Prediction of perinatal death using machine learning models: a birth registry-based cohort study in northern Tanzania

Innocent B Mboya, Michael J Mahande, Mohanad Mohammed, Joseph Obure, Henry G Mwambi

ABSTRACT

Objective We aimed to determine the key predictors of perinatal deaths using machine learning models compared with the logistic regression model.

Design A secondary data analysis using the Kilimanjaro Christian Medical Centre (KCMC) Medical Birth Registry cohort from 2000 to 2015. We assessed the discriminative ability of models using the area under the receiver operating characteristics curve (AUC) and the net benefit using decision curve analysis.

Setting The KCMC is a zonal referral hospital located in Moshi Municipality, Kilimanjaro region, Northern Tanzania. The Medical Birth Registry is within the hospital grounds at the Reproductive and Child Health Centre.

Participants Singleton deliveries (n=42 319) with complete records from 2000 to 2015.

Primary outcome measures Perinatal death (composite of stillbirths and early neonatal deaths). These outcomes were only captured before mothers were discharged from the hospital.

Results The proportion of perinatal deaths was 3.7%. There were no statistically significant differences in the predictive performance of four machine learning models except for bagging, which had a significantly lower performance (AUC 0.76, 95% CI 0.74 to 0.79, p=0.006) compared with the logistic regression model (AUC 0.78, 95% CI 0.76 to 0.81). However, in the decision curve analysis, the machine learning models had a higher net benefit (ie, the correct classification of perinatal deaths considering a trade-off between false-negatives and false-positives)—over the logistic regression model across a range of threshold probability values.

Conclusions In this cohort, there was no significant difference in the prediction of perinatal deaths between machine learning and logistic regression models, except for bagging. The machine learning models had a higher net benefit, as its predictive ability of perinatal death was considerably superior over the logistic regression model. The machine learning models, as demonstrated by our study, can be used to improve the prediction of perinatal deaths and triage for women at risk.

INTRODUCTION

Neonatal survival is at the heart of Sustainable Development Goals agenda. The Every Newborn Action Plan to end Preventable

Deaths set a goal for all countries to reach the target of ten or less newborn deaths per 1000 live births and 10 or less stillbirths per 1000 total births by the year 2035. Furthermore, the United Nations set the target of reducing neonatal mortality to 12 deaths per 1000 live births or fewer by 2030. Globally, neonatal deaths declined by 51% from 5 million in 1990 to 2.5 million in 2017. But this decline has not been realised in low-income and middle-income countries, which carries the highest burden of neonatal deaths, with south Asia and sub-Saharan Africa accounting for 79% of the total burden of neonatal deaths in 2017. Furthermore, the under-5 mortality rate has decreased almost across the world, but the proportions of neonatal deaths remained high in this group. Neonatal deaths accounted for 47% of all under-5 deaths in 2018, and it has increased from 40% in 1990, with sub-Saharan Africa bearing the highest
burden.\textsuperscript{6} Globally 2.5 million children died in the first month of life in 2018, with approximately 7000 newborn deaths every day.\textsuperscript{6} Nearly three-quarters of these deaths occur during the first week, with about one million dying on the first day and close to one million dying within the next 6 days.\textsuperscript{6}

Globally, more than five million perinatal deaths occur each year.\textsuperscript{3} The majority (95\%) of these deaths occur in sub-Saharan Africa and Southern Asia.\textsuperscript{7} According to the Tanzania Demographic and Health Survey, the perinatal mortality rate has slightly increased from 36 to 39 deaths per 1000 live births between 2010–11 and 2015–16 survey rounds, respectively, relative to under-5 mortality.\textsuperscript{3} In addition, perinatal mortality rate in Tanzania is the highest in East Africa.\textsuperscript{7}

Early identification of pregnant women at risk for adverse maternal and perinatal outcomes during the prenatal period and timely provision of high-quality healthcare services have been reported to improve maternal and newborn survival.\textsuperscript{3} Machine learning (hereafter denoted as ‘ML’) models are methodologies for developing algorithms that learn from existing data to make predictions on new data.\textsuperscript{9} ML models have shown better predictive performance over the classical or conventional regression models,\textsuperscript{10} and they can better handle a significant number of potential predictors. However, there is conflicting evidence of the performance of these models. Previous investigators have demonstrated that, compared with the classical regression models, ML models have superior performance for early differentiation of sepsis and non-infectious systemic inflammatory response syndrome in critically ill children,\textsuperscript{11} in predicting neonatal and under-5 mortality,\textsuperscript{12-16} and critical care and hospitalisation outcomes.\textsuperscript{17-18} In contrast, other studies have shown no predictive performance benefit of the ML models in prediction of clinical outcomes.\textsuperscript{9,19}

The first step in addressing high perinatal mortality is the accurate capture and classification of the causes of those deaths across all settings.\textsuperscript{20} WHO International Classification of Diseases (ICD-10) is a standardised tool used for the classification of deaths occurring during the perinatal period: ICD-PM.\textsuperscript{2,20,21} ML models may be an essential tool in the assessment of risk factors for deaths during the perinatal period and triage pregnant women at high risk of experiencing adverse perinatal outcomes, especially in low-resourced settings where the majority of perinatal deaths occur at home.\textsuperscript{22,23} Capturing the chain of events that led to the perinatal mortality, from both the maternal and the perinatal side, informs the design and development of preventative and therapeutic measures.\textsuperscript{2}

Using data from the medical birth registry at Kilimanjaro Christian Medical Centre (KCMC) referral hospital in northern Tanzania, we aimed to determine the key predictors of perinatal death using ML models. Previous studies using the same data\textsuperscript{26-31} applied standard regression models to assess risk factors for adverse perinatal outcomes. A major weakness of conventional regression analysis, as opposed to ML models, is that many covariates are excluded based on specific model assumptions. In contrast, ML techniques which are non-parametric in nature find the most predictive groupings of factors based on their frequency and strength of association, with no particular model assumptions.\textsuperscript{32} In this study, we compared the predictive performance of the ML models with the conventional regression analysis, particularly logistic regression (Lreg).

**METHODS**

**Study design, setting and population**

We conducted a secondary analysis of birth cohort data from the KCMC referral hospital, situated in the Moshi Municipality of Kilimanjaro region, Northern Tanzania. The hospital receives deliveries from nearby communities and referral cases from other healthcare facilities inside the region and the neighbouring regions.\textsuperscript{33} The hospital has an average annual delivery rate of 4000 births.\textsuperscript{31,33,34} The study population was women who delivered singleton babies. We included 42 319 deliveries with complete records between 2000 and 2015. We excluded records with missing values on the outcome (perinatal status) and the covariates as well as pregnancies with multiple gestations to avoid over-representation of high-risk pregnancies.\textsuperscript{31}

**Data source**

We used data from the KCMC referral hospital medical birth registry between the years 2000 and 2015, which were collected among mothers who delivered at the department of obstetrics and gynaecology. More description of the KCMC medical birth registry is also available elsewhere.\textsuperscript{26,28,31,35,36} Briefly, the KCMC medical birth registry is within hospital grounds at the Reproductive and Child Health Centre. The birth registry has been in operation since the year 2000, established to serve both clinical, administrative and research purposes.\textsuperscript{35,36} Trained midwives collected data using a standardised questionnaire (within 24 hours after delivery or later in case a mother had recovered from complications), after which data are entered into a computerised database located at the birth registry. Also, additional data were abstracted from the antenatal care (ANC) cards and the hospital medical records of the mother.\textsuperscript{26}

A unique hospital identification number was assigned to each woman at first admission and used to trace her medical records at later admissions, and further to link records of successive births of the same woman.\textsuperscript{36} Data captured information on the background characteristics of mother and father, mother’s health before and during present pregnancy, information about delivery including complications, and child characteristics including their status (ie, whether dead or alive).

**Study variables**

The main outcome variable in this study was perinatal death which was defined as the number of stillbirths
Briefly, ANN is a method constructed from three layers of connected nodes: input, hidden and output. The input where each input variable appears as a node; the hidden layer contains several nodes determined during the model tuning phase. In contrast, the output layer contains several nodes equal to the number of classes to be predicted. Between these layers, there are weighted links, the hidden layer receives a sum of the multiplication of the input variables with associated weights values plus the bias. This value is entered into an activation function, such as a logistic or sigmoid function, to decide the class prediction. Outputs of the network are interpreted as class probabilities and sum to one. We used nnet package to construct the ANN model.

