processing techniques have made the data outliers free and more solid, and it also increases the accuracy. As a result, we used several preprocessing steps such as cleaning data, missing value handling, categorical data handing, feature selection, feature scaling. Having completed all the preprocessing steps, the data becomes ready for the machine learning models. We have used several groups of supervised learning classifiers such as Tree, Ensemble, Neighbors, Naive Bayes, Calibration, Discriminant Analysis, SVM, Linear model, Gaussian Process, and Deep Neural Network. Most of those classifiers have shown good performance with this preprocessed training and test dataset. The methodology of our proposed model is depicted in Fig. 1.

A. Dataset Description

We have used a dataset that is containing the medical records of 13527 women. It has 21 diverse observation values for every pregnant woman counting title, age, address, admission time and time, ANC number of shrouds (by therapeutically prepared supplier), para, the reason of confirmation, amid pregnancy (week), cesarean, breech conveyance, partograph, blood circulation, AMTSL, birth weight, PNC-1(postnatal administrations and the status of the patient), PNC1 (postnatal delivery administrations), BP, introduction, layer and cervix (OS).

Para alludes to the total number count of pregnancies that the lady has carried on the last twenty weeks of pregnancy. This number includes both live births and pregnancy misfortunes after twenty weeks. Gravida referred to the number of times of affirmed pregnancies of a lady, including both live birth and interrupted pregnancies. ANC (ante-natal check-up) implies a routine checkup for the mother to ensure appropriate facility for further safety. ANC is usually conducted in three stages, the first one is between 4^{th} to 28^{th} week, the second one is between 28^{th} to 36^{th} week, and the last one is between 36^{th} to 48^{th} week. There is a possibility that the infant can open its mouth almost 10 cm or over, then it can be conveyed in an ordinary way. On the other hand, if curved (OS) isn't 10 cm, even it is over 12 hours, this patient requires a cesarean conveyance.

PNC implies postnatal care. After conveyance, this is often done by checking the typical state of the mother. Cephalic is for typical conveyance, but in some cases, there's a breach at multi-case, but typical conveyance is done. Cesarean conveyance happens sideways or transverse. Pantographs are utilized to decide the physical condition of the mother and child. After children's birth, Placenta is extricated within the AMTCL strategy. Blood circulation is given on the off chance that the quiet is Iron deficient. Most of the time, the layer remains completely intact. Some of the time, it endures from spilling, burst. Blood weight is watched to screen the typical arrangement of the mother.

B. Data Pre-processing

In machine learning, data preprocessing is in the approach of transferring or encoding the raw data in a phase where algorithms can use the data for building the model. We need to preprocess the data accordingly to make it fit for the machine learning model. A well-processed data gives high accuracy and makes the model more reliable. Here, we have used several stages of preprocessing, which have been illustrated in Fig. 2.

In our dataset, there are lots of incomplete, null, and duplicate values. For this reason, we took three steps to correct these data. Those three steps are described below:

- We have noticed that there are many data points that are repeated in a row. Therefore, we simply removed all the duplicate data other than one single observation.
- We also notice some of the rows and columns are empty. We also erased the entire empty rows or columns from our dataset.
- There were some rows and columns that have 50% or more incomplete or null values. We removed the entire rows and columns to fix this issue.
- There were some columns that have very low variance. We also ignored those columns to make the dataset better suited for our machine learning model.

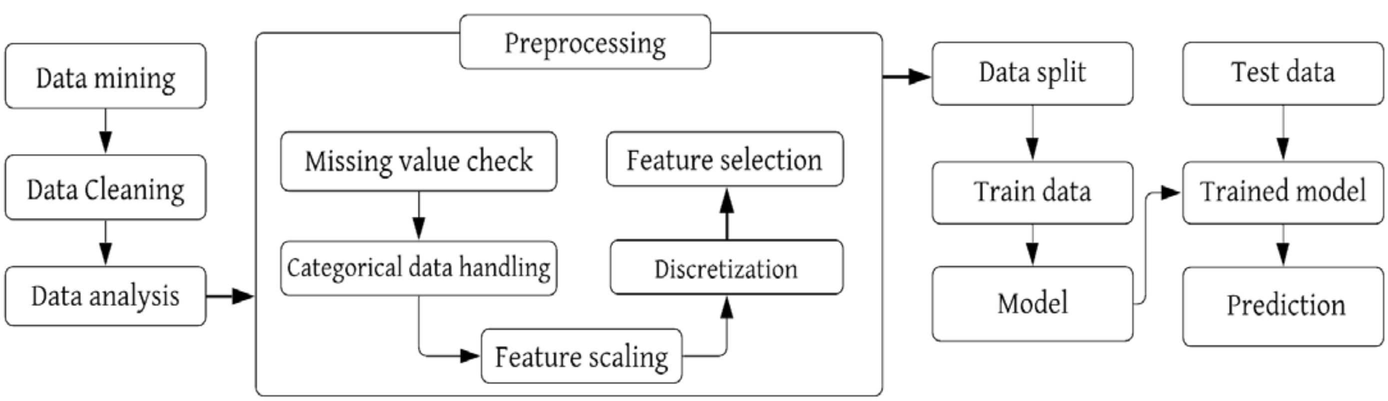

Fig. 1. Workflow of the Proposed Model.

