

Predicting the Use Labor Induction Intervention: A Machine Learning Approach for the Birth-Cohort Registry at A Tertiary Hospital in North Tanzania.

Clifford Tarimo

College of Public Health, Zhengzhou University 100 Kexue Avenue, Zhengzhou, Henan 450001 China

Soumitra Bhuyan

Rutgers University New Brunswick, Edward J. Bloustein, School of Planning and Public Policy

Quanman Li

College of Public Health, Zhengzhou University 100 Kexue Avenue, Zhengzhou, Henan 450001 China

Michael Mahande

Institute of Public Health, Kilimanjaro Christian Medical University College, P.O. Box 2240, Moshi, Tanzania.

Jian Wu (✉ jianwuzzu@163.com)

College of Public Health, Zhengzhou University 100 Kexue Avenue, Zhengzhou, Henan 450001 China

Research Article

Keywords: Machine learning models, predictors, labor induction, north-Tanzania

Posted Date: January 22nd, 2021

DOI: <https://doi.org/10.21203/rs.3.rs-147720/v1>

License:  This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

Predicting the use labor induction intervention: A machine learning approach for the birth-cohort registry at a tertiary hospital in north Tanzania.

Clifford Silver Tarimo^{1, 2}, Soumitra S. Bhuyan³, Quanman Li², Michael Johnson Mahande⁴, Jian Wu^{2*}

1. Department of Science and Laboratory Technology, Dar es salaam Institute of Technology, P.O. Box 2958, Dar es Salaam, Tanzania

2. College of Public Health, Zhengzhou University 100 Kexue Avenue, Zhengzhou, Henan 450001 China

3. Rutgers University New Brunswick, Edward J. Bloustein, School of Planning and Public Policy

4. Institute of Public Health, Kilimanjaro Christian Medical University College, P.O. Box 2240, Moshi, Tanzania.

*Corresponding author
E-mail: wujian@zzu.edu.cn

Abstract

Introduction: Following an increased use of labor induction procedure to prevent adverse maternal and fetal outcomes in Sub-Saharan Africa, hitting the best algorithm that accurately classify subjects in need of the intervention is of paramount importance. This study aimed at comparing the potential benefits of applying machine learning (ML) algorithms over the conventional logistic regression model in predicting the use of labor induction intervention in pregnant women attending one of the tertiary hospitals in north Tanzania for delivery.

Methods: We conducted a secondary data analysis of the Kilimanjaro Christian Medical Centre (KCMC) birth registry database for women with uncomplicated pregnancies from the year 2000 to 2015. We excluded observations with non-vertex presentation and those with

missing information on labor induction status. Model accuracy and Area under the receiver operating characteristic curve (AUC - ROC) were used to assess the discriminative ability of the selected models. We plotted the decision curve analysis (DCA) to assess the clinical utility of the models under observation.

Results: A total of 21,578 deliveries were analyzed. Among these, 8814 (41%) were induced during the study period. Among the selected machine learning models, Random forest algorithm exhibited the best performance in terms of accuracy [0.75; 95%CI (0.73 – 0.76)] and AUC-ROC [AUC-ROC: 0.75; 95% CI (0.74 – 0.76)] compared to other models including logistic regression. Among assessed maternal attributes, parity, maternal age, body mass index, gestational age and birthweight were deemed most important predictors for labor induction intervention.

Conclusion: The selected machine learning methods offered better computational performance compared to the conventional logistic regression model in predicting the use of labor induction intervention. The current study lends substantial support to the use of machine learning models in predicting the use of labor induction intervention.

Keywords: Machine learning models, predictors, labor induction, north-Tanzania

Background

Labor induction (IOL) is one among the most frequent obstetric intervention which include artificial stimulation of uterine contraction before its spontaneous onset [1 - 3]. The procedure can be achieved either by mechanical means or by the use of pharmaceuticals readily available in the market [4]. Mechanical induction methods include amniotomy, membrane sweeping and the use of balloon catheter [5 - 6]. Major predictors of IOL procedure may be categorized as maternal, fetal, social or combination of all or some of these factors [7]. Given the increasing attention to reduce perinatal morbidity and mortality, the rates of IOL have continued to rise over the past few decades [8]. Globally, the prevalence of IOL varies greatly between countries and regions but higher rates have been more reported in developed countries than in developing countries [9 - 10]. The IOL accounts for approximately 20% of deliveries in the UK and USA, but the rates have been rising steadily since the past decade [11, 12]. The rate of IOL for Africa region currently stands at 4.4%, confirming the lowest rates for this important intervention in this region [13]. The low rates of IOL in Africa may be reflecting the existing high perinatal and maternal mortality rates in the region whereby 94% of all pregnancy-related deaths worldwide are recorded in this region [14].

Major indications for IOL comprises of postdates, intra uterine growth restrictions (IUGR), fetal macrosomia, oligohydramnios, isoimmunization, gestational diabetes, chorioamnionitis, pre-labor rupture of membranes (PROM) and hypertensive disorders of pregnancy to mention a few [15 - 17]. However, the World Health Organization recommends IOL as a therapeutic option only when the benefits of pregnancy termination surpass the risks of its continuation [18]. Early identification of subject in need of the pregnancy intervention has been shown to positively affect the pregnancy outcomes for the mother and the newborn [19]. While there is increasing clinical and administrative interest in predicting the use of IOL intervention, there is no any study that has made use of machine learning algorithm to model this event to-date [20]. Machine learning is a method of data analysis that

automates analytical model building. It is a branch of artificial intelligence based on the idea that systems can learn from data, identify patterns and make decisions with minimal human intervention[58]. We thought that machine learning (ML) approach in healthcare data is important due to its wide range of applications and its incredible ability to adapt and provide solutions to complex problems efficiently, effectively, and quickly [21 – 23]. However, myriad of studies has shown similar capabilities of logistic regression and machine learning (ML) models in terms of their performance but reported loss of interpretability in some of ML models [24, 25]. Selecting the best classifier algorithm to execute for a particular healthcare intervention based on performance still remains an ad-hoc process using fundamental benchmarks such as overall loss function and misclassification metrics. In this article, we compare the performance of ML algorithms (Boosting, Random Forest, Bagging, Artificial Neural Network and Naïve Bayes) with that of logistic regression in predicting the utilization of IOL intervention based on social and clinical attributes. Using efficient algorithm to predict IOL intervention may be useful to different domains such as risk management, tailored health communications, decision support system and personalized medicine. This may also play a significant role in reshaping and optimizing the current policy governing IOL practices by channeling specific intervention for individuals who will be most likely in need of the intervention.

METHODS

Study Setting, data source and study design.

The current study was conducted at the Kilimanjaro Christian Medical Centre (KCMC), one of the four (4) tertiary hospitals in Tanzania. This facility is located in Moshi urban district, in northern part of Tanzania, serving the residents of Kilimanjaro as well as the nearby regions. The hospital records more than 2000 deliveries annually, 20% of these admissions are referral cases while the remaining

proportion comprises of self-referrals. Since the year 2000, the facility has been recording information on pregnancy, delivery as well as information of the new-born in a specific database. Specially trained nurses conduct personal interviews on daily basis after each uncomplicated delivery or within three days in case of complicated ones. Interviews are conducted using the standardized questionnaire. Records from the hospital birth registry database covers socio-demographic information as well as mother's health status before and after delivery. Socio demographic information in this database include maternal age, occupation, education level, marital status, tribe, religion and maternal residence. Coverage on clinical information include parity status, use of labor induction, referral status, indications for IOL, induction methods used, history of previous pregnancies, and others. More details on the KCMC medical birth registry procedures has been described elsewhere [26]. We excluded observations with missing information on IOL status. Deliveries with missing information on the covariates under study or with data inconsistencies (example; age below 14 years) were also excluded (**Figure 1**). We remained with 21,578 deliveries that constituted to our final sample size.

