

Exploring the Performance of Machine Learning Models and Predictive Factors for Fetal Death: Preliminary Results



Maria Eduarda Ferro de Mello, Élisson da Silva Rocha, Flávio Leandro de Moraes, Barbara de Queiroz Figueiroôa, Marília Santana da Silva, Waldemar Brandão Neto, Theo Lynn, and Patricia Takako Endo

Abstract This study investigates the effectiveness of machine learning models in predicting fetal death and identifying significant predictive factors. The study utilized a dataset from the Programa Mãe Coruja Pernambucana (PMCP) that includes socio-demographic, prenatal, maternal, and family health history data. The data underwent pre-processing and was explored using four tree-based machine learning models, each of which was evaluated based on their performance and feature importance. The attributes that significantly impacted the learning process were the first prenatal week, maternal age, and months between pregnancies. The application of predictive models for fetal deaths in this context can enhance the ability to detect such occurrences thus representing a pivotal support tool for the PMCP to identify mothers with high risk of adverse outcomes and promote targeted interventions of monitoring during pregnancy, and ultimately increase the likelihood of positive outcomes for mothers and babies.

M. E. F. de Mello · É. da Silva Rocha · F. L. de Moraes · W. B. Neto · P. T. Endo (✉)
Universidade de Pernambuco, Recife, Brazil
e-mail: patricia.endo@upe.br

M. E. F. de Mello
e-mail: mefm@ecom.poli.br

É. da Silva Rocha
e-mail: esr2@ecom.poli.br

F. L. de Moraes
e-mail: flavio.leandromoraes@upe.br

W. B. Neto
e-mail: waldemar.neto@upe.br

B. de Queiroz Figueiroôa · M. S. da Silva
Programa Mãe Coruja Pernambucana, Recife, Brazil
e-mail: barbarafigueiroa@gmail.com

T. Lynn
Dublin City University, Dublin, Ireland
e-mail: theo.lynn@dcu.ie

1 Introduction

Fetal death is a significant public health issue that affects millions of parents and families worldwide. Primary care for prenatal and neonatal health has a significant impact on the lives and health of mothers and developing babies. Increased investment in the provision of routine and emergency prenatal and neonatal care, basic sanitation, immunizations, and access to skilled health care has reduced the likelihood of neonatal and maternal death. Notwithstanding this, over two million pregnancies resulted in stillbirths in 2020; over 40% of which occurred during childbirth [1]. In the same year, a further 2.4 million children died in the first 28 d of life representing 47% of deaths of children under 5 years old [2].

Fetal and early neonatal mortality share the same etiology and conditions that result in the death of the fetus or newborn in the first hours of life. According to the World Health Organization (WHO), fetal death includes babies who die after the 22nd week of gestation, before expulsion or complete extraction from the mother's body [2, 3]. They can be classified as early or late (after the 28th week). The United Nations 2030 Agenda for Sustainable Development has specific targets for reducing global maternal mortality [4], and ending preventable deaths of newborns and children under 5 years of age [5]. Fetal Mortality Rate (FMR) is one of the indicators that assess the quality of health care provided to pregnant women during pregnancy and childbirth. This index expresses the number of fetal deaths with fetuses weighing