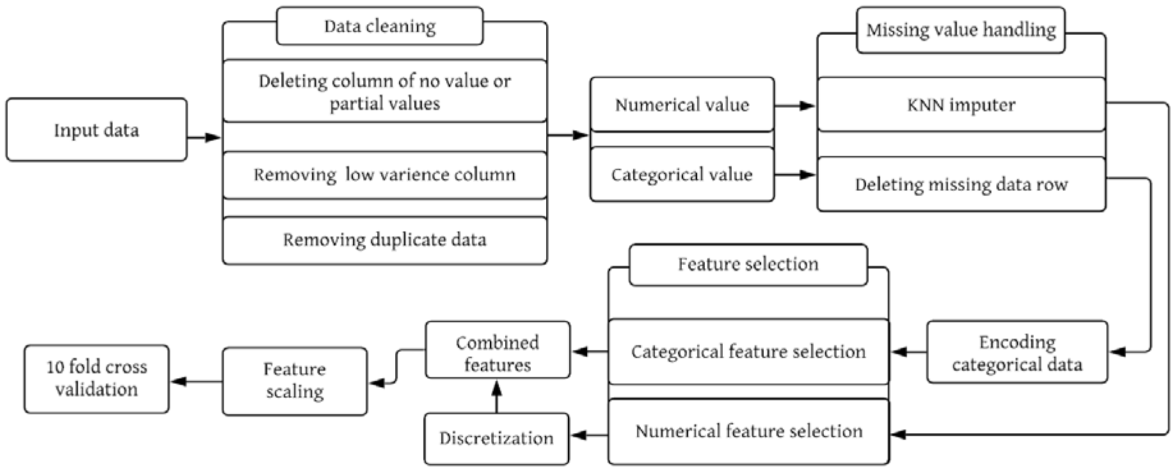

Fig. 2. Data Preprocessing Stages.

Generally, a missing value is defined as the value which is either not stored in the sample or missing partial information. The missing value is commonly seen in any kind of dataset. Our dataset also had some missing values. However, most predictive modeling methods can't handle any missing value. Hence, this issue must be solved before we feed this data into the machine learning model. Sometimes, median, mean, mode methods are used to update a missing value. The selection of the methods depends on the data types and the totals number of observations. However, the most straightforward procedure for dealing with the missing value is to remove the whole row for categorical features and replace the missing value by selecting the nearest neighbors for numerical data. We have used the K Nearest Neighbor aka KNN based method for a more accurate missing value imputation and replaced the NAN data by getting the nearest neighbor value. We have considered three neighbors for KNN algorithm implementation and completed all the missing values using KNN imputer to build a perfect feature matrix. The following Fig. 3 has illustrated the working procedure of the KNN imputer.

Categorical data is a qualitative feature whose values are taken based on the value of labels. So, we need to encode this type of data into numbers so that the machine learning model can implement mathematical operations on it. In our dataset, there are a total of three categorical variables, including "PRESENTATION", "REASON" and "MEMBRANE". We have used one-hot encoding, one of the most popular encoding algorithms, to encode the categorical values into numbers. It is the most general approach, and it works well unless any categorical variable takes a large number of different values. After this encoding, a binary matrix is formed where 1 indicates the presence of any value and 0 indicates the absence of the value.

Feature selection is a critical stage of implementing a machine learning model. It is the process of determining the mathematical relation between the feature variable and the target variable. We have kept the most significant features and dropped some features with less significance on the targeted variable. Reduction in features reduces the computational cost. Our dataset contains 21 features, and we have considered the p-value for finding the probability of the null hypothesis. The features with a p-value less than 0.05have been taken out. After checking multicollinearity, we have maintained a strategic distance from those components, which show repetition and don't back the p-value assumption. Besides, to handle the numerical feature, we took the help of the Pearson correlation coefficient, which is defined in equation -1, and for categorical features, we used the ANOVA F measurement, which is described in equation -2.

$$r = \frac{\sum_{i=1}^{n}(x_i-\bar{x})(y_i-\bar{y})}{\sqrt{\sum_{i=1}^{n}(x_i-\bar{x})^2}\sqrt{\sum_{i=1}^{n}(y_i-\bar{y})^2}} \tag{1}$$

$$F = \frac{n\sum_{i=1}^{n}(\overline{x_k}-\overline{x_G})^2/(k-1)}{\sqrt{\sum_{i=1}^{n}(x_i-\overline{x_k})^2/(N-k)}} \tag{2}$$

After performing the feature selection, we had the most relevant nine features, including Para, Age, Cervix (OS), Gravida, Systolic, Diastolic, Reason, FHR(BPM), and Presentation.