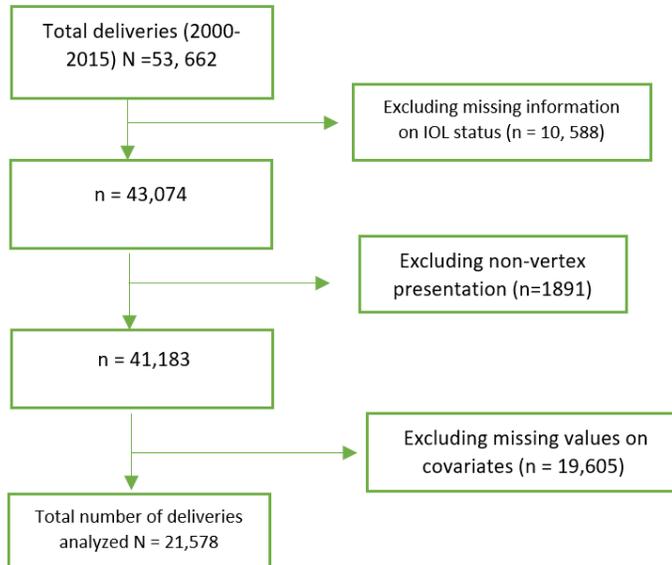


Figure 1: Schematic diagram for sample size determination

Study Outcome

The primary outcome of the study was defined as either the pregnancy was intervened by labor induction process by any means (mechanically or by the use of pharmaceuticals) or it achieved spontaneously. Instantaneous and accurate prediction of subject requiring the IOL intervention enables clinicians not only to allocate the IOL resources but also to timely intervene on high-risk patients.

Statistical analyses

Data analysis was performed using R package Version 4.0.3. Mean and standard deviation were used to summarize the continuous variables while categorical variables were summarized using frequency and percentages. Pearson chi-squared test was used to determine association between a set of independent variables and IOL status. We compared the performance of logistic regression (Lreg),

random forest (RF), naïve bayes (NB), Artificial neural networks (ANN), Boosting and Bagging algorithms in predicting the use of IOL.

RF is an ensemble technique that has been shown to have the best predictive performance and well suited for medium and large datasets. It is the tree-based learning algorithm which aggregates multiple generated decision trees formed on bootstrap samples aiming at reducing model overfitting while improving its accuracy [27]. We also used this algorithm to identify important variable for IOL intervention [28]. Prior to building ML models, we split data into “training” and “testing” datasets in 70%: 30% ratio. Randomization was performed prior to splitting so as to avoid possibility of data dependency. We estimated Out-of-Bag error (OOB) (tested against training data subsets that are not included in sub-tree construction) and validation error (tested against the test data) to come up with the best possible predictive model [29]. We used the “*RandomForest*” package to implement this model in R.

Naïve Bayes (NB) is a powerful classifier algorithm belonging to a family of simple probabilistic classifiers based on Bayes’ theorem with strong independence as well as equal-importance assumptions among the features under observation [30, 31]. It is somehow computationally expensive but can handle an enormous amount of high dimensional data. Even though the naïve assumptions are rarely true, the algorithm performs surprisingly good in many cases [32, 33]. We used “*naiveBayes*” function in R-package to fit NB models.

Artificial Neural Network (ANN) is a computational model inspired by the biological neural networks aiming at simulating the functioning of the human brain [34, 35]. The algorithms learn from inputs, hidden and output layers which are interconnected to produce the desired outputs. Backpropagation is a set of learning rules that are used to guide these complex networks. The input units receive information based on the internal weighting system and the neural network attempts to learn about

them and eventually produce the desired results. This algorithm can learn by themselves and produce the output that is not limited to the input provided to them. Another fascinating feature is that, even if a neuron is not responding or a piece of information is missing, the network can detect the fault and still produce the output [36, 37]. In addition, ANN can perform multiple tasks in parallel without affecting the system performance. We used “*nnet*” package to fit the ANN model in R.

Boosting is an ensemble meta-algorithm which works by combining weak learners to form a strong rule for classification by performing several iterations, a process which improves the prediction accuracy [38, 39]. These algorithms seek to improve the prediction power by training a sequence of weak models, each compensating the weaknesses of its predecessors.

Bagging or Bootstrap aggregation also uses an ensemble learning to evolve machine learning models. The bagging technique is useful for both regression and statistical classification [40]. This algorithm is used with decision trees, where it significantly raises the stability of models by the reduction of variance and it also eliminates the challenge of model overfitting. Briefly, the base algorithm reads the data and assigns equal weight to each covariate under observation. Thereafter, the incorrect predictions made by the base learner are identified. In the next iteration, the false predictions are assigned to the next base learner with a higher weightage on these incorrect predictions. This process is repeated until the algorithm can correctly classify the output [41].

Logistic regression is one of the simplest machine learning algorithms that has been used often in low dimension data for binary classification problems. This algorithm uses the sigmoid function to perform prediction task [42, 43]. We used generalized linear model found in “*glm*” package to execute logistic regression algorithm.

After training the selected models, we computed the prediction performance of each model using testing dataset (which is the 30% proportion of the main data set reserved and unseen by the model)

for validation task. The main purpose of using the testing data set is to test the generalization ability of the trained models (assessing whether the model overfits). To assess the validity and performance of the models, we used “area under the receiver operating characteristic curve” (AUC-ROC). The AUC-ROC is a performance measurement used in machine learning for classification problem that uses the true positive and false positive rate that represents the degree or measure of separability and describes how much the model is capable of distinguishing between classes [44, 45]. To compare the ROC curve between models, Delong’s test was used [63]. In these machine learning models, we applied techniques to minimize potential overfitting. These include Out-of-bag estimation, 10-fold cross validation and hold-out method.

Variable importance, Decision curve analysis (DCA) and model validation

We identified covariates which had the most predictive power compared to others. “*VarImp*” function was used to execute this process. This algorithm adds randomness to the data by creating shuffled copies of all features followed by training a random forest classifier on extended dataset while applying a feature importance measures to evaluate the importance of each feature [46, 47]. At every iteration, it assesses whether a real feature has a higher importance than the rest of other features. Important variables are the drivers of the outcome and their values have a significant impact in the overall model accuracy.

To compare the performance and clinical utility of these models at a given threshold probability, we plotted a decision curve analysis (DCA). This is a common framework in which a clinical judgement of the relative value of benefits and harms associated with prediction model is made. It calculates the “net-benefit” as a parameter of interest for each threshold probability [48, 49]. In other words, the DCA incorporates the information about the benefits of correctly identifying the induced deliveries (true positives) and the relative harm of incorrectly identifying the same (false positives). We therefore

presented the net benefit of each model through the range of threshold probabilities in a decision curve. Model is said to be superior to another at the chosen threshold if its net benefit surpasses the net benefit of other models for a given value of threshold probability.

RESULTS

Characteristics of study participants

During 2000 – 2015, KCMC hospital recorded about 53,662 deliveries. Of these, we excluded entries (n=10,588) that had no information on whether the labor was induced or not. We also left out pregnancies that had non-vertex presentation (n=1891) to avoid overrepresentation of the use of IOL intervention at the study site. The mean maternal age of study participants was 28 (SD = 6) years. About half of deliveries were from mothers aged between 25 – 35 years. Our study population had a good balance between nulliparous (46%) and multiparous (54%) mothers. Majority of deliveries (78%) were at term while post term deliveries constituted of about 8% and preterm accounted for 14% of all deliveries. Sociodemographic and clinical characteristics of study participants are clearly displayed in

Table 1.