RF is an extension of classification and regression trees. RF performance is better compared with bagged trees because it decorrelates the trees, hence improves accuracy. Several forests of decision trees are grown using a random bootstrapped training sample. Also, instead of using all the variables/features in each tree, a random sample of variables are selected and tested at each split in each tree. The prediction is made for unobserved data by taking a majority vote of the individual trees. We used RandomForest package to construct the RF model. NB is an effective classifier due to its simplicity, exhibiting a surprisingly competitive predictive accuracy. NB uses probability theory to find the most possible sample class in a classification problem. NB has two assumptions: (1) each attribute is conditionally independent of the other attributes given the class and (2) all the attributes have an impact on the class. We used naivebayes package to construct the NB model.

Lreg is a standard multivariate classification method. It arises from the desire to model the posterior probabilities via linear functions in covariates, such that besides predicting class labels, it provides a probabilistic interpretation of this labeling. Lreg uses a sigmoid function instead of a linear function to map predictions to probabilities between 0 and 1. We used glm method to construct the Lreg model. Bagging, or bootstrap aggregation and boosting are general techniques for improving prediction rules and accuracy of the resulting predictions, by reducing the associated variance of prediction. Bagging divides the available data into many bootstrap samples and then train a separate model for each bootstrap, and then make a final prediction by averaging and voting for regression and classification, respectively. Boosting, on the other hand, is a committee-based approach that uses a weighted average of prediction from various samples. The incorrectly predicted cases from a given step are given a higher weight during the next step. Thus, it is an iterative procedure, incorporating weights, as opposed to simple averaging of predictions.

In the training set, parameter tuning and cross-validation aim to find a balance between building a model that can classify the training data effectively without overfitting to
the random fluctuations. For each ML model, we used 10-fold cross-validation as a resampling method, where the training set is divided equally into 10 parts (folds). Therefore, every nine folds are used together for training the model and the remaining one fold for testing. This training-testing process is repeated ten times. We performed feature selection using the RF algorithm. After selecting the most important features, we retained in the dataset and used them for analysis in both the training and testing data for all models. We used the Synthetic Minority Over-sampling Technique (SMOTE) method \(^{58,59}\) to address the class imbalance in the outcome (ie, the low proportion of perinatal deaths), by specifying the additional sampling to be ‘smote’ on train control parameter specifications. SMOTE is a method that produces artificial minority samples by interpolating between existing minority samples and their nearest minority neighbors. \(^{58,59}\)

Using the testing set (30% of the remaining randomly selected sample), we computed the predictive performance of the six models (including Lreg model) from the training set using the area under the receiver-operating-characteristics curve (AUC ROC). We used the ROCR package for plotting ROC curves, obtaining the AUC values and comparison of models using AUC values. We also used measures from the confusion matrix results (ie, accuracy, sensitivity, specificity, positive and negative predictive values (NPV)), and the net benefit through decision curve analysis \(^{60,61}\)—which quantifies whether a machine provides a relevant improvement in the prediction. We used epiR package to obtain confidence intervals for the performance measures and DCA package (http://www.decisioncurveanalysis.org) for decision curve analysis. We further used ggplot2 package to plot the decision curves. A good model will have a higher net benefit. \(^{60}\) We used Delong’s test to compare the ROC between models, where, a p<0.05 was considered statistically significant. The variable importance is a scaled measure with a maximum value of 100.\(^1\) The R code used for this analysis is shown in online supplemental material 1.

**Patient and public involvement**

There was no patient and public involvement.

---

**RESULTS**

**Characteristics of study participants**

The characteristics of the participants are shown in Table 1. A total of 55,003 total deliveries were recorded at the KCMC medical birth registry from 2000 to 2015. Of these, we excluded 3316 (6%) multiple gestations (to avoid over-representation of high-risk pregnancies), \(^{31}\) 49 (0.1%) records missing maternal identification numbers (hence could not be linked to child records), 791 (1.4%) records with a mismatch between the date of birth and unknown sequence (ie, singleton vs multiple births). We further excluded a total of 8528 (15.5%) observations with missing values in both the outcome (perinatal status) and covariates. We, therefore, analysed data for a total of 42,319 singleton deliveries with complete records (figure 1).

The overall proportion of perinatal death among 42,319 singleton deliveries in this study was 3.7%. The proportion of perinatal deaths among mothers aged 20–34, 35–39 and 40+ years was 3.5%, 4.7% and 5.8%, respectively. Mothers with no education (5.6%) and those with primary education level (4.4%), who resided in rural areas (4.8%), had less than four ANC visits (5.9%), and those referred for delivery (7.8%) had a higher proportion of perinatal death. Among fathers, a higher proportion of perinatal death is among those aged 35+years (4.1%), with no (9.6%) or with primary education level (4.6%) as well as those who were unemployed (5.7%), Table 1.

Furthermore, the most common obstetric care and complications in this birth cohort included induction of labour (22.7%), malaria (13.2%), preterm birth (10.8%) and LBW (10.2%). About 4% of mothers in this cohort experienced pre-eclampsia/eclampsia during pregnancy. Less than half of all children were females. The proportion of perinatal death among women who experienced induction of labour, with malaria, delivered preterm, delivered LBW baby and experienced pre-eclampsia/eclampsia during pregnancy was 4.8%, 3.8%, 14.2%, 17.6% and 12.5%, respectively. The proportion of perinatal death is almost similar among males (3.8%) compared with females (3.6%) children in this cohort (Table 1).

The trends in the proportion of perinatal deaths that occurred at KCMC between the years 2000–2015 are shown in figure 2. Overall, the proportion of perinatal deaths has slightly declined over the years by 6% (95% CI, 0.3% to 12.3%), though this decline was not statistically significant (p=0.06).

**Variable importance**

We used the RF algorithm for feature/variable selection. This model selected a total of 20 important predictors (figure 3) based on its threshold measure of importance out of the 32 variables. We used these 20 variables in all the subsequent analysis for all models in both training and testing sets.