In data analysis, it can often be observed that the numerical data are mostly like skewed or non-standard deviation due to outliers, multi distributions, very exponential distributions, and more. We converted the numeric value into categorical behavior to solve this problem. We have applied the discretization method to converts the numerical value into a distribution function.

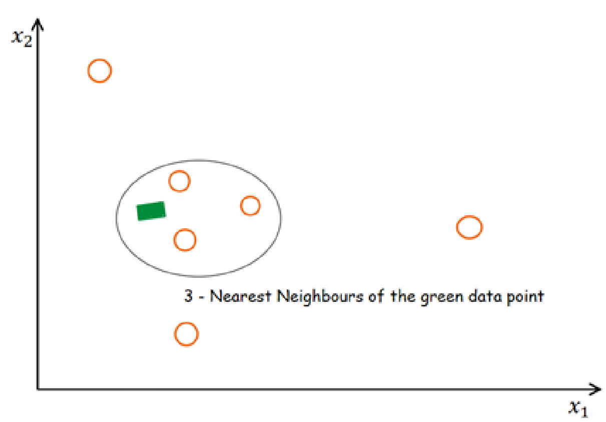

Fig. 3. The KNN Imputer for Missing Value Handling

Feature scaling is one of the crucial techniques that are mandatory to standardize the working data's independent features. Nevertheless, there exist various methods like Min-Max Scaling, Variance Scaling, Standardization, Mean Normalization, and Unit vectors for feature scaling. In our work, we have applied the Min-Max scaling as a feature scaling technique. Here, the transferred range between 0 and 1. The min-max scaling can be written as shown in equation- 3 below,

$$\acute{x} = \frac{x - \min(x)}{max(x) - \min(x)} \tag{3}$$

Cross-Validation is used to assess the models' predictive performance and judge the performance of a new data set. This is often fundamentally a variation of Recursive Feature Elimination (RFE) with cross-validation in each iteration. In RFE, the features are disposed of using a backward selection of the features. It uses a base classifier for selecting the features of data. It starts by building a model with the complete set of features and computes an importance score for each variable. The features with the least imperative score are deleted. The method recursively finds the optimal set of features that gives the most excellent model accuracy. With the cross-validation, the training and test set is divided into k number of folds where the k-1 fold of them is used for training and the rest 1 fold of data is used for testing. After running iteratively for K times, we actually get the average accuracy. This method takes more computation, but we can ensure good accuracy with comparatively fewer data. In our project, we used ten-fold cross-validated features to build our classification models.

C. Model Description

Implementing a lot of diverse models can ensure the best possible accuracy. So, we have applied the 32 most sensible machine learning classifiers, including Tree, Ensemble classifier, Neighbors, Naive Bayes, Calibration, Discriminant Examination, SVM, Linear model, to predict the birth mode Gaussian Process, and deep neural networks. All models are shown in Fig. 4.

Tree-based algorithms are considered to be one of the leading and most used supervised learning methods. In this work, we have implemented a decision trees and additional tree classifiers. In these two algorithms, we utilized "gini" for the Gini impurity, and the splitter is chosen as 'best' to select the part at each node.

Ensemble methods are procedures that make multiple models and combine them to create moves forward. Here, we utilized five ensemble-based classifiers [21]. These are AdaBoost, Stowing, Gradient Boosting, enable hist gradient boosting, Random Forests classifiers. In all these classifiers, in AdaBoost classifiers, the number of boosting estimators is 50 with the SAMME.R as a real boosting algorithm. In the bagging classifier, we used ten estimators. The loss function of Gradient Boosting is 'deviance', logistic regression with probabilistic outputs with the 100 boosting stages. Besides, in the section of random forest three, we use 100 trees as a forest with Gini impurity.

Afterward, we have utilized three neighbors' classifiers of statistical pattern recognition [22]. These are radius neighbor, k-neighbor, and nearest centroid. In KNN, we used five neighbors for every iteration. Besides, the Makowski metric is chosen for all neighbor classifiers.

Naïve Bayes is based on an estimate of the age of the "naive" with a set of learning calculations guided by an application and an estimate of the accuracy between them [23].