Table 1: Socio-demographic characteristics of study participant (N=21,578)

Characteristics	Induced delivery n (%)	Spontaneous delivery n (%)	χ^2 p-value
Maternal age			
<25	3728 (42.3)	4304 (33.72)	<0.001
25-35	4162 (47.22)	6122 (47.96)	
>35	924 (10.48)	2338 (18.32)	
Maternal religion			
Muslim	3464 (39.3)	4830 (37.84)	0.03
Christian	5350 (60.7)	7934 (62.16)	
Gestational age			
Preterm	1018 (11.55)	1802 (14.12)	<0.001
Term	7029 (79.75)	9906 (77.61)	
Post term	767 (8.70)	1056 (8.27)	
Maternal residence			
Rural	2904 (32.95)	4795 (37.57)	<0.001
Urban	5910 (67.05)	7969 (62.43)	
Parity status			
Nulliparous	6036 (68.48)	3964 (31.06)	<0.001
Multiparous	2778 (31.52)	8800 (68.94)	
Multiple gestation			
No	8466 (96.05)	12171 (95.35)	0.014
Yes	348 (3.95)	593 (4.65)	
PROM			
No	8623 (97.83)	12567 (98.46)	0.001
Yes	191 (2.17)	197 (1.54)	
Child Sex			
Male	4275 (48.5)	6243 (48.91)	0.555
Female	4539 (51.50)	6521 (51.09)	
Circumcision			
No	7810 (88.61)	10785 (84.50)	<0.001
Yes	1004 (11.39)	1979 (15.50)	
Referred for delivery.			
No	7418 (84.16)	10517 (82.4)	0.001
Yes	1396 (15.84)	2247 (17.60)	
Alcohol use during pregnancy			
No	6999 (79.41)	9578 (75.04)	<0.001
Yes	1815 (20.59)	3186 (24.96)	
Maternal occupation			
Employed	6570 (55.64)	4210 (43.09)	<0.001
Unemployed	5237 (44.36)	5561 (56.91)	
Maternal education			
None	1378 (18.97)	3547 (24.78)	<0.001
Primary	3231 (44.48)	5879 (41.07)	
Secondary	1756 (24.17)	2877 (20.10)	
Higher	899 (12.38)	2011 (14.05)	

Number of ANC visits			
≥4	4874 (62.72)	9765 (70.72)	
<4	2897 (37.28)	4042 (29.28)	<0.001
Marital status			
Married	7479 (84.85)	11459 (89.78)	
Not married	1335 (15.15)	1305 (10.22)	<0.001

Important variables for IOL intervention

We used the “VarImp” function in “RandomForest” package to obtain important/significant characteristics that plays a major role in predicting the utilization IOL intervention among the 14 variables in the main dataset. We used Mean Decrease Gini parameter to achieve the variable importance (Figure 2). Variable with high importance have strong association with the prediction results. We found that body mass index, maternal age, gestational age, parity and birthweight were the important features to consider when predicting IOL intervention.

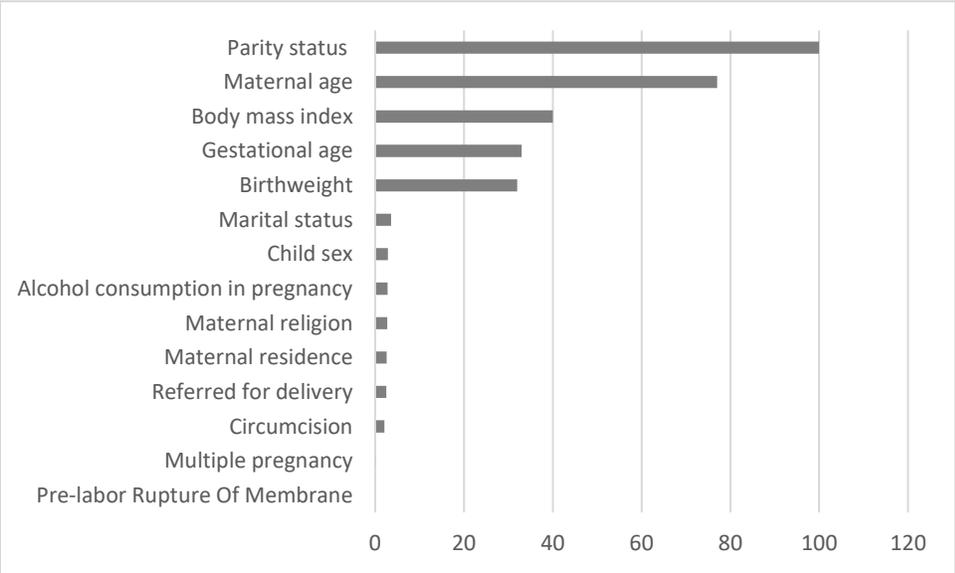


Figure 2: Variable importance measures for prediction of labor induction intervention

Predicting the use of IOL intervention

The performance of the selected ML models versus that of logistic regression can be visualized in **Figure 3** as well as **Table 2**. It shows that there is a significant difference in AUC-ROC between logistic regression model and the selected ML algorithms ($p < 0.001$). Among the ML models, Naïve Bayes (NB) was outperformed by other algorithms.

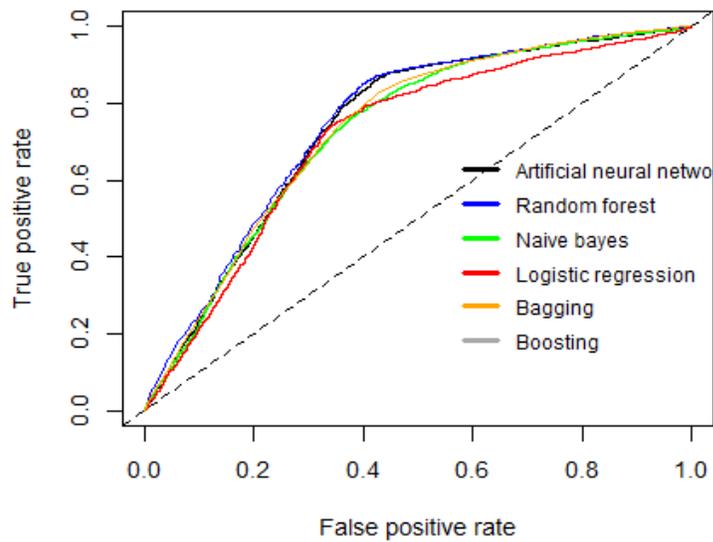


Figure 3: ROC curve for comparing the performance of ML algorithms compared to Logistic regression.

Results from the DCA (**Figure 4**) has shown that the selected ML models demonstrated high net benefit over the range of threshold probabilities when compared to that of logistic regression model.

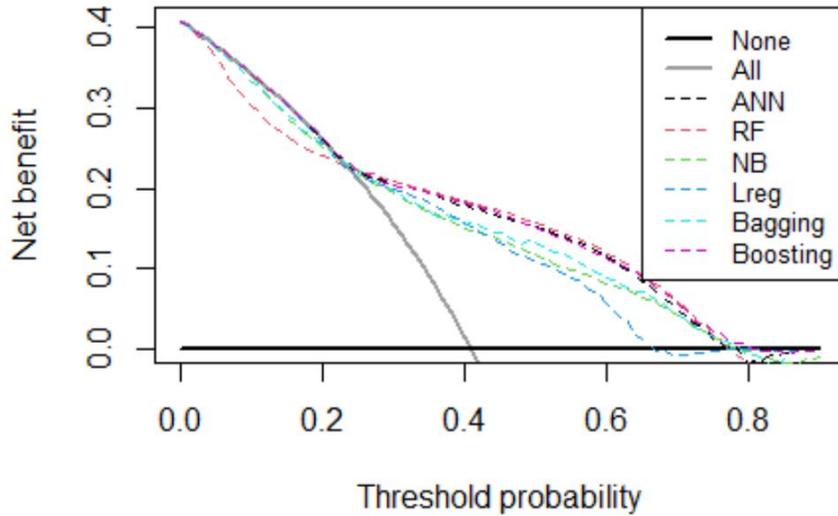


Figure 4: Comparison of machine learning models versus Logistic regression over the range of threshold probabilities.