**Predicting perinatal deaths**

The discriminatory abilities of all models for the prediction of perinatal death are in figure 4A and Table 2. There were no significant differences (p>0.05) in the AUC the ROC curve between Lreg with RF, ANN, boosting and NB. However, bagging had significantly lower predictive performance (AUC 0.76, 95% Cl 0.74 to 0.79, p=0.006) compared with the Lreg model (AUC 0.78, 95% CI 0.76 to 0.81). Furthermore, the ANN model (sensitivity 0.60, 95% CI 0.55 to 0.64) and NB model (Sensitivity 0.57, 95% CI 0.52 to 0.62) had slightly higher sensitivity compared with Lreg (sensitivity 0.56, 95% CI 0.51 to 0.60) while boosting (Specificity 0.89 95% CI 0.88 to 0.89) and RF
<table>
<thead>
<tr>
<th>Characteristics</th>
<th>Total</th>
<th>Perinatal death</th>
<th>P value*</th>
<th>Characteristics</th>
<th>Total</th>
<th>Perinatal death</th>
<th>P value*</th>
</tr>
</thead>
<tbody>
<tr>
<td>Maternal</td>
<td></td>
<td></td>
<td></td>
<td>Obstetrics</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Age (years)</td>
<td></td>
<td></td>
<td></td>
<td>Gestational diabetes</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>15–19</td>
<td>3470</td>
<td>99 (2.9)</td>
<td>&lt;0.001</td>
<td>No</td>
<td>42 288</td>
<td>1560 (3.7)</td>
<td>0.89</td>
</tr>
<tr>
<td>20–34</td>
<td>32 675</td>
<td>1158 (3.5)</td>
<td></td>
<td>Yes</td>
<td>31</td>
<td>1 (3.2)</td>
<td></td>
</tr>
<tr>
<td>35–39</td>
<td>4984</td>
<td>235 (4.7)</td>
<td></td>
<td>Diabetes</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>40+</td>
<td>1190</td>
<td>69 (5.8)</td>
<td></td>
<td>No</td>
<td>42 240</td>
<td>1553 (3.7)</td>
<td>0.002</td>
</tr>
<tr>
<td>Education level</td>
<td></td>
<td></td>
<td></td>
<td>Hypertension</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>None</td>
<td>567</td>
<td>32 (5.6)</td>
<td>&lt;0.001</td>
<td>Yes</td>
<td>79</td>
<td>8 (10.1)</td>
<td></td>
</tr>
<tr>
<td>Primary</td>
<td>23 010</td>
<td>1019 (4.4)</td>
<td></td>
<td>No</td>
<td>42 241</td>
<td>1550 (3.7)</td>
<td></td>
</tr>
<tr>
<td>Secondary</td>
<td>5275</td>
<td>159 (3.0)</td>
<td></td>
<td>Yes</td>
<td>78</td>
<td>11 (14.1)</td>
<td></td>
</tr>
<tr>
<td>Higher</td>
<td>13 467</td>
<td>351 (2.6)</td>
<td></td>
<td>Bleeding</td>
<td></td>
<td></td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>Occupation</td>
<td></td>
<td></td>
<td>0.37</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Unemployed</td>
<td>9316</td>
<td>365 (3.9)</td>
<td></td>
<td>No</td>
<td>422</td>
<td>33 (7.8)</td>
<td></td>
</tr>
<tr>
<td>Employed</td>
<td>30 061</td>
<td>1085 (3.6)</td>
<td></td>
<td>Yes</td>
<td>5573</td>
<td>209 (3.8)</td>
<td></td>
</tr>
<tr>
<td>Others</td>
<td>2942</td>
<td>111 (3.8)</td>
<td></td>
<td>No</td>
<td>41 661</td>
<td>1523 (3.7)</td>
<td></td>
</tr>
<tr>
<td>Marital status</td>
<td></td>
<td></td>
<td>0.89</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Single</td>
<td>4954</td>
<td>186 (3.8)</td>
<td></td>
<td>Malaria</td>
<td></td>
<td></td>
<td>0.79</td>
</tr>
<tr>
<td>Married</td>
<td>37 300</td>
<td>1372 (3.7)</td>
<td></td>
<td>No</td>
<td>36 746</td>
<td>1352 (3.7)</td>
<td></td>
</tr>
<tr>
<td>Widowed/divorced</td>
<td>65</td>
<td>3 (4.6)</td>
<td></td>
<td>Yes</td>
<td>5573</td>
<td>209 (3.8)</td>
<td></td>
</tr>
<tr>
<td>Area of residence</td>
<td></td>
<td></td>
<td>&lt;0.001</td>
<td></td>
<td></td>
<td></td>
<td>0.43</td>
</tr>
<tr>
<td>Urban</td>
<td>25 056</td>
<td>725 (2.9)</td>
<td></td>
<td>No</td>
<td>41 588</td>
<td>1538 (3.7)</td>
<td></td>
</tr>
<tr>
<td>Rural</td>
<td>17 263</td>
<td>836 (4.8)</td>
<td></td>
<td>Yes</td>
<td>731</td>
<td>23 (3.1)</td>
<td></td>
</tr>
<tr>
<td>Alcohol consumption during pregnancy</td>
<td></td>
<td></td>
<td>0.001</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>No</td>
<td>30 759</td>
<td>1191 (3.9)</td>
<td></td>
<td>Complications</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Yes</td>
<td>11 560</td>
<td>370 (3.2)</td>
<td></td>
<td>Pre-eclampsia/eclampsia</td>
<td></td>
<td></td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>Smoking during pregnancy</td>
<td></td>
<td></td>
<td>0.97</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Yes</td>
<td>53</td>
<td>2 (3.8)</td>
<td></td>
<td>Induction of labour</td>
<td></td>
<td></td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>No</td>
<td>42 266</td>
<td>1559 (3.7)</td>
<td></td>
<td>No</td>
<td>32 732</td>
<td>1105 (3.4)</td>
<td></td>
</tr>
<tr>
<td>No of ANC visits</td>
<td></td>
<td></td>
<td>&lt;0.001</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>≥4</td>
<td>28 742</td>
<td>760 (2.6)</td>
<td></td>
<td>PROM</td>
<td></td>
<td></td>
<td>0.006</td>
</tr>
<tr>
<td>&lt;4</td>
<td>13 577</td>
<td>801 (5.9)</td>
<td></td>
<td>No</td>
<td>41 416</td>
<td>1543 (3.7)</td>
<td></td>
</tr>
<tr>
<td>Referred for delivery</td>
<td></td>
<td></td>
<td>&lt;0.001</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>No</td>
<td>32 762</td>
<td>819 (2.5)</td>
<td></td>
<td>PPH</td>
<td></td>
<td></td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>Yes</td>
<td>9557</td>
<td>742 (7.8)</td>
<td></td>
<td>No</td>
<td>42 091</td>
<td>1516 (3.6)</td>
<td></td>
</tr>
<tr>
<td>Paternal characteristics</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Age (years)</td>
<td></td>
<td></td>
<td></td>
<td>3–4 degree tear</td>
<td></td>
<td></td>
<td>0.49</td>
</tr>
<tr>
<td>&lt;25</td>
<td>3938</td>
<td>122 (3.1)</td>
<td></td>
<td>No</td>
<td>42 305</td>
<td>1560 (3.7)</td>
<td></td>
</tr>
<tr>
<td>25–29</td>
<td>10 593</td>
<td>346 (3.3)</td>
<td></td>
<td>Yes</td>
<td>14</td>
<td>1 (7.1)</td>
<td></td>
</tr>
</tbody>
</table>

Continued
Table 1  Continued

<table>
<thead>
<tr>
<th>Characteristics</th>
<th>Total</th>
<th>Perinatal death</th>
<th>P value*</th>
<th>Characteristics</th>
<th>Total</th>
<th>Perinatal death</th>
<th>P value*</th>
</tr>
</thead>
<tbody>
<tr>
<td>Maternal</td>
<td></td>
<td></td>
<td></td>
<td>Obstetrics</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>30–34</td>
<td>12 303 (29.1)</td>
<td>457 (3.7)</td>
<td></td>
<td>Abruption placenta</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>35+</td>
<td>15 485 (36.6)</td>
<td>636 (4.1)</td>
<td></td>
<td>No</td>
<td>42 193 (99.7)</td>
<td>1490 (3.5)</td>
<td></td>
</tr>
<tr>
<td>Education level</td>
<td></td>
<td></td>
<td>&lt;0.001</td>
<td>Placenta previa</td>
<td></td>
<td></td>
<td>0.04</td>
</tr>
<tr>
<td>None</td>
<td>281 (0.7)</td>
<td>27 (9.6)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Primary</td>
<td>18 987 (44.9)</td>
<td>868 (4.6)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Secondary</td>
<td>4565 (10.8)</td>
<td>154 (3.4)</td>
<td></td>
<td>Yes</td>
<td>74 (0.2)</td>
<td>6 (8.1)</td>
<td></td>
</tr>
<tr>
<td>Higher</td>
<td>18 486 (43.7)</td>
<td>512 (2.8)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Occupation</td>
<td></td>
<td></td>
<td>&lt;0.001</td>
<td>Presentation</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Unemployed</td>
<td>5710 (13.5)</td>
<td>323 (5.7)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Employed</td>
<td>36 102 (85.3)</td>
<td>1218 (3.4)</td>
<td></td>
<td>Term birth (&gt;37 weeks)</td>
<td>37 764 (89.2)</td>
<td>914 (2.4)</td>
<td></td>
</tr>
<tr>
<td>Others</td>
<td>507 (1.2)</td>
<td>20 (3.9)</td>
<td></td>
<td>Preterm birth (&lt;37 weeks)</td>
<td>4555 (10.8)</td>
<td>647 (14.2)</td>
<td></td>
</tr>
<tr>
<td>Gestational age at birth</td>
<td></td>
<td></td>
<td>&lt;0.001</td>
<td>Birth weight</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>42 319</td>
<td>1561 (3.7%)</td>
<td></td>
<td>Normal birth weight</td>
<td>37 991 (89.8)</td>
<td>801 (2.1)</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Low birth weight</td>
<td>4328 (10.2)</td>
<td>760 (17.6)</td>
<td></td>
</tr>
<tr>
<td>Child’s sex</td>
<td></td>
<td></td>
<td>0.42</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Female</td>
<td>20 430 (48.3)</td>
<td>738 (3.6)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Male</td>
<td>21 889 (51.7)</td>
<td>823 (3.8)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

*P value based on the χ² test.
ANC, antenatal care; PPH, postpartum haemorrhage; PROM, premature rupture of the membranes.