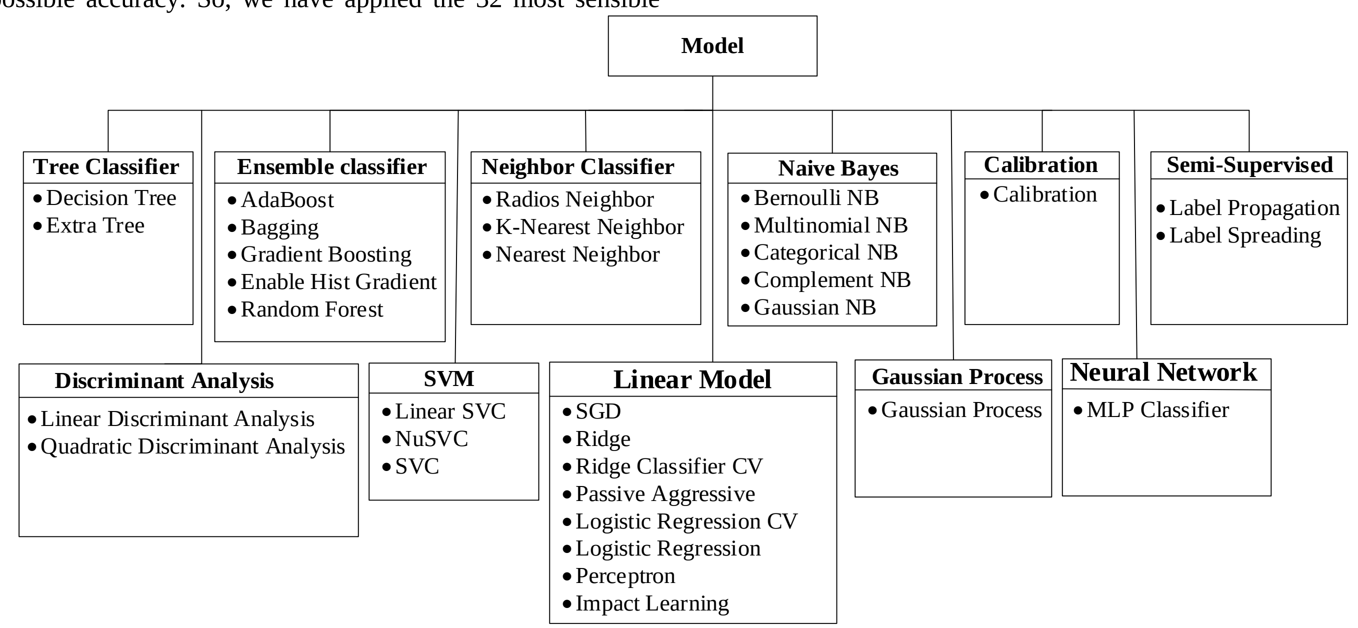

Fig. 4. Tabulation of all the Models.

Here Bernoulli, Multinomial, Categorical, Complement, and Gaussian Naive Bayes are executed to compare the Bayes algorithm on childbirth mode detection. For each method, we have also used the additive smoothing parameter.

Model calibration implies the process where we take a model that is already trained and apply a post-processing operation, which improves its probability estimation. Thus, if we were to inspect the samples that were estimated to be positive with a probability of 0.85, we would expect that 85% of them are in fact positive.

The label propagation and label spreading are used in the area of semi-supervised algorithms. The Gamma Parameter for RBF bit is utilized as 20. The maximum emphasis is 1000, and the neighbor parameters are 7.

The researchers commonly use discriminant analysis to analyze the data when the criterion or the dependent variable is categorical, and the predictor or the independent variable is the interval in nature. Dependent variables should categorize at the moment as well as include predictive or distinct variable natures such that researchers can use them to analyze research data (quadratic inequality analysis (QDA) and linear inequality analysis (LDA) [24].

The support vector machine, aka SVM, is used mainly for exploring a hyperplane in d-dimensional space that notably classifies the data points [25]. In the linear SVC, we used hinge as loss function with l_2 penalty. The numerical value three is used as the polynomial kernel in NuSVC with the RBF kernel type.

The linear model could be a module lesson if it contains diverse functions for performing machine learning linearly [26]. We utilized the eight classifiers such as SGD, Ridge Classifier, Ridge Classifier CV, Passive-Aggressive Classifier, Logistic Regression, Logistic Regression CV, Perceptron, and Impact Learning [27].

The Gaussian process is a stochastic method in probabilistic hypotheses and statistics, such as a common multivariate distribution containing which is a limited random sample collection. The kernel of the gaussian process classifier specifies the covariance function, and the accessible internal optimizers are 'fmin_l_bfgs_b'.

A neural network could be an arrangement of algorithms that endeavors to recognize basic relationships in a set of data through a method that imitates the way the human brain works. The Artificial Neural Network (ANN) is a computing system where neurons inspire people [28]. There are three layers, and these are the input layer, hidden layer, and output layer. The input layer usually takes the input data into the network. The hidden layer is the layer where input and output are connected based on conditions. The output level is decided by considering the respect action, weight, and hidden level. There's no rule of the thumb to select the hidden layer in ANN. We have used sixty-four hidden layers between the input and output layers.