Table 2: Prediction performance of the conventional logistic regression models as compared to machine learning models

Model	Logistic regression	Artificial Neural Network	Neural	Random Forest	Naïve Bayes	Bagging	Boosting
ACC	0.70 (0.69-0.71)	0.74 (0.73-0.75)		0.75 (0.73-0.76)	0.71 (0.70-0.72)	0.72 (0.71-0.73)	0.74 (0.73-0.75)
AUC	0.71 (0.70-0.73)	0.74 (0.73-0.75)		0.75 (0.74-0.76)	0.73 (0.72-0.74)	0.74 (0.72-0.75)	0.75 (0.73-0.76)
P-value	Reference	<0.001		<0.001	<0.001	<0.001	<0.001
Sensitivity	0.72 (0.70-0.73)	0.86 (0.84-0.87)		0.86 (0.85-0.87)	0.79 (0.78-0.80)	0.72 (0.70-0.73)	0.83 (0.81-0.84)
Specificity	0.67 (0.65-0.69)	0.58 (0.56-0.60)		0.59 (0.57-0.61)	0.59 (0.57-0.61)	0.67 (0.65-0.69)	0.57 (0.56-0.59)
PPV	0.76 (0.75-0.77)	0.75 (0.73-0.76)		0.75 (0.74-0.77)	0.74 (0.72-0.75)	0.76 (0.75-0.77)	0.74 (0.72-0.75)
NPV	0.62 (0.60-0.64)	0.74 (0.72-0.75)		0.74 (0.72-0.76)	0.66 (0.64-0.68)	0.62 (0.60-0.64)	0.69 (0.67-0.71)

DISCUSSION

The current study made use of ML models to predict the use of labor induction intervention at tertiary hospital using maternal birth registry database. We observed a significant advantage of using ML models over the conventional logistic regression in terms of sensitivity, accuracy, and area under the receiver operating characteristic curve (AUC-ROC). Few important factors may be responsible for observing much improvements in predictive performance of the ML models. These include the ability of ML algorithms in handling high-order interactions between predictors and non-linear relationships

with outcomes. Additionally, while model overfitting is still a problem in conventional models including logistic regression, ML models have adopted multiple approaches such as k-fold cross validation and hold out methods that may have played a significant role in alleviating this problem in the current study. However, logistic regression model seemed to have approximately equal sensitivity to that of Bagging classifier in this occasion. Despite the fact that bagging classifier is a powerful ensemble method which reduces variance and model overfitting, it sometimes gives its final prediction based on the mean predictions from the subset trees, rather than outputting the precise values for the classification task. This may be one of the reasons that this ML algorithm was outperformed by others. Important variables for IOL intervention were maternal age, maternal body mass index, parity, gestational age, and birthweight. In other words, these are the attributes with the most predictive power of IOL utilization. Studies elsewhere have shown a significant relationship between these attributes and the risk of obstetric interventions including IOL [50 – 55]. The decision curve analysis which portrayed the impact of false-negative and false-positive misclassification errors. From the DCA, “Net benefit” is calculated across a range of threshold probabilities, which is defined as the minimum probability of disease at which further intervention would be warranted, as net benefit = sensitivity × prevalence – (1 – specificity) × (1 – prevalence) × w where w is the odds at the threshold probability. For a prediction model that gives predicted probability of disease \hat{p} , sensitivity and specificity at a given threshold probability is calculated by defining test positive as $\hat{p} \geq p_t$. We observed that the net benefit for RF model surpassed that of all other models under investigation, which mean higher accuracy in predicting the likelihood of IOL intervention. Higher net benefits in predictive ability of ML models have been documented elsewhere [56, 57]. The main reason for the higher benefits may be explained by the fact that ML models can accommodate high order nonlinear interactions among the covariates under study, a characteristic which is not featured in conventional logistic regression models. In addition, logistic regression model works under strict distributional assumptions while ML models are

mostly non-parametric and hence robust [59, 60]. Another fascinating ML feature is their ability to efficiently handle model overfitting. Overfitting is the scenario when a model only learns the details and noise in the training data to the extent that it negatively impacts the performance of the model on the testing data [61, 62]. To our knowledge, the current study is the first study that applied and compared most popular ML algorithms to predict the use of IOL intervention in Tanzania. We enrolled a number of deliveries from an extended period of time (15 years), a sample that may have accommodated a diversified groups of study participants with contrasting characteristics. We were also able to derive the most important predictors of IOL intervention from diverse socio and clinical features that are routinely recorded in hospital setting. The DCA revealed that all ML models yielded a larger net benefit compared to that of logistic regression model. This explains the trade-off between the appropriate classification of induced and spontaneous deliveries throughout a wide range of threshold probabilities. However, our study had some limitations that should be taken into consideration during interpretations of the results. Observations with missing values in both the outcome and predictors were excluded from our analyses, a situation which may mean underestimation or overestimation of proportion of induced and spontaneous deliveries across the study period. A prospectively designed study (which we call upon) may handle this challenge. There is also a risk of selection bias since the study recruited participants who were only admitted at this facility and hence it may not be generalizable to other localities in Tanzania. Finally, we assessed the validity of our models to ensure that they are not overfitting the training dataset by testing on the test data as well as performing cross-validation.

Conclusion

Prediction of labor induction intervention based on sociodemographic and clinical information may be of limited value when relying on logistic regression models. ML methods has proved to have an extended net benefit over the range of threshold probabilities, indicating an extended clinical utility over that of logistic regression. Therefore, ML models offers new approach and direction for enhancing clinician's judgement for the sake of improving as well as optimizing utilization of resources in a resource-constrained setting.

Declarations

Ethics approval and consent to participate.

This study sought and was granted an ethical approval from Kilimanjaro Christian Medical University College Research and Ethics Committee with reference number **985**. The registry project obtained informed verbal consents from the study subjects during development of the medical registry database and was approved by the Ministry of Health of Tanzania and the National Ethics Committee in Norway prior to its commencement. The midwife nurse gave every woman oral information about the birth registry, the data needed to be collected from them and the use of the data for research purposes. Following the consent, the woman could still opt not to reply to individual questions. All consent procedures were approved by the Kilimanjaro Christian Medical Centre ethical committee and the administrative permission to access the data was provided by the KCMC hospital. Furthermore, confidentiality and privacy were assured as per the protocol of the birth registry. Patients' names were coded by the unique hospital registration numbers to ensure anonymity. We declare that all methods adopted in this research were carried out in accordance with the guideline and regulations for involving human participants.

Consent for publication

Not Applicable

Availability of data and materials

The data used/or analyzed during the current study is available from the corresponding author on a reasonable request.

Conflict of Interests

The authors declare to have no competing interests.

Funding

Research on CDC-Hospital-Community Trinity Coordinated Prevention and Control System for Major Infectious Diseases ", Zhengzhou University 2020 Key Project of Discipline Construction, XKZDQY202007. We declare that the funder had no influence on the study design, collection, analysis, and interpretation of data and on writing the manuscript.

Abbreviations

ACC: Accuracy; AUC: Area under curve; AUC-ROC: Area under the receiver-operating-characteristics curve; ANN: Artificial neural network; DCA: Decision curve analysis; IOL:

Induction of labor; KCMC: Kilimanjaro Christian Medical Center; ML: Machine learning; NB: Naïve Bayes; NPV: Negative predictive value; PPV: Positive predictive value; RF: Random forest.