(Specificity 0.88, 95% CI 0.88 to 0.89) had slightly higher specificity compared with Lreg (specificity 0.87, 95% CI 0.86 to 0.88). Due to the low prevalence of perinatal death the sensitivity was lower for both approaches.

Figure 1  Schematic diagram showing the number of singleton deliveries analysed, KCMC medical birth registry data, 2000–2015. KCMC, Kilimanjaro Christian Medical Centre.

Figure 2  Trends of perinatal death, KCMC medical birth registry data, 2000–2015. KCMC, Kilimanjaro Christian Medical Centre.
deaths (3.7%), all models had high (NPV 0.98, 95% CI 0.98 to 0.98).

With regard to the number of actual and predicted outcomes (table 3), all models correctly predicted perinatal deaths by more than half of 468 deaths in the testing set. The numbers of correct classification were higher in the ANN 280 (59.8%) and NB 267 (57.1%), followed by the Lreg model 261 (55.8%) and bagging 260 (55.6%). The decision curve analysis (figure 4B) demonstrated that the net benefit of the RF model surpassed that of other ML models, including Lreg for all threshold values, indicating that the RF model is more superior in predicting the risk of perinatal deaths in this cohort. The accuracy of the RF model was 0.87, 95% CI (0.86 to 0.87), compared with 0.87, 95% CI (0.87 to 0.88) for boosting and 0.86, 95% CI (0.85 to 0.86) for the Lreg model (table 2). Furthermore, other ML models (except for boosting) demonstrated high net benefit over a range of threshold probability values relative to that of the Lreg model. Also, the RF model had a superior net benefit over all models (figure 4B).

FIGURE 3 Variable importance of predictors for perinatal death in the random forest model scaled to have a maximum value of 100. ANC, antenatal care.

FIGURE 4 Prediction ability of perinatal deaths comparing different machine learning models in the test set: (A) Receiver operating characteristics curves. The corresponding values of the area under the receiver operating characteristics curve for each model are in table 2. (B) Decision curve analysis. The net benefit of the machine learning models (except for boosting) is larger over a range of threshold probability values compared with that of the logistic regression model.

DISCUSSION
In this study, the perinatal death was predicted using five ML models (ANN, RFs, NB, bagging and boosting). There were no differences in the predictive performance between ML models except for bagging, which had a lower predictive performance. The ANN and NB had higher sensitivity compared with the Lreg, and other ML models. Specificity for all models was high, mainly due to the low prevalence of perinatal deaths in this cohort. Additionally, results from the decision curve analysis revealed that the ML models (except for boosting) had a higher net benefit over a range of threshold probability values compared with the Lreg model, indicating high accuracy. The RF model demonstrated a superior net benefit over other models.

In the present study, maternal characteristics before and during pregnancy, pregnancy history, and paternal characteristics identified pregnancies at high risk of experiencing adverse perinatal outcomes that might need close clinical follow-ups. It is worth noting here that paternal age and education level were highly predictive of perinatal death more than the known pregnancy-related conditions or complications such as prematurity. Previous literature shows that paternal characteristics, particularly advanced paternal age, increase the risk of adverse perinatal outcomes, such as low birth weight, prematurity, small for gestational age and low Apgar scores,62-64 despite conflicting evidence from other studies.65 Furthermore, studies using data from the KCMC Medical Birth Registry (same data source to the current study) focused on modelling the association between maternal
and pregnancy-related characteristics and complications during pregnancy and childbirth with the risk of adverse perinatal outcomes but ignored paternal characteristics. Despite challenges in male involvement in pregnancy and childbirth in Tanzania, their participation is critical to improving maternal and child health outcomes.

On top of clinicians’ judgement, previous investigators applied standard regression models in prediction of risk for adverse perinatal outcomes, particularly perinatal death. We found no differences in the predictive performance of the ML models, except for bagging, which had lower predictive capacity. The sensitivity of the ML models was also almost comparable to that of Lreg, which indicates that both models correctly classified perinatal deaths. Our finding is consistent with a recent systematic review that showed no performance benefit of ML models over Lreg for the prediction of clinical outcomes. The possible explanation for lack of differences in the performance between the compared models could be attributed to the low proportion of outcome and exposures in this study, as well as data quality and recording challenges inherent in registry-based studies.

In contrast, some previous investigators have demonstrated that ML models could better predict outcomes in pregnant women compared to traditional models. However, we found no differences in the predictive performance of the ML models, excluding bagging, which had lower predictive capacity. The sensitivity of the ML models was also almost comparable to that of Lreg, which indicates that both models correctly classified perinatal deaths. Our finding is consistent with a recent systematic review that showed no performance benefit of ML models over Lreg for the prediction of clinical outcomes. The possible explanation for lack of differences in the performance between the compared models could be attributed to the low proportion of outcome and exposures in this study, as well as data quality and recording challenges inherent in registry-based studies.

Table 2: Prediction performance of the reference and machine learning models in the test set

<table>
<thead>
<tr>
<th>Model</th>
<th>Logistic regression</th>
<th>Artificial neural network</th>
<th>Random forests</th>
<th>Naïve bayes</th>
<th>Bagging</th>
<th>Boosting</th>
</tr>
</thead>
<tbody>
<tr>
<td>ACC</td>
<td>0.86 (0.85 to 0.86)</td>
<td>0.83 (0.82 to 0.83)</td>
<td>0.87 (0.86 to 0.87)</td>
<td>0.84 (0.83 to 0.85)</td>
<td>0.82 (0.81 to 0.83)</td>
<td>0.87 (0.87 to 0.88)</td>
</tr>
<tr>
<td>AUC</td>
<td>0.78 (0.76 to 0.81)</td>
<td>0.78 (0.76 to 0.80)</td>
<td>0.79 (0.76 to 0.81)</td>
<td>0.79 (0.76 to 0.81)</td>
<td>0.76 (0.74 to 0.79)</td>
<td>0.79 (0.76 to 0.81)</td>
</tr>
<tr>
<td>P value*</td>
<td>Reference</td>
<td>0.59</td>
<td>0.37</td>
<td>0.65</td>
<td>0.006</td>
<td>0.20</td>
</tr>
<tr>
<td>Sensitivity</td>
<td>0.56 (0.51 to 0.60)</td>
<td>0.60 (0.55 to 0.64)</td>
<td>0.54 (0.49 to 0.58)</td>
<td>0.57 (0.52 to 0.62)</td>
<td>0.55 (0.50 to 0.59)</td>
<td>0.54 (0.49 to 0.58)</td>
</tr>
<tr>
<td>Specificity</td>
<td>0.87 (0.86 to 0.88)</td>
<td>0.84 (0.83 to 0.84)</td>
<td>0.88 (0.88 to 0.89)</td>
<td>0.85 (0.84 to 0.86)</td>
<td>0.83 (0.82 to 0.84)</td>
<td>0.89 (0.88 to 0.89)</td>
</tr>
<tr>
<td>PPV</td>
<td>0.14 (0.12 to 0.16)</td>
<td>0.12 (0.11 to 0.14)</td>
<td>0.15 (0.13 to 0.17)</td>
<td>0.13 (0.11 to 0.14)</td>
<td>0.11 (0.10 to 0.12)</td>
<td>0.15 (0.14 to 0.17)</td>
</tr>
<tr>
<td>NPV</td>
<td>0.98 (0.98 to 0.98)</td>
<td>0.98 (0.98 to 0.98)</td>
<td>0.98 (0.98 to 0.98)</td>
<td>0.98 (0.98 to 0.98)</td>
<td>0.98 (0.98 to 0.98)</td>
<td>0.98 (0.98 to 0.98)</td>
</tr>
</tbody>
</table>

*We calculated p values to compare the AUC the receiver operating characteristics curve of logistic with each machine learning model.