IV. Experiment

In furtherance of our experiment from the proposed work, we have first assembled the model and trained it. Thirty-two classifiers from supervised learning based on different learning methodologies have been implemented to predict childbirth's most applicable mode. This section described different experimental tasks for the performance analysis and evaluation and compared all algorithms. Besides, we have illustrated the experimental setup used to execute the whole task and used 11 statistical evaluation metrics for analysis performance. Finally, we have also compared with other works related to this issue regarding the best version of our work.

A. Experimental Setup

We have completed the whole computation in google colab, a python simulation environment provided by Google. This environment comes with parallel computation facilities for fast execution. We have used the most popular libraries to make easy and expressive data structures work well and intuitively with fast, flexible, and time-series data. Finally, the scikit-learn Library contains specialized machine learning and statistical modeling tools, including classification, regression, and clustering algorithms for modeling. We have used a machine learning framework named sci-kit learn and deep learning framework Keras to implement the classification algorithm. Finally, we used Matplotlib for data visualization, graphical representation, and also for data analysis.

B. Measurement Metrics

We have used several Statistical metrics [29] for measurements, evaluation, and analysis of the performances and compared all the algorithms. We will define and describe all of those in the following section.

We have analyzed the 11 statistical measurements, including accuracy, F1 score, precision, recall, and so on. Accuracy and F1 score are the most important of them.

Accuracy is a metric that evaluates the matric for the correct prediction rate for the positive class. The expression is shown below in equation 4.

$$Accuracy = \frac{True\ Positive(TP)}{True\ Positive(TP) + False\ Positive(FP)} \tag{4}$$

F1 score conveys the balance between the precision and the recall. It is also called the F Score or the F Measure. A good F1 score indicates that we have low false positives and low false negatives in the results. The expression of the F1 score is shown in equation 5.

$$F1 = \frac{TP}{TP + 0.5(FP + FN)} \tag{5}$$

Recall considers the percentage of correct predictions for all the positive categories. In other words, recall is how many of the true positives were recalled (found), i.e., how many of the correct hits were also found. The expression of recall score is shown below in equation 6.

$$RS = \frac{TP}{TP + FN} \tag{6}$$

The F-beta score is evaluated in the binary classification model based on a configurable single-score for the positive class's forecasts. It's also calculated utilizing precision and recall. The value of the F-beta score can be calculated using equation 7 below.

$$FBS = \frac{(1+\beta^2)\,.(Precision\,.Recall)}{(\beta^2\,.Precision + Recall\,)} \quad (7)$$

Hamming loss is designed for multiclass while Precision, Recall, F1-beta score represents one clear single-presentation-value for multiple-label cases compared to the precision/recall/f1beta score that can be assessed only for independent binary classifiers for each label. The expression of Hamming loss is shown below in equation 8.

$$HL = \frac{1}{|D|}\sum_{i=1}^{|D|} \frac{|Yi\,\Delta\,Zi|}{|L|} \quad (8)$$

In the Jaccard similarity coefficient, the union of the two label sets is used to compare the set of labels predicted in y_true to mark the separate intersection as a measure by calculation. The equation of Jaccard similarity coefficient is shown below in equation 9.

$$\mathrm{J}(A,B) = \frac{|A \cap B|}{|A \cup B|} \quad (9)$$

Matthews Correlation Coefficient, aka MCC, is used as a standard for binary and multiclass classification in machine learning. The equation of Matthews Correlation Coefficient is shown below in equation 10.

$$\mathrm{MCC} = \frac{TP\,.FN - FP\,.FN}{\sqrt{(TP+FP).(TP+FN).(TN+FP).(TN+FN)}} \quad (10)$$

AUC stands for "Area under the ROC Curve." That is, AUC measures the entire two-dimensional area underneath the entire ROC curve (think integral calculus) from (0,0) to (1,1). AUC provides an aggregate measure of performance across all possible classification thresholds. One way of interpreting AUC is as the probability that the model ranks a random positive example more highly than a random negative example.

The balanced accuracy in binary (BAC) and multiclass classification is usually used to measure the performance if the dataset is imbalanced. The equation of balance accuracy is shown below in equation 11.

$$\mathrm{BAC} = \frac{1}{2}\left(\frac{TP}{P} + \frac{TN}{N}\right) \quad (11)$$

Cohen's kappa (CKS) is a statistic that measures inter-annotator agreement. We can consider Cohen's Kappa as a quantitative measure of reliability for two raters that are rating the same thing. The equation of Cohen's kappa (CKS) is shown below in equation 12.

$$\mathrm{K} = \frac{P_0 - P_E}{1 - P_E} \quad (12)$$

V. Result Analysis

Our experiment improved the methods of decision-making. We have implemented thirty-two classifier parameters to gain the best possible performance. After getting the performance matrix from all the models, we have tabulated the data into a table. We have analyzed the 11 statistical measurements, including accuracy, F1 score, RS, PS, FBS, HL, JS, MS, AUC, BAC and CKS. The statistical measure and for each of the algorithms is shown in Table I below. In addition, we have compared the performance of all the proposed models and determine the best suitable model that can be used in real-life decision-making or selecting the most suitable mode of childbirth.