Acknowledgements

We would like to thank the staff of the Birth Registry, Department of Obstetrics & Gynecology of the Kilimanjaro Christian Medical Centre and the Department of Epidemiology and Applied Biostatistics of the Kilimanjaro Christian Medical University College for their substantial support during this study. Special thanks to women who participated in the KCMC birth registry study and the Norwegian birth registry for partnering with us in providing the limited dataset used for this study. Finally, we would like to acknowledge Mr. Innocent B. Mboya (Lecturer at Kilimanjaro Christian Medical University College (KCMUCo) - Tanzania) for the motivation he provided on machine learning techniques.

Author Contributions

CST conceived the idea, designed the study, analyzed and drafted the manuscript. SB co-supervised the project and provided technical assistance on global health perspective. QL reviewed the final manuscript. MJM data management and retrieval, read and approved the final manuscript. JW supervised the project and reviewed the final manuscript. All authors read and approved the final version of the manuscript.

References

1. American College of Obstetricians and Gynecologists (ACOG practice bulletin). Induction of labor. *Obstet Gynecol.* 2009; 114:386–397.
2. MacKenzie JZ. Induction of labor at the start of the new millennium. *Reproduction.* 2006; 131:989–998.
3. Sanchez-Ramos L. Induction of labor. *Obstet Gynecol Clin N Am.* 2005; 32: 181–200.
4. Mozurkewich EL, Chilimigras JL, Berman DR, Perni UC, Romero VC, King VJ *et al.* Methods of induction of labour: a systematic review. *BMC Pregnancy Childbirth.* 2011; **11**, 84.
5. Heinemann J, Gillen G, Sanchez-Ramos L, Kaunitz AM: Do mechanical methods of cervical ripening increase infectious morbidity? A systematic review. *American Journal of Obstetrics & Gynecology.* 2008; 199(2):177-187.
6. Vaknin Z, Kurzweil Y, Sherman D. Foley catheter balloon vs locally applied prostaglandins for cervical ripening and labor induction: a systematic review and meta-analysis. *American journal of obstetrics and gynecology.* 2010; 203(5):418-429
7. Lueth GD, Kebede A, Medhanyie AA. Prevalence, outcomes and associated factors of labor induction among women delivered at public hospitals of MEKELLE town: A hospital based cross sectional study. *BMC Pregnancy Childbirth.* 2020; 20: 203.
8. Robson S, Pridmore B, Dodd J. Outcomes of induced labour. *Aust N Z J Obstet Gynaecol.* 1997;37:16–9.
9. Lydon-Rochelle MT, Ca'rdenas V, Nelson JC, Holt VL, Gardella C, Easterling TR. Induction of labor in the absence of standard medical indications: Incidence and correlates. *Med Care.* 2007;45:505–12.
10. Coonrod DV, Bay RC, Kishi GY. The epidemiology of labor induction: Arizona, 1997. *Am J Obstet Gynecol.* 2000; 182:1355–62.
11. Zhang J, Yancey MK, Henderson CE. National trends in labor induction 1989-1998. *J Reprod Med.* 2002; 47: 120 – 124.
12. Laughon SK, Zhang J, Grewal J, Sundaram R, Beaver J, Reddy UM. Induction of labor in a contemporary obstetric cohort. *Am J Obstet Gynecol.* 2012; 206: 486-489.

13. Fawole B, Nafiou I, Machoki M, Wolomy-Molondo J, Mugerwa K, Neves I et al. Unmet need for Induction of Labor in Africa: Secondary analyses from the 2004 – 2005 WHO Global Maternal and Perinatal Health Survey (A cross sectional survey). *BMC Public Health*. 2012; 12:722.
14. World Health Organization. Neonatal and perinatal mortality: country, regional and global estimates. Available from: <https://apps.who.int/iris/handle/10665/43800> . Accessed 30 December 2020.
15. Thomas J, Kavanagh J, Anthony K. RCOG Evidence-based Clinical Guidelines Induction of labour. 2001.
16. Yeast JD, Jones A, Poskin M. Induction of labor and the relationship to cesarean delivery: a review of 7001 consecutive inductions. *Am J Obstet Gynecol*. 1999;180:628–33.
17. Sanchez-Ramos L. Induction of labour. *Obstet Gynecol Clin North Am*. 2005;32:181–200.
18. The World Health Organization. Recommendations for induction of labor. Available from: https://www.who.int/reproductivehealth/publications/maternal_perinatal_health/9789241501156/en/. Accessed 30 April 2020.
19. Lassi, Z.S., Mansoor, T., Salam, R.A. et al. Essential pre-pregnancy and pregnancy interventions for improved maternal, newborn and child health. *Reprod Health* 11, S2 (2014). <https://doi.org/10.1186/1742-4755-11-S1-S2>
20. Iftikhar P, Kuijpers MV, Khayyat A, Iftikhar A, DeGouvia De Sa M. Artificial Intelligence: A New Paradigm in Obstetrics and Gynecology Research and Clinical Practice. *Cureus*. 2020 Feb 28;12(2):e7124. doi: 10.7759/cureus.7124. PMID: 32257670; PMCID: PMC7105008.
21. McCoy LG, Banja JD, Ghassemi M, et al. Ensuring machine learning for healthcare works for all *BMJ Health & Care Informatics* 2020;27:e100237. doi: 10.1136/bmjhci-2020-100237
22. Davenport T, Kalakota R. The potential for artificial intelligence in healthcare. *Future Healthc J*. 2019 Jun;6(2):94-98. doi: 10.7861/futurehosp.6-2-94. PMID: 31363513; PMCID: PMC6616181.
23. Benjamin A Goldstein, Ann Marie Navar, Michael J Pencina, John P A Ioannidis, Opportunities and challenges in developing risk prediction models with electronic health records data: a systematic

- review, *Journal of the American Medical Informatics Association*, Volume 24, Issue 1, January 2017, Pages 198–208, <https://doi.org/10.1093/jamia/ocw042>
24. Lynam, A.L., Dennis, J.M., Owen, K.R. *et al.* Logistic regression has similar performance to optimised machine learning algorithms in a clinical setting: application to the discrimination between type 1 and type 2 diabetes in young adults. *Diagn Progn Res* 4, 6 (2020). <https://doi.org/10.1186/s41512-020-00075-2>
 25. Shillan D, Sterne JAC, Champneys A, Gibbison B. Use of machine learning to analyse routinely collected intensive care unit data: a systematic review. *Crit Care*. 2019;23(1):284.
 26. Mahande JM, Daltveit AK, Mmbaga BT, Obure J, Masenga G, Manongi R, *et al.* Recurrence of perinatal death in northern Tanzania: a registry-based study. *BMC Pregnancy Childbirth*. 2013; 13:166.
 27. Jiang, R., Tang, W., Wu, X. *et al.* A random forest approach to the detection of epistatic interactions in case-control studies. *BMC Bioinformatics* 10, S65 (2009). <https://doi.org/10.1186/1471-2105-10-S1-S65>
 28. Strobl, C., Boulesteix, AL., Zeileis, A. *et al.* Bias in random forest variable importance measures: Illustrations, sources and a solution. *BMC Bioinformatics* 8, 25 (2007). <https://doi.org/10.1186/1471-2105-8-25>
 29. Song, L., Langfelder, P. & Horvath, S. Random generalized linear model: a highly accurate and interpretable ensemble predictor. *BMC Bioinformatics* 14, 5 (2013). <https://doi.org/10.1186/1471-2105-14-5>
 30. Langarizadeh M, Moghbeli F. Applying Naive Bayesian Networks to Disease Prediction: a Systematic Review. *Acta Inform Med*. 2016 Oct;24(5):364-369. doi: 10.5455/aim.2016.24.364-369. Epub 2016 Nov 1. PMID: 28077895; PMCID: PMC5203736.
 31. Sharma RK, Sugumaran V, Kumar H, Amarnath M. A comparative study of naïve Bayes classifier and Bayes net classifier for fault diagnosis of roller bearing using sound signal. *International Journal of Decision Support Systems*. 2015 Jan 1;1(1):115–29.