ACC, accuracy; AUC, area under the curve; NPV, negative predictive value; PPV, positive predictive value.
or adverse pregnancy outcomes compared with classical regression models. The application of ML models may improve the classification of adverse events occurring during the perinatal period and, therefore, assist in triaging and provision of close clinical follow-up for women at high risk. Other studies also provide evidence of improved prediction of under-5 and neonatal mortality, using ML models. The utility of these models may, therefore, improve the prediction of adverse pregnancy outcomes as opposed to standard regression models.

In this study, the decision curve analysis that accounts for the impact of false-negative and false-positive misclassification errors showed superior predictive performance of the ML approaches over the Lreg model. This demonstrates a higher net benefit for the prediction of perinatal deaths. The higher net benefit in the prediction ability of the ML approaches has also been documented elsewhere. This is because ML approaches can incorporate the high order nonlinear interactions between predictors, which cannot be addressed by traditional modelling approaches, including the Lreg model. Furthermore, the use of cross-validation is also known to reduce potential overfitting in ML models. It is important to note that ML approaches are, to a large extent, non-parametric as opposed to the Lreg model that relies on strong distributional assumptions.

The strength of this study is that it is the first to apply modern ML approaches to predict perinatal deaths, particularly in Tanzania and to a large extent sub-Saharan Africa, compared with the classical Lreg model. Our study demonstrated that ML models might be used to improve the prediction of perinatal deaths and triage of women at risk. We also used the SMOTE balancing technique to avoid the bias of the model toward skewed data (reduce overfitting), hence improving the prediction accuracy of the ML algorithm. However, SMOTE is not very effective for high dimensional data. Our study also had some limitations that are worth considering when interpreting the results. First, we excluded observations with missing values in both the outcome and exposures from the analysis, a problem inherent in cohort studies, including birth registries, which may lead to underestimation of the proportion of perinatal death. Two excluded variables (maternal BMI and HIV status) have been associated with perinatal and under-5 deaths; hence their exclusion might increase the risk of residual confounding bias. The effect of exclusion of these two variables and missing values to predict perinatal deaths remains unquantified.

Second, selection bias/or referral bias is a common problem in hospital-based studies, which affects the generalisation of findings to the general population. This might also be the case in the present study. However, our findings might reflect a similar setting in Tanzania and probably in other sub-Saharan African countries. Third, the KCMC Medical Birth Registry cohort only captures perinatal deaths occurring in the health facility (KCMC hospital), which may underestimate the observed perinatal deaths in the wider population. Currently, the hospital has no mechanisms to follow-up the birth outcomes from deliveries that occur at home and post-discharge outcomes of the babies after mothers are discharged from the hospital within the first week, especially within the KCMC hospital catchment area. Future extensions include ways of handling missing values before applying the ML algorithms to predict perinatal death and other adverse pregnancy outcomes.

CONCLUSION

The ML models (except for bagging) performed equally with the Lreg model to predict perinatal deaths using maternal, paternal and obstetric factors in this cohort. The ML models, however, have a higher net benefit, demonstrating superiority in the prediction of perinatal death. Furthermore, the RF model also demonstrated superior performance over other ML models. These models are a useful and alternative strategy over the standard Lreg model to predict perinatal deaths, considering the richness of the medical birth registries. Moreover, the ML models are capable of handling many predictors at the same time, which is crucial in capturing multiple risk factors for adverse perinatal outcomes such as perinatal deaths. The application of ML models may, therefore, increase the prediction ability of adverse perinatal outcomes and thereby helping in triage women most at risk.

Acknowledgements We would like to acknowledge the midwives who participated in data collection and all women and children whose information enabled the availability of data used in this study. The authors also thank the staff at the Birth Registry for capturing these data in the electronic system. We also appreciate the Centre for International Health at the University of Bergen in Norway and the Department of Obstetrics and Gynecology of the KCMC hospital in Tanzania for establishing the KCMC medical birth registry, which facilitated the availability of data to conduct this study.

Contributors All authors made a substantial contribution to this study. IBM, MJM, JO and HGM designed the study. IBM and MJM analysed the data. MJM, JO and HGM commented on the manuscript. IBM drafted the manuscript and had primary responsibility for the final content. All authors reviewed the drafts of this manuscript and approved the final version for submission.

Funding This work was funded by GSK Africa Non-Communicable Disease Open Lab through the DELTAS Africa Sub-Saharan African Consortium for Advanced Biostatistics (SSACAB) Grant No. 107754/2/15/Z- training programme. The views expressed in this publication are those of the author(s) and not necessarily those of GSK.

Disclaimer The views expressed in this publication are those of the author(s) and not necessarily those of AAS, NEPAD Agency, Wellcome Trust or the UK government. The Norwegian Council for Higher Education’s Programme for Development Research (NUFU) funded the establishment of the birth registry at the Kilimanjaro Christian Medical Centre (KCMC). The funders had no role in study design, data collection, and analysis, decision to publish, or preparation of the manuscript.

Competing interests None declared.

Patient consent for publication Not required.

Ethics approval This study was approved by the Kilimanjaro Christian Medical College Research Ethics and Review Committee (KCMU-CRERC) with approval number 2424. Before data collection, mothers provided oral informed consent after receiving information about the purpose of the study, questionnaire, and its contents as well as the intention to gather new knowledge, which will, in turn,
benefit mothers and children in the future. Mothers were informed that participation was voluntary and had no implications on the care they would receive. Following consent, mothers were free to refuse to reply to single questions. For privacy and confidentiality, unique identification numbers were used to identify and then link mothers with child records. There was no person-identifiable information in any electronic database, and instead, unique identification numbers were used. Necessary measures were taken by midwives to ensure privacy during the interview process.

Provenance and peer review  Not commissioned; externally peer reviewed.

Data availability statement  Data are available on reasonable request. The KCMC medical birth registry data contains potentially identifying and sensitive patient information. This has also been stipulated by the local Institutional Review Board of KCMC hospital and the National Ethics Committee in Tanzania when establishing this birth registry. Permission to use the data in this study was made through the Kilimanjaro Christian Medical University College Research and Ethics Review Committee, and received an approval number 2424. The authors do not have the legal right to share the data publicly. All data requests can be sent to the Executive Director of the KCMC referral hospital, P. O. Box 3010, Moshi, Tanzania, Email: kcmcadmin@kcmc.ac.tz or through the corresponding author.

Open access  This is an open access article distributed in accordance with the Creative Commons Attribution 4.0 Unported (CC BY 4.0) license, which permits others to copy, redistribute, remix, transform and build upon this work for any purpose, provided the original work is properly cited, a link to the licence is given, and any changes made are indicated. See: https://creativecommons.org/licenses/by/4.0/.

ORCID iDs  Innocent B Mboya http://orcid.org/0000-0001-9861-5879
Michael J Mahande http://orcid.org/0000-0002-7750-7657

REFERENCES


48 StataCorp. Stata statistical software: release 15. College Station, TX: StataCorp LLC, 2017.


54 IEEE. Bayesian networks classifiers for gene-expression data. 2011 11th International Conference on intelligent systems design and applications, 2011.


64 Peneza AK, Maluka SO. ‘Unless you come with your partner you will be sent back home’: strategies used to promote male involvement in antenatal care in southern Tanzania. *Glob Health Action* 2018;11:1449724.