From Table I, we can see that the decision tree classifier has predicted the best accuracy of 0.918307, and the F1 score is 0.918198 from the branch of the Tree. Secondly, the hist gradient boosting classifier has gained the best accuracy of 0.959158 with an F1 score is 0.959071 from the section of Ensemble's algorithm. Thirdly, the KNN classifier has acquired better accuracy, which is 0.961015, along with an F1 score of 0.960853 from the area of neighbor's classifiers. Also, from the section of naive Bayes algorithms, we can figure out that Gaussian naive Bayes has shown the best accuracy of 0.874381 and its F1 score is 0.872635 among all naive Bayes classifiers. Next, calibration also has placed the best accuracy of 0.877063, and its F1 score is 0.875664 from the branch of naive Bayes. After that, we can see from the semi-supervised classifiers, the label propagation has gained the best accuracy of 0.906972, and the F1 score is 0.906223. Besides, it can also be seen from the discriminant analysis section that the quadratic discriminant analysis has proved the best position of accuracy 0.979992 with an F1 score of 0.979962. Moreover, SVC is the best for predicting childbirth mode with an accuracy of 0.956477 and an F1 score of 0.956302 from all algorithms of SVM. Furthermore, we also can find out that the Gaussian process classifier placed the accuracy of 0.891708 and the F1 score is 0.890333. In the neural network, the multilayer perceptron classifier has acquired a good performance with an accuracy of 0.954404, and the F1 score is 0.954299.

Overall, by considering all the sections of algorithms for the prediction of childbirth mode, we can observe that the quadratic discriminant analysis is the winner with an accuracy of 0.979992 and the F1 score is 0.979962. The neighbors' classifier also comes up with the second-best with an accuracy of 0.961015, and the F1 score is 0.960853.