32. Wolfson J, Bandyopadhyay S, Elidrissi M, Vazquez-Benitez G, Vock DM, Musgrove D, et al. A Naive Bayes machine learning approach to risk prediction using censored, time-to-event data. *Statist Med.* 2015 Sep 20;34(21):2941–57.
33. Kazmierska J, Malicki J. Application of the Naïve Bayesian Classifier to optimize treatment decisions. *Radiother Oncol.* 2008 Feb;86(2):211–6.
34. Shahid N, Rappon T, Berta W (2019) Applications of artificial neural networks in health care organizational decision-making: A scoping review. *PLoS ONE* 14(2): e0212356. <https://doi.org/10.1371/journal.pone.0212356>
35. Lisboa PJ, Taktak AFG. The use of artificial neural networks in decision support in cancer: a systematic review. *Neural Netw.* 2006;19:408–15.
36. Kawauchi, K., Furuya, S., Hirata, K. et al. A convolutional neural network-based system to classify patients using FDG PET/CT examinations. *BMC Cancer* 20, 227 (2020). <https://doi.org/10.1186/s12885-020-6694-x>
37. Almeida, P.P., Cardoso, C.P. & de Freitas, L.M. PDAC-ANN: an artificial neural network to predict pancreatic ductal adenocarcinoma based on gene expression. *BMC Cancer* 20, 82 (2020). <https://doi.org/10.1186/s12885-020-6533-0>
38. Komori, O., Eguchi, S. A boosting method for maximizing the partial area under the ROC curve. *BMC Bioinformatics* 11, 314 (2010). <https://doi.org/10.1186/1471-2105-11-314>
39. Freund Y, Schapire RE: A decision-theoretic generalization of on-line learning and an application to boosting. *Journal of computer and system sciences* 1997, 55: 119–139. 10.1006/jcss.1997.1504
40. Zhang, H., Song, Y., Jiang, B., Chen, B., Shan, G. Two-Stage Bagging Pruning for Reducing the Ensemble Size and Improving the Classification Performance. *Hindawi Mathematical Problems in Engineering* Volume 2019, Article ID 8906034, 17 pages <https://doi.org/10.1155/2019/8906034>

41. Datta, S., Pihur, V. & Datta, S. An adaptive optimal ensemble classifier via bagging and rank aggregation with applications to high dimensional data. *BMC Bioinformatics* 11, 427 (2010). <https://doi.org/10.1186/1471-2105-11-427>
42. Choi, S.H., Labadorf, A.T., Myers, R.H. et al. Evaluation of logistic regression models and effect of covariates for case-control study in RNA-Seq analysis. *BMC Bioinformatics* 18, 91 (2017). <https://doi.org/10.1186/s12859-017-1498-y>
43. Dreiseitl S, Ohno-Machado L. Logistic regression and artificial neural network classification models: a methodology review. *J Biomed Inform* 2002;35:352-9.
44. Florkowski CM. Sensitivity, Specificity, Receiver-Operating Characteristic (ROC) Curves and Likelihood Ratios: Communicating the Performance of Diagnostic Tests. *Clin Biochem Rev.* 2008 Aug;29(Suppl 1):S83-7.
45. Zweig MH, Campbell G: Receiver-operating characteristic (ROC) plots: a fundamental evaluation tool in clinical medicine. *Clinical Chemistry* 1993, 39(4):561-577.
46. Strobl, C., Boulesteix, AL., Kneib, T. et al. Conditional variable importance for random forests. *BMC Bioinformatics* 9, 307 (2008). <https://doi.org/10.1186/1471-2105-9-307>
47. Strobl C, Zeileis A: Danger: High Power! – Exploring the Statistical Properties of a Test for Random Forest Variable Importance. *Proceedings of the 18th International Conference on Computational Statistics, Porto, Portugal 2008.*
48. Vickers AJ, Elkin EB. Decision curve analysis: a novel method for evaluating prediction models. *Med Decis Making.* 2006 Nov-Dec;26(6):565-74. doi: 10.1177/0272989X06295361. PMID: 17099194; PMCID: PMC2577036.
49. Vickers AJ, Elkin EB. Decision Curve Analysis: A Novel Method for Evaluating Prediction Models. *Medical Decision Making.* 2006;26(6):565-574. doi:10.1177/0272989X06295361

50. Hsieh TT, Liou JD, Hsu JJ, Lo LM, Chen SF, Hung TH. Advanced maternal age and adverse perinatal outcomes in an Asian population. *Eur J Obstet Gynecol Reprod Biol* 2010;148:21–6.
51. Ahveci B, Melekoglu R, Evruke IC, Cetin C. The effect of advanced maternal age on perinatal outcomes in nulliparous singleton pregnancies. *BMC Pregnancy Childbirth*. 2018;18(1):343. Published 2018 Aug 22. doi:10.1186/s12884-018-1984-x
52. Yogev Y, Melamed N, Bardin R, Tenenbaum-Gavish K, Ben-Shitrit G, Ben-Haroush A. Pregnancy outcome at extremely advanced maternal age. *Am J Obstet Gynecol*. 2010;203:558:1–7.
53. Dammer U, Bogner R, Weiss C, Faschingbauer F, Pretscher J, Beckmann MW, Sütterlin M, Kehl S. Influence of body mass index on induction of labor: A historical cohort study. *J Obstet Gynaecol Res*. 2018 Apr;44(4):697-707. doi: 10.1111/jog.13561. Epub 2018 Jan 8. PMID: 29316054.
54. Teal, Elizabeth Nicole MD, MPH; Lewkowitz, Adam MD; Koser, Sarah BS; Tran, Carol BS; Gaw, Stephanie MD, PhD Relationship Between Maternal BMI and Labor Induction Outcomes [14T], *Obstetrics & Gynecology*: May 2019 - Volume 133 - Issue - p 216S doi: 10.1097/01.AOG.0000559116.12776.c2
55. Athukorala C, Rumbold AR, Willson KJ, Crowther CA. The risk of adverse pregnancy outcomes in women who are overweight or obese. *BMC Pregnancy Childbirth* 2010; 10: 56
56. Raita Y, Goto T, Faridi MK, et al. Emergency department triage prediction of clinical outcomes using machine learning models. *Crit Care* 2019;23:64.
57. Goto T, Camargo CA, Faridi MK, et al. Machine Learning-Based prediction of clinical outcomes for children during emergency department triage. *JAMA Netw Open* 2019;2:e186937.
58. Kuhle, S., Maguire, B., Zhang, H. et al. Comparison of logistic regression with machine learning methods for the prediction of fetal growth abnormalities: a retrospective cohort study. *BMC Pregnancy Childbirth* 18, 333 (2018). <https://doi.org/10.1186/s12884-018-1971-2>

59. Levy, J.J., O'Malley, A.J. Don't dismiss logistic regression: the case for sensible extraction of interactions in the era of machine learning. *BMC Med Res Methodol* **20**, 171 (2020). <https://doi.org/10.1186/s12874-020-01046-3>
60. Kupek, E. Beyond logistic regression: structural equations modelling for binary variables and its application to investigating unobserved confounders. *BMC Med Res Methodol* **6**, 13 (2006). <https://doi.org/10.1186/1471-2288-6-13>
61. Frank E, Hall M, Trigg L, Holmes G, Witten IH. Data mining in bioinformatics using Weka. *Bioinformatics*. 2004;20(15):2479–81
62. Gómez D, Rojas A. An empirical overview of the no free lunch theorem and its effect on real-world machine learning classification. *Neural Comput*. 2016;28(1):216–28.
63. Heaman M, Kingston D, Chalmers B, Sauve R, Lee L, Young D. Risk factors for preterm birth and small-for-gestational-age births among Canadian women. *Paediatr Perinat Epidemiol*. 2013;27:54–61.