73 Blagus R, Lusa L. Improved shrunk centroid classifiers for high-dimensional class-imbalanced data. *BMJ Bioinformatics* 2013;14:64.
# input Stata file
library(foreign)
data <- read.dta("D:/Documents/Innocent/PhD 2018/PhD project/Data/Results/Machine Learning Paper/complete_data_withlabels.dta")
utils::View(head(data))
summary(data)
#############################################################################
# Divide data into training and testing sets
library(caret)
set.seed(100) # For reproducibility

trainIndex <- createDataPartition(data$perinatal_death, p=.70, list=FALSE) # divide the data based on the outcome variable in the same ratio
trainData <- data[trainIndex,]
testData <- data[-trainIndex,]
table(data$perinatal_death)
table(trainData$perinatal_death)
table(testData$perinatal_death)
#############################################################################
# Variable/Feature Selection
#############################################################################

# install.packages("mlbench")
# install.packages("Hmisc")
# install.packages("randomForest")

# RandomForests
# install.packages("caret", "mlbench", "mlbench", "randomForest")

library(mlbench)
library(Hmisc)
library(randomForest)
library(caret)

# speed up the performance of the analysis
# library(doParallel)
# ncores <- detectCores() - 1
# cl <- makeCluster(ncores)
# registerDoParallel(cl)

# Parameter tuning and Cross validation
ctrl <- trainControl(method="CV", # 10fold cross validation
number = 10,
#repeats = 10, #has no meaning for this resampling method
savePredictions = TRUE,
classProbs=TRUE,
allowParallel = TRUE,
summaryFunction = twoClassSummary
#search = "random" # Random searching of hyperparameters?
)

#Feature selection using random forests algorithm
RFMod <- train(x=trainData[,-which(names(trainData) %in% c("perinatal_death")]),
y=trainData[,which(names(trainData) %in% c("perinatal_death"))],
method="rf", metric="ROC", trControl=ctrl)

#select and plot the Important variables from the RF model
RFImp <- varImp(RFMod, scale=T)
RFImp
plot(RFImp, main='Variable Importance')
plot(RFImp, top = 20, main='Variable Importance') #all 20 selected features

#Plot of important features with proper names
impvarsdata<-data.frame(impvars=c("Year of birth", "Paternal age groups", "Birth weight",
"Paternal education level", "Maternal education level", "Maternal occupation",
"Maternal age groups", "Sex of the child", "Mode of delivery", "Number of ANC visits",
"Area of residence", "Alcohol use during pregnancy", "Paternal occupation",
"Gestational age at birth", "Malaria during pregnancy", "Induction of labor",
"Maternal marital status", "Abruptio placenta", "Referral status",
"Pre-eclampsia/ eclampsia"),
rfvalues=c(100.00, 41.12, 31.51, 28.49, 26.82, 26.32, 26.18, 25.00, 22.22,
21.13, 19.53, 19.39, 17.64, 17.29, 14.31, 13.96, 12.57, 11.96,
11.11, 10.69))

utils::View(impvarsdata) #compare with RF below
RFImp
class(impvarsdata$rfvalues)
class(impvarsdata$impvars)

# arrange in descending order
# https://rpubs.com/Felix/7644

impvarsdata$impvars <- factor(impvarsdata$impvars,
levels = impvarsdata[order(impvarsdata$rfvalues), "impvars"]

ggplot(impvarsdata, aes(x = impvars, y = rfvalues)) + geom_bar(stat = "identity") +
  coord_flip() + xlab("Predictors") + ylab("Importance") +
  theme(axis.text = element_text(colour = "black")) +
  theme(axis.line = element_line(colour = "black"))

ggplot(RFImp, top = 20, main='Variable Importance') #all 20 selected features

# Remove the variables which were not selected from both training and testing datasets

#training
trainDataIMP <- trainData[, c("year", "fage_grp", "lbw", "feducation", "meducation2", "mocup_grp2",
  "mage_grp2", "sex2", "demode_grp2", "ancvisit_grp", "residence_grp2", "alcopreg2", "PTB_LMP_grp",
  "focup_grp", "INDUCT", "MHDP10", "marstat_grp2", "CMPL04", "refdel2", "PE_E")]
trainDataIMP$perinatal_death <- trainData$perinatal_death[match(row.names(trainDataIMP),
  row.names(trainData))]

#testing
testDataIMP <- testData[,c("year", "fage_grp", "lbw", "feducation", "meducation2", "mocup_grp2",
  "mage_grp2", "sex2", "demode_grp2", "ancvisit_grp", "residence_grp2", "alcopreg2", "PTB_LMP_grp",
  "focup_grp", "INDUCT", "MHDP10", "marstat_grp2", "CMPL04", "refdel2", "PE_E")]
testDataIMP$perinatal_death <- testData$perinatal_death[match(row.names(testDataIMP),
  row.names(testData))]

#save both data to CSV
write.csv(x = trainDataIMP, file = "D:/Documents/Innocent/PhD 2018/PhD project/Data/Results/Machine Learning Paper/trainDataIMP.csv", row.names = TRUE)
write.csv(x = testDataIMP, file = "D:/Documents/Innocent/PhD 2018/PhD project/Data/Results/Machine Learning Paper/testDataIMP.csv", row.names = TRUE)

#Classification - use selected variables (from training data) to predict perinatal death (in testing data)

#Reload the saved data
TrainData <- read.csv(file = "D:/Documents/Innocent/PhD 2018/PhD project/Data/Results/Machine Learning Paper/trainDataIMP.csv", header = TRUE, row.names = 1, stringsAsFactors = TRUE)
TestData <- read.csv(file = "D:/Documents/Innocent/PhD 2018/PhD project/Data/Results/Machine Learning Paper/testDataIMP.csv", header = TRUE, row.names = 1, stringsAsFactors = TRUE)

# 1) Logistic regression

library(DMwR)
ctrlLReg <- trainControl(## 10-fold CV
method = "cv",
number = 10,
classProbs=TRUE,
savePredictions = TRUE,
allowParallel = TRUE,
sampling = "smote",
# To correct for class imbalances (more alive than death in perinatal_death)
summaryFunction = twoClassSummary)

# Train
set.seed(1239)
lreg <- train(x=TrainData[-which(names(TrainData) %in% c("perinatal_death"))],
y= TrainData$perinatal_death, method="glm",family=binomial(),
trControl=ctrlLReg)
lreg

#-------------------------------------------------------------
# Test
set.seed(1239)
lreg.pred <- predict(lreg, TestData[-which(names(TestData) %in% c("perinatal_death"))],
y= TestData$perinatal_death, method="glm",family=binomial(),
trControl=ctrlLReg)
lreg.tab = table(pred = lreg.pred, true = TestData[,c("perinatal_death")])
lreg.Conf = confusionMatrix(lreg.pred, TestData[,c("perinatal_death")], positive =
levels(TestData[,c("perinatal_death")])[2])
lreg.Conf

# 2) Naive Bayes

# Load required packages
library(dplyr)       # Used by caret
library(nnet)       # support vector machine
library(pROC)       # plot the ROC curves
library(e1071)
# install.packages("naivebayes")
library(naivebayes)
set.seed(1236)
# Setup for cross validation
ctrlNB <- trainControl(method="CV", # 10fold cross validation
    number = 10,
    savePredictions = TRUE,
    classProbs=TRUE,
    allowParallel = TRUE,
    sampling = "smote",
    summaryFunction = twoClassSummary
)
set.seed(1236)

# Train
NBModel.tune <- train(x=TrainData[-which(names(TrainData) %in% c("perinatal_death"))],
    y= TrainData$perinatal_death,
    method = "naive_bayes", # Naive Bayes
    tuneLength = 5, # 9 values of the cost function
    preProc = c("center","scale"), # Center and scale data
    not needed for this analysis
    metric="ROC",
    trControl=ctrlNB)

NBModel.tune
plot(NBModel.tune)

# Test
NBModel.pred <- predict(NBModel.tune, TestData[-which(names(TestData) %in% c("perinatal_death"))])
NBModel.tab = table(pred = NBModel.pred, true = TestData[,c("perinatal_death")])
NBModel.Conf = confusionMatrix(NBModel.pred, TestData[,c("perinatal_death")], positive = levels(TestData[,c("perinatal_death")])[2])
NBModel.Conf