TABLE I. PERFORMANCE ANALYSIS

Name	Accuracy	F1S	RS	PS	FBS	HL	JS	MCC	AUC	BAC	CKS
Tree Classifiers											
DtC	0.918317	0.918198	0.918236	0.918166	0.918178	0.081683	0.850266	0.877476	0.951797	0.918236	0.877473
ETC	0.902228	0.902208	0.902168	0.902268	0.902242	0.097772	0.823259	0.853352	0.93791	0.902168	0.853342
Ensemble Classifiers											
AdC	0.772896	0.760404	0.77305	0.79518	0.774519	0.227104	0.637223	0.681965	0.857822	0.77305	0.659515
BC	0.94637	0.946173	0.946315	0.946145	0.946142	0.05363	0.898545	0.919609	0.970062	0.946315	0.919553
GBC	0.950083	0.94991	0.950032	0.949875	0.949878	0.049917	0.905279	0.925165	0.971867	0.950032	0.925123
HGBC	0.959158	0.959071	0.959129	0.959056	0.959057	0.040842	0.921829	0.938758	0.977144	0.959129	0.938737
RFC	0.956271	0.95605	0.956208	0.956053	0.956032	0.043729	0.916453	0.934484	0.975667	0.956208	0.934404
Neighbors Classifiers											
RNC	0.796823	0.793171	0.796323	0.795863	0.794037	0.203177	0.667912	0.697534	0.878434	0.796323	0.695157
KNC	0.961015	0.960893	0.960944	0.960897	0.960889	0.038985	0.925345	0.941547	0.977879	0.960944	0.94152
NC	0.808375	0.803996	0.80814	0.804359	0.803641	0.191625	0.681652	0.714516	0.886066	0.80814	0.712552
Naive Bayes Classifiers											
BNB	0.331271	0.165918	0.333126	0.110469	0.127515	0.668729	0.110446	-0.00883	0.499843	0.333126	-0.00031
MNB	0.808168	0.803797	0.808056	0.805796	0.804182	0.191832	0.681258	0.715031	0.894046	0.808056	0.71227
CNB	0.336015	0.16767	0.333333	0.112005	0.129157	0.663985	0.112005	0	0.5	0.333333	0
CoNB	0.718647	0.681511	0.718514	0.74856	0.699551	0.281353	0.539377	0.609911	0.834682	0.718514	0.577959
GNB	0.874381	0.872635	0.874237	0.872807	0.872524	0.125619	0.777803	0.81239	0.93513	0.874237	0.81156
Calibration Classifier											
CC	0.877063	0.875664	0.87692	0.87552	0.875444	0.122937	0.782584	0.816105	0.935097	0.87692	0.815584
Semi-Supervised Classifier											
LP	0.906972	0.906223	0.906743	0.90729	0.906667	0.093028	0.831271	0.861163	0.947339	0.906743	0.860439
LS	0.902847	0.901956	0.902615	0.902912	0.902333	0.097153	0.824296	0.854988	0.945149	0.902615	0.85425
Discriminant Analysis Classifiers											
LDA	0.870462	0.868522	0.870233	0.868103	0.868115	0.129538	0.774095	0.806278	0.922529	0.870233	0.805682
QDA	**0.979992**	**0.979962**	0.979964	0.979987	0.979974	0.020008	0.960818	0.97	0.990303	0.979964	0.969987
SVM Classifiers											
LSVC	0.880982	0.879968	0.880805	0.879558	0.879671	0.119018	0.790406	0.821664	0.930283	0.880805	0.821466
NuSVC	0.924711	0.923732	0.924601	0.924279	0.923889	0.075289	0.860045	0.887741	0.960271	0.924601	0.887061
SVC	0.956477	0.956302	0.956403	0.956278	0.956279	0.043523	0.917007	0.934751	0.975629	0.956403	0.934713
Linear Model Classifiers											
SGDC	0.871493	0.869171	0.871364	0.870912	0.869713	0.128507	0.774591	0.809034	0.922391	0.871364	0.807246
RdC	0.865924	0.863561	0.865752	0.864239	0.863613	0.134076	0.764468	0.800186	0.919536	0.865752	0.798875
RdCV	0.865924	0.86358	0.865754	0.864204	0.863609	0.134076	0.764519	0.800154	0.920406	0.865754	0.798875
PAC	0.828383	0.82831	0.82787	0.867552	0.846668	0.171617	0.714729	0.761057	0.893194	0.82787	0.742496
LRCV	0.891502	0.890906	0.891326	0.89062	0.890718	0.108498	0.808021	0.837313	0.932856	0.891326	0.837249
LR	0.883457	0.882394	0.883287	0.882026	0.88211	0.116543	0.79431	0.825425	0.931482	0.883287	0.82518
Pr	0.837871	0.836795	0.837884	0.842252	0.839304	0.162129	0.727938	0.760099	0.903731	0.837884	0.756858
IL	0.880982	0.879634	0.880853	0.879964	0.879639	0.119018	0.790078	0.822198	0.930677	0.880853	0.821475
Gaussian Process Classifiers											
GPC	0.891708	0.890444	0.891566	0.890383	0.890278	0.108292	0.80634	0.838053	0.938326	0.891566	0.837557
Neural Network Classifier											
MLPC	0.954414	0.954259	0.954337	0.954276	0.954258	0.045586	0.913242	0.931665	0.975341	0.954337	0.931619

VI. Conclusion and Future Work

Selection of the best baby delivery methods is crucial for protecting both the mother and the newborn baby. But it still remains to explore the best sets of features when making this decision in a computerized way. That's why we try to leverage AI to recognize the best mode of baby delivery. Nowadays, machine learning, deep learning, and other computerized computation models are ubiquitously being used in medical decision-making. Here, we have used machine learning-based binary classification algorithms for decision-making between two methods of childbirth. This model will assist the doctor in making a more accurate decision within a very short time. This machine-learning-aided decision will not replace the necessity of the doctors for decision-making. Instead, it will help the physician to gain a deeper insight into the patient's information available. The way of decision-making using this model is very computerized and less likely to have an error.

The dataset we have used in this project is not very robust. In the future, we want to add much more observation to our dataset and make this model much more general. We believe a large set of data will produce better accuracy and less overfitting. Besides, we will implement a more in-depth learning-based classification to expand the investigation and make the top choice for record-breaking performance. After childbirth, we plan to implement this system to predict other real-life biomedical factors in advance. In addition, the data can be collected during the whole nine months of the mother's pregnancy. In the future, we plan to make a GUI of this model available to physicians, who can use it as like medical device for decision making.

References

[1] Maeda, Eri, et al. "Cesarean delivery rates for overall and multiple pregnancies in Japan: A descriptive study using nationwide health insurance claims data." Journal of Obstetrics and Gynaecology Research (2021).

[2] Chen, Qian, et al. "The impact of cesarean delivery on infant DNA methylation." BMC pregnancy and childbirth 21.1 (2021): 1-8.

[3] Ali, Nazia Binte, et al. "Are childbirth location and mode of delivery associated with favorable early breastfeeding practices in hard to reach areas of Bangladesh?." Plos one 15.11 (2020): e0242135.

[4] Serçekuş, Pınar, Okan Vardar, and Sevgi Özkan. "Fear of childbirth among pregnant women and their partners in Turkey." Sexual & Reproductive Healthcare 24 (2020): 100501.