Figures

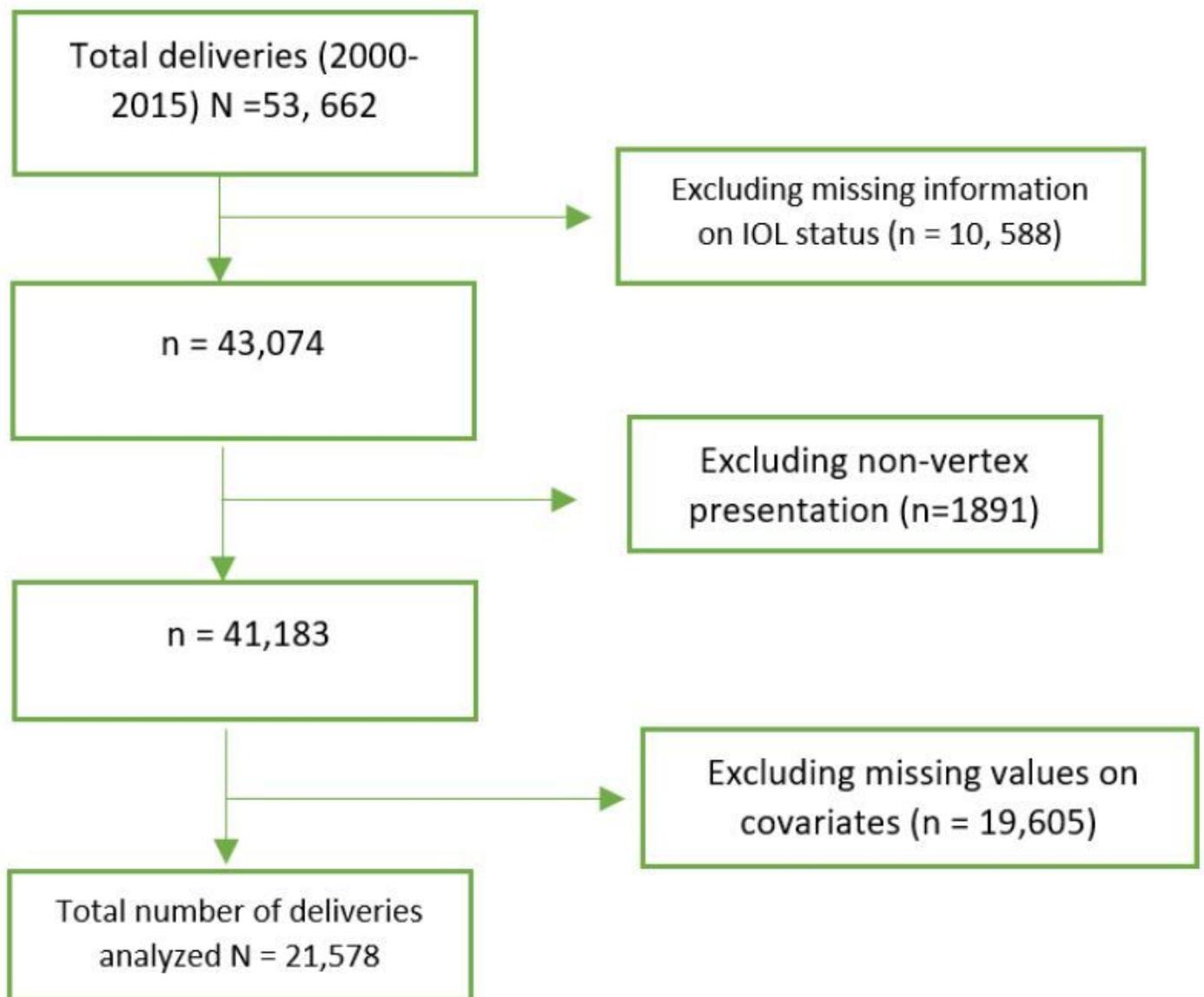


Figure 1

Schematic diagram for sample size determination

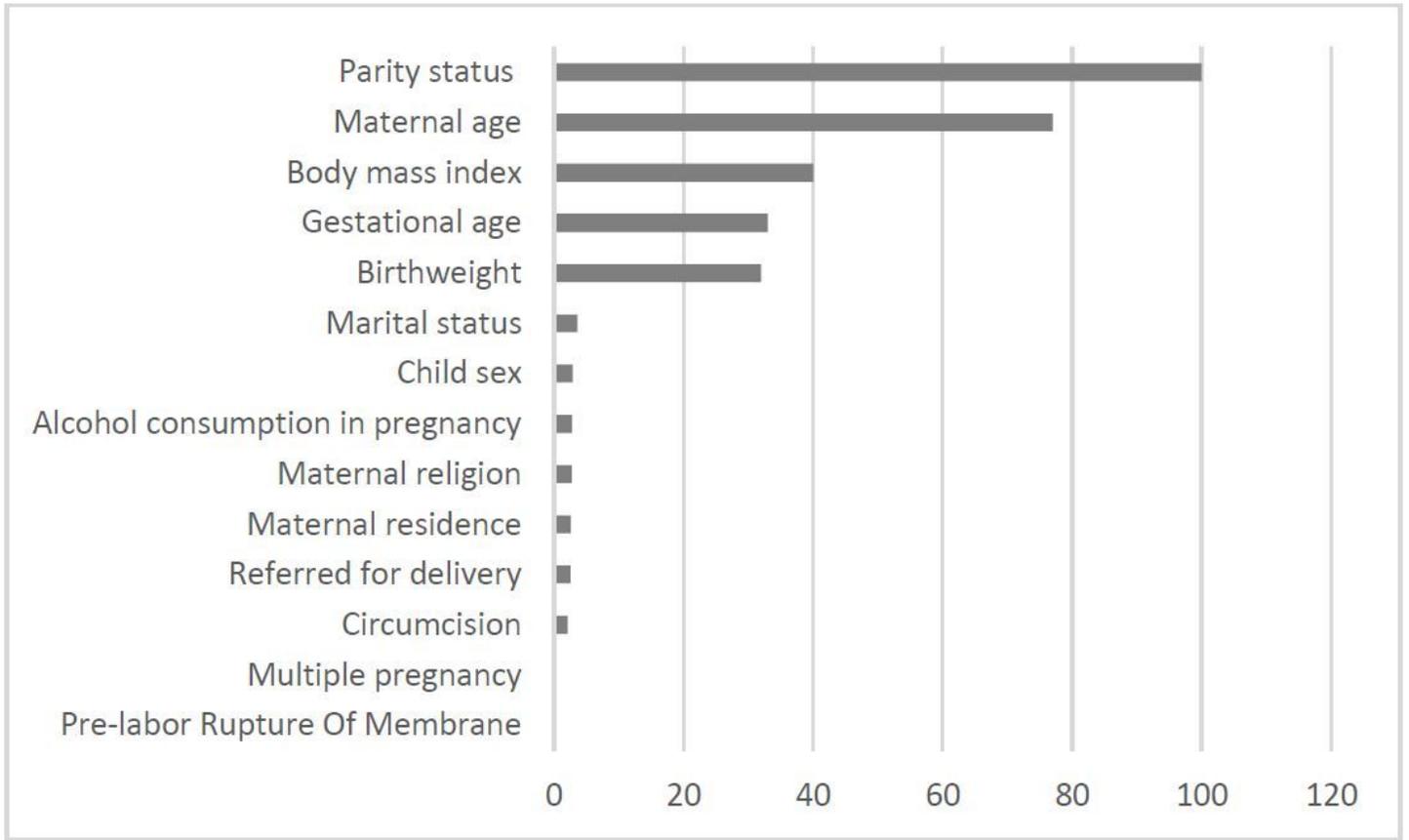


Figure 2

Variable importance measures for prediction of labor induction intervention

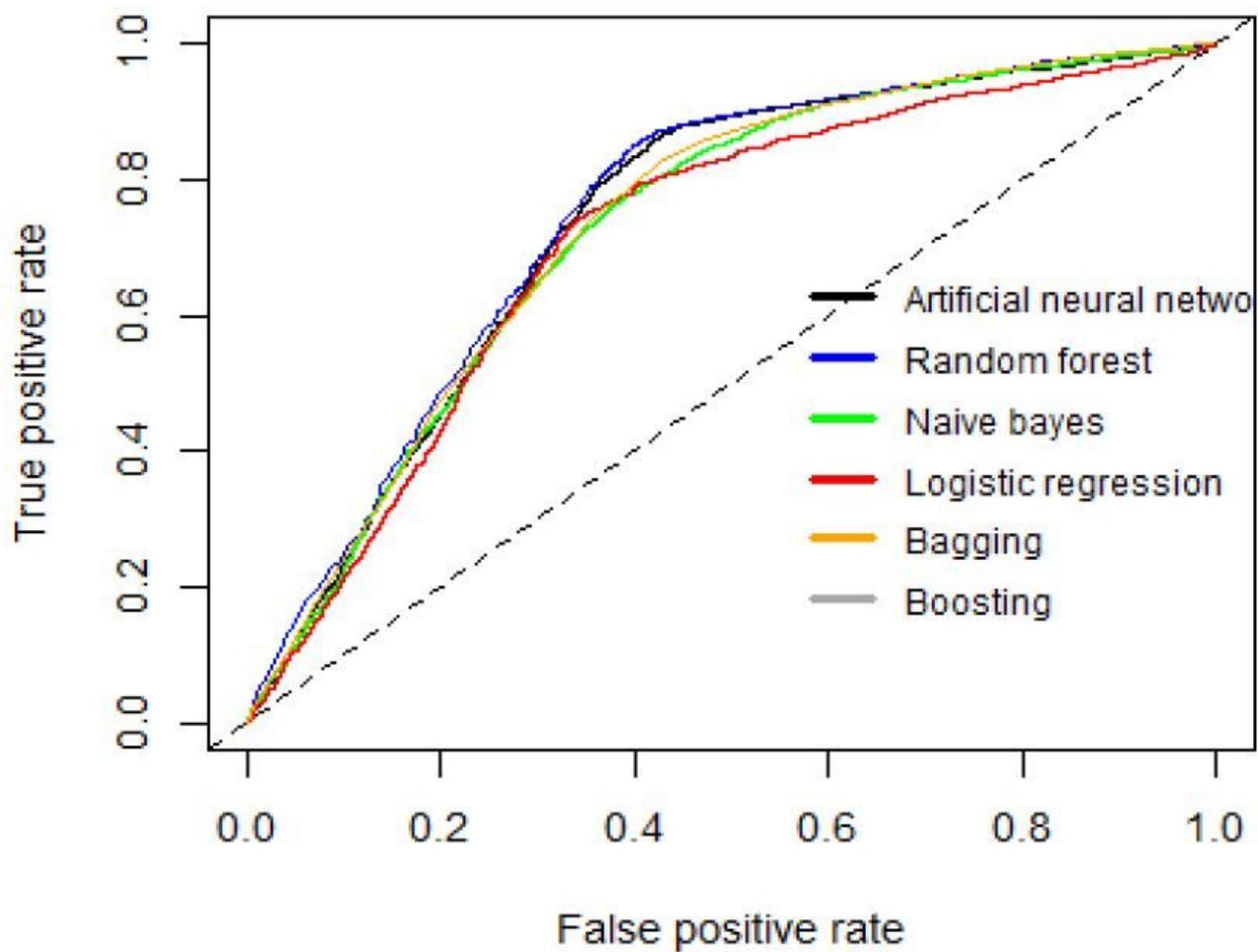


Figure 3

ROC curve for comparing the performance of ML algorithms compared to Logistic regression.

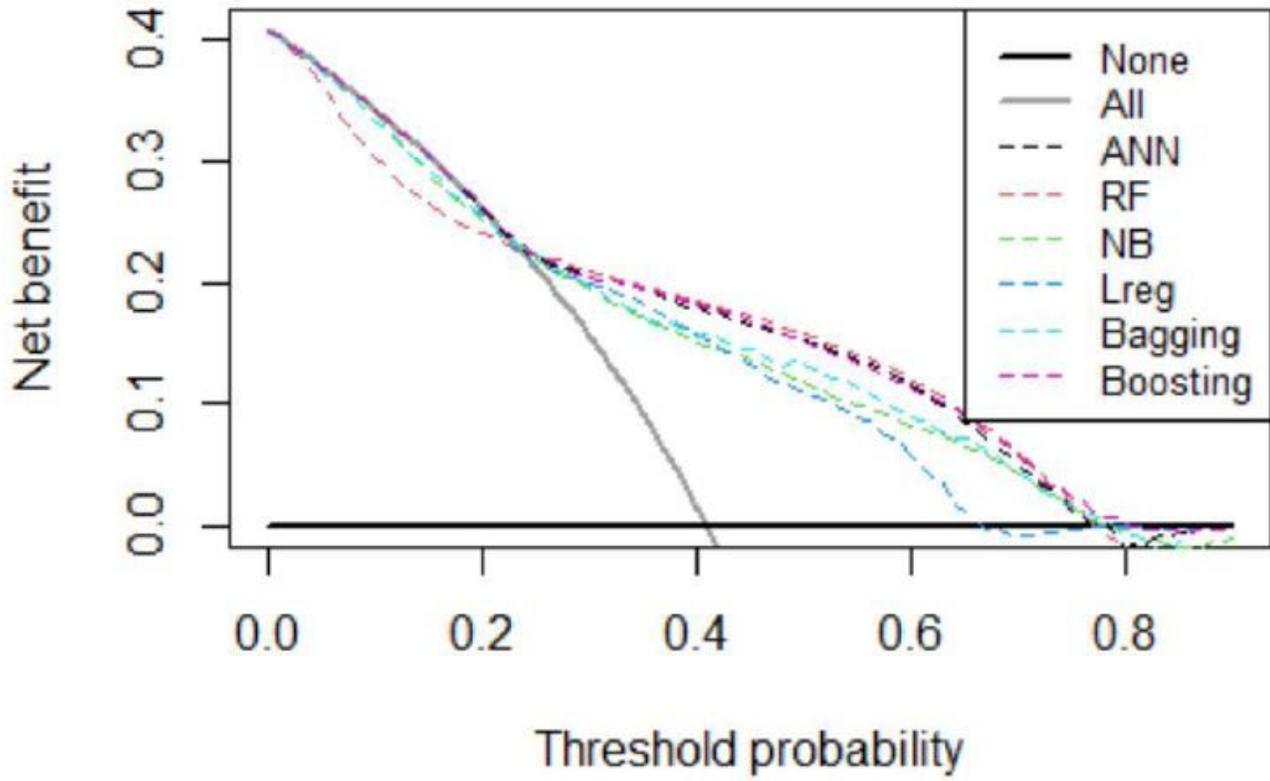


Figure 4

Comparison of machine learning models versus Logistic regression over the range of threshold probabilities.