# 3) Artificial neural network (ANN)

# Load important packages
# library(caret)
# library(dplyr)     # Used by caret
library(nnet)      # support vector machine
# library(pROC)     # plot the ROC curves
```r
set.seed(1234)
# set.seed(123)
# Setup for cross validation
ctrlANN <- trainControl(method="CV", # 10fold cross validation
 number = 10,
 # savePredictions = TRUE,
 classProbs=TRUE,
 allowParallel = TRUE,
 savePredictions = TRUE,
 sampling = "smote",
 summaryFunction = twoClassSummary
)

# Train: Artificial neural network (ANN) model
set.seed(1234)
ANNModel.tune <- train(x=TrainData[-which(names(TrainData) %in% c("perinatal_death"))],
y= TrainData$perinatal_death,
 method = "nnet", # neural network
 # tuneLength = 5, # 9 values of the cost function
 # preProc = c("center","scale"), # Center and scale data
 # metric="ROC",
 # linout=TRUE,
 # maxit = 1000,
 # tuneGrid = my.grid,
 # trace = F,
 # tuneLength = 10,
 # linout = 1,
 # trControl=ctrlANN,
 metric="ROC")

ANNModel.tune
plot(ANNModel.tune)

# Test
set.seed(1234)
ANNModel.pred <- predict(ANNModel.tune, TestData[-which(names(TestData) %in% c("perinatal_death"))])
ANNModel.tab = table(pred = ANNModel.pred, true = TestData[,c("perinatal_death")])
ANNModel.Conf = confusionMatrix(ANNModel.pred, TestData[,c("perinatal_death")],
 positive = levels(TestData[,c("perinatal_death")])[2])
ANNModel.Conf
```
# 4) Bagged Tree

```r
# Specify the training configuration
ctrlBagTree <- trainControl(method = "cv",
    number = 10,
    classProbs=TRUE,
    savePredictions = TRUE,
    allowParallel = TRUE,
    sampling = "smote",
    summaryFunction = twoClassSummary)  # For AUC

# Cross validate the credit model using "treebag" method;
# Track AUC (Area under the ROC curve)
set.seed(1237)  # for reproducibility
BaggedTree <- train(x=TrainData[-which(names(TrainData) %in% c("perinatal_death"))],
    y= TrainData$perinatal_death,
    method = "treebag",
    metric = "ROC",
    trControl = ctrlBagTree)
BaggedTree
```

# Test
```r
set.seed(1237)
BaggedTree.pred <- predict(BaggedTree, TestData[, -which(names(TestData) %in% c("perinatal_death"))],
    pred = BaggedTree.pred, true = TestData[, c("perinatal_death")],
    metric = "ROC",
    trControl = ctrlBagTree)
BaggedTree.Conf = confusionMatrix(BaggedTree.pred, TestData[, c("perinatal_death")], positive =
    levels(TestData[, c("perinatal_death")])[2])
# if you need to predict alive change to 1 at the end of the syntax
BaggedTree.Conf
```

# 5) Boosting

```r
#install gbm package
# install.packages("gbm")
library(gbm)

# Specify the training configuration
ctrlBoosting <- trainControl(method = "cv",
    number = 10,
    classProbs=TRUE,
    savePredictions = TRUE,
```

BMJ Publishing Group Limited (BMJ) disclaims all liability and responsibility arising from any reliance placed on this supplemental material which has been supplied by the author(s)
allowParallel = TRUE,
sampling = "smote",
summaryFunction = twoClassSummary)  # For AUC

# Cross validate the credit model using "treebag" method;
# Track AUC (Area under the ROC curve)
set.seed(1238)  # for reproducibility
Boosting <- train(x=TrainData[-which(names(TrainData) %in% c("perinatal_death"))],
y= TrainData$perinatal_death,
method = "gbm",
metric = "ROC",
trControl = ctrlBoosting)
plot(Boosting)

#-------------------------------------------------------------
# Test
set.seed(1238)
Boosting.pred <- predict(Boosting, TestData[-which(names(TestData) %in% c("perinatal_death"))])
Boosting.tab = table(pred = Boosting.pred, true = TestData[,c("perinatal_death")])
Boosting.Conf = confusionMatrix(Boosting.pred, TestData[,c("perinatal_death")], positive =
levels(TestData[,c("perinatal_death")])[2])
Boosting.Conf

# 6) Random Forests (RF)
# Load additional libraries
library(caTools)
library(mlbench)

#set.seed(123)
# Setup for cross validation
ctrlRF <- trainControl(method="CV",  # 10fold cross validation
   number = 10,
savePredictions = TRUE,
classProbs=TRUE,
allowParallel = TRUE,
sampling = "smote",
summaryFunction = twoClassSummary
)

#Train
set.seed(1235)
RFModel.tune <- train(x=TrainData[-which(names(TrainData) %in% c("perinatal_death"))],
y = TrainData$perinatal_death,
method = "rf", # Random Forest
#tuneLength = 5,       # 9 values of the cost function
#preProc = c("center","scale"), # Center and scale data
metric="ROC",
trControl=ctrlRF)

RFModel.tune
plot(RFModel.tune)

# #-------------------------------------------------------------
# # Test
set.seed(1235)
RFModel.pred <- predict(RFModel.tune, TestData[,which(names(TestData) %in% c("perinatal_death")])]
RFModel.tab = table(pred = RFModel.pred, true = TestData[,c("perinatal_death")])
RFModel.Conf = confusionMatrix(RFModel.pred, TestData[,c("perinatal_death")], positive =
levels(TestData[,c("perinatal_death")])[[2]]
RFModel.Conf

# # RESULTS
# #-------------------------------------------------------------

# # Plot ROC curves
# #-------------------------------------------------------------
library(ROCR)
probANN <- predict(ANNModel.tune, newdata=TestData[,which(names(TestData) %in%
c("perinatal_death")])], type="prob")
predANN <- prediction(probANN[,2], TestData[,c("perinatal_death")])
perfANN <- performance(predANN, measure = "tpr", x.measure = "fpr")
plot(perfANN, col="Black", lwd=2)

probRF <- predict(RFModel.tune, newdata=TestData[,which(names(TestData) %in%
c("perinatal_death")])], type="prob")
predRF <- prediction(probRF[,2], TestData[,c("perinatal_death")])
perfRF <- performance(predRF, measure = "tpr", x.measure = "fpr")
plot(perfRF, col="Blue", lwd=2, add = TRUE)

probNB <- predict(NBModel.tune, newdata=TestData[,which(names(TestData) %in%
c("perinatal_death")])], type="prob")
predNB <- prediction(probNB[,2], TestData[,c("perinatal_death")])
perfNB <- performance(predNB, measure = "tpr", x.measure = "fpr")
plot(perfNB, col="Green", lwd=2, add = TRUE)
probLreg <- predict(lreg, newdata=TestData[,which(names(TestData) %in% c("perinatal_death"))], type="prob")
predLreg <- prediction(probLreg[,2], TestData[,c("perinatal_death")])
perfLreg <- performance(predLreg, measure = "tpr", x.measure = "fpr")
plot(perfLreg, col="Red", lwd=2, add = TRUE)

probBaggedTrees <- predict(BaggedTree, newdata=TestData[,which(names(TestData) %in%
c("perinatal_death"))], type="prob")
predBaggedTrees <- prediction(probBaggedTrees[,2], TestData[,c("perinatal_death")])
predBoosting <- performance(predBaggedTrees, measure = "tpr", x.measure = "fpr")
plot(predBaggedTrees, col="Orange", lwd=2, add = TRUE)

probBoosting <- predict(Boosting, newdata=TestData[,which(names(TestData) %in%
c("perinatal_death"))], type="prob")
predBoosting <- prediction(probBoosting[,2], TestData[,c("perinatal_death")])
predBoosting <- performance(predBoosting, measure = "tpr", x.measure = "fpr")
plot(predBoosting, col="DarkGray", lwd=2, add = TRUE)

# plots together
par(mar=c(5,5,1,1), cex=0.9) # cex = text font size, mar=margins (left, bottom, top, right)
plot(perfANN, col="Black", lwd=2)
plot(perfRF, col="Blue", lwd=2, add = TRUE)
plot(perfNB, col="Green", lwd=2, add = TRUE)
plot(perfLreg, col="Red", lwd=2, add = TRUE)
plot(perfBaggedTrees, col="Orange", lwd=2, add = TRUE)
plot(perfBoosting, col="DarkGray", lwd=2, add = TRUE)

legend(0.55,0.7, legend = c("Artificial neural networks", "Random forest", "Naive bayes",
"Logistic regression", "Bagging", "Boosting"),
col = c("Black", "Blue", "Green", "Red", "Orange", "DarkGray"),
lt = 1,lw = 3,bty = "n", y.intersp = 1.5, cex=0.8)