[5] Shaheen, Razia, et al. "Prevalence of Low Birth Weight in Urban Dhaka and its Association with Maternal Age and Socioeconomic Status." Dr. Sulaiman Al Habib Medical Journal 2.4 (2020): 162-166.

[6] Mboya, Innocent B., et al. "Prediction of perinatal death using machine learning models: a birth registry-based cohort study in northern Tanzania." BMJ Open 10.10 (2020): e040132.

[7] Abraham, Abin, et al. "Dense phenotyping from electronic health records enables machine-learning-based prediction of preterm birth." medRxiv (2020).

[8] Islam, Muhammad Nazrul, et al. "Exploring Machine Learning Algorithms to Find the Best Features for Predicting Modes of Childbirth." IEEE Access (2020).

[9] Kowsher, M., Prottasha, N. J., Tahabilder, A., & Islam, M. B. (2020, February). Machine Learning Based Recommendation Systems for the Mode of Childbirth. In International Conference on Cyber Security and Computer Science (pp. 295-306). Springer, Cham.

[10] Khan, Nafiz Imtiaz, et al. "Prediction of Cesarean Childbirth using Ensemble Machine Learning Methods." (2020).

[11] Fu, Yuanqing, et al. "Integration of an interpretable machine learning algorithm to identify early life risk factors of childhood obesity among preterm infants: a prospective birth cohort." BMC medicine 18.1 (2020): 1-10.

[12] Siddiqui, Mohammad Khubeb, et al. "A review of epileptic seizure detection using machine learning classifiers." Brain informatics 7 (2020): 1-18.

[13] Soh, Yan Xi, et al. "Determinants of childbirth self-efficacy among multi-ethnic pregnant women in Singapore: a structural equation modeling approach." Midwifery 87 (2020): 102716.

[14] Chen, Yanfang, et al. "Association between mode of birth and post-traumatic stress disorder following childbirth: a prospective cohort study of Chinese women." (2020).

[15] Zhang, Yiye, et al. "Development and validation of a machine learning algorithm for predicting the risk of postpartum depression among pregnant women." Journal of Affective Disorders 279 (2020):1-8.

[16] Lipschuetz, Michal, et al. "Prediction of vaginal birth after cesarean deliveries using machine learning." American journal of obstetrics and gynecology 222.6 (2020): 613-e1.

[17] Serçekuş, Pınar, Okan Vardar, and Sevgi Özkan. "Fear of childbirth among pregnant women and their partners in Turkey." Sexual & Reproductive Healthcare 24 (2020): 100501.

[18] Onchonga, David, et al. "Midwife-led integrated pre-birth training and its impact on the fear of childbirth. A qualitative interview study." Sexual & Reproductive Healthcare 25 (2020): 100512.

[19] Liu, Lijue, et al. "Machine learning algorithms to predict early pregnancy loss after in vitro fertilization-embryo transfer with fetal heart rate as a strong predictor." Computer Methods and Programs in Biomedicine 196 (2020): 105624.

[20] Lindblad Wollmann, Charlotte, et al. "Predicting vaginal birth after previous cesarean: using machine ‐ learning models and a population ‐ based cohort in Sweden." Acta obstetricia et gynecologica Scandinavica (2020).

[21] Pintelas, Panagiotis, and Ioannis E. Livieris. "Special issue on ensemble learning and applications." (2020): 140.

[22] Wang, Yizhen, Somesh Jha, and Kamalika Chaudhuri. "Analyzing the robustness of nearest neighbors to adversarial examples." International Conference on Machine Learning. PMLR, 2018.

[23] Zhang, Huan, Liangxiao Jiang, and Liangjun Yu. "Class-specific attribute value weighting for Naive Bayes." Information Sciences 508 (2020): 260-274.

[24] Charles, Vincent, Juan Aparicio, and Joe Zhu. "The curse of dimensionality of decision-making units: A simple approach to increase the discriminatory power of data envelopment analysis." European Journal of Operational Research 279.3 (2019): 929-940.

[25] Suykens, Johan AK, and Joos Vandewalle. "Least squares support vector machine classifiers." Neural processing letters 9.3 (1999): 293-300.

[26] Stroup, Walter W. Generalized linear mixed models: modern concepts, methods and applications. CRC press, 2012.

[27] Kowsher, Md, Anik Tahabilder, and Saydul Akbar Murad. "Impact-learning: a robust machine learning algorithm." Proceedings of the 8th International Conference on Computer and Communications Management. 2020.

[28] Kukreja, Harsh, et al. "An introduction to artificial neural network." Int J Adv Res Innov Ideas Educ 1 (2016): 27-30.

[29] Carvalho, Diogo V., Eduardo M. Pereira, and Jaime S. Cardoso. "Machine learning interpretability: A survey on methods and metrics." Electronics 8.8 (2019): 832.