# Compute AUC values

ANN_auc <- performance(predANN, measure = "auc")
ANN_auc <- ANN_auc@y.values[[1]]

RF_auc <- performance(predRF, measure = "auc")
RF_auc <- RF_auc@y.values[[1]]

NB_auc <- performance(predNB, measure = "auc")
NB_auc <- NB_auc@y.values[[1]]
Lreg_auc <- performance(predLreg , measure = "auc")
Lreg_auc <- Lreg_auc@y.values[[1]]

Bagged_auc <- performance(predBaggedTrees , measure = "auc")
Bagged_auc <- Bagged_auc@y.values[[1]]

Boosting_auc <- performance(predBoosting , measure = "auc")
Boosting_auc <- Boosting_auc@y.values[[1]]

#an alternative and short way compared to the above syntax
aucANN<-auc(as.numeric(TestData$perinatal_death), probANN[,2])#ANN
aucRF<-auc(as.numeric(TestData$perinatal_death), probRF[,2])#RF
aucNB<-auc(as.numeric(TestData$perinatal_death), probNB[,2])#NB
aucLreg<-auc(as.numeric(TestData$perinatal_death), probLreg[,2])#Lreg
aucBagging<-auc(as.numeric(TestData$perinatal_death), probBaggedTrees[,2])#Bagging
aucBoosting<-auc(as.numeric(TestData$perinatal_death), probBoosting[,2])#Boosting

#============================================================================
#Compute CI for AUC
#============================================================================
library(pROC)
ci.auc(as.numeric(TestData$perinatal_death), probANN[,2])#ANN
ci.auc(as.numeric(TestData$perinatal_death), probRF[,2])#RF
ci.auc(as.numeric(TestData$perinatal_death), probNB[,2])#NB
ci.auc(as.numeric(TestData$perinatal_death), probLreg[,2])#Lreg
ci.auc(as.numeric(TestData$perinatal_death), probBaggedTrees[,2])#Bagging
ci.auc(as.numeric(TestData$perinatal_death), probBoosting[,2])#Boosting

#============================================================================
#Comparison of AUC for different machines using Delong's Test
#============================================================================
library(ROCR)
roc.test(aucLreg, aucANN, method = "delong") #delong test is the default
roc.test(aucLreg, aucNB)
roc.test(aucLreg, aucBagging)
roc.test(aucLreg, aucBoosting)
roc.test(aucLreg, aucRF)

#============================================================================
#Compute CT for Sens, Spec, Prevelance ... etc
#============================================================================
library(epiR)
epi.tests(ANNModel.Conf$table, conf.level = 0.95)
epi.tests(RFModel.Conf$table, conf.level = 0.95)
epi.tests(NBModel.Conf$table, conf.level = 0.95)
epi.tests(lreg.Conf$table, conf.level = 0.95)
epi.tests(BaggedTree.Conf$table, conf.level = 0.95)
epi.tests(Boosting.Conf$table, conf.level = 0.95)

# Summary statistics of the Results
library(caret)
results <- resamples(list(ANN=ANNModel.tune, RF=RFModel.tune,
                         NB=NBModel.tune, BaggedTree=BaggedTree, Lreg=lreg, Boosting=Boosting))
# Table comparison
summary(results)

# boxplot comparison
bwplot(results)
# Dot-plot comparison
dotplot(results)

# Decision curve analysis
library("tidyverse")
library("dplyr")

#Predict Perinatal death
ANNModel.pred <- predict(ANNModel.tune, TestData[, -which(names(TestData) %in%
c("perinatal_death")]), type = "prob")
RFModel.pred <- predict(RFModel.tune, TestData[, -which(names(TestData) %in% c("perinatal_death")]),
type = "prob")
NBModel.pred <- predict(NBModel.tune, TestData[, -which(names(TestData) %in%
c("perinatal_death")]), type = "prob")
lreg.pred <- predict(lreg, TestData[, -which(names(TestData) %in% c("perinatal_death")]), type = "prob")
BaggedTree.pred <- predict(BaggedTree, TestData[, -which(names(TestData) %in%
c("perinatal_death")]), type = "prob")
Boosting.pred <- predict(Boosting, TestData[, -which(names(TestData) %in% c("perinatal_death")]), type = "prob")
dcadata <- TestData %>% select(c(perinatal_death))
source("D:/Documents/Innocent/PhD 2018/PhD project/Data/Results/Machine Learning Paper/dca.r")
# This file is available at http://www.decisioncurveanalysis.org
library(reshape2)
dcadata$ANN<-as.numeric(ANNModel.pred$Died)
dcadata$RF <- as.numeric(RFModel.pred$Died)
dcadata$NB <- as.numeric(NBModel.pred$Died)
dcadata$Lreg <- as.numeric(lreg.pred$Died)
dcadata$Bagging <- as.numeric(BaggedTree.pred$Died)
dcadata$Boosting <- as.numeric(Boosting.pred$Died)
data.set <- dcadata

attach(data.set)
data.set$perinatal_death <- as.numeric(data.set$perinatal_death)
data.set$perinatal_death <- data.set$perinatal_death - 1 # 0 = alive, 1 = died (in the testing data)

# Plot the DCA curves
library(DCA)
dca(data=data.set, outcome="perinatal_death",
    predictors=c("ANN", "RF", "NB", "Lreg", "Bagging", "Boosting"), xstart=0, ymin=0)

dcaoutput <- dca(data=data.set, outcome="perinatal_death",
    predictors=c("ANN", "RF", "NB", "Lreg", "Bagging", "Boosting"), xstart=0, xstop=0.3, ymin=0)

dcadf <- data.frame(dcaoutput$net.benefit)
temp <- melt(dcadf, id="threshold",
    measure=c("ANN", "RF", "NB", "Lreg", "Bagging", "Boosting"))

library(ggplot2)

ggplot(temp,
   aes(x=threshold,
       y=value,
       colour=variable,
       group=variable)) + geom_line() +
   geom_line(size=0.8) + # Thicker line
   coord_cartesian(xlim = c(0, 0.075), ylim= c(0, 0.04)) +
   labs(x="Threshold probability (%)") + labs(y="Net benefit") +
   theme_minimal() + theme(legend.position = c(0.8, 0.8),
       text = element_text(size=12),
       panel.grid.major = element_blank(),
       panel.grid.minor = element_blank(),
       axis.line = element_line(colour = "black"),
       axis.ticks = element_line(size = 1),
       panel.border = element_rect(linetype = "solid"),
       fill = NA, colour = "black"),
   axis.text = element_text(color = "black"),
   scale_color_manual(values = c("Black", "Blue", "Green", "Red", "Orange",
                                "DarkGray"),
                      labels = c("Artificial neural networks",
                                 "Random Forest",
                                 "Naive Bayes",
                                 "Logistic Regression",
                                 "Bagged Trees",
                                 "Boosting")
)
"Random forest",
"Naive bayes",
"Logistic regression",
"Bagging",
"Boosting") + labs(color="")

# Combining the a) ROC and b) DC plots on top of one another.

# install.packages("magick")
library(magick)

Figure3_roc<-image_read("D:/Documents/Innocent/PhD 2018/PhD project/Data/Results/Machine Learning Paper/Figure3_roc.tiff")
Figure3_roc<-image_annotate(Figure3_roc, "a", font = 'Times', size = 30)
Figure3_roc

Figure3_deccurve2<-image_read("D:/Documents/Innocent/PhD 2018/PhD project/Data/Results/Machine Learning Paper/Figure3_deccurve.tiff")
Figure3_deccurve2<-image_annotate(Figure3_deccurve2, "b", font = 'Times', size = 30)
Figure3_deccurve2

img <- c(Figure3_roc, Figure3_deccurve2) # combine the two images

img <- image_scale(img) # size of the images

img

image_info(img)

img2<-image_append(image_scale(img), stack = TRUE) %>%
  image_border("grey", "1x1")

img2

image_info(img2)

# export the image to tiff format (better quality)
image_write(img2, path = "D:/Documents/Innocent/PhD 2018/PhD project/Data/Results/Machine Learning Paper/Figure3.tiff",
             format = "tiff")

# ===========================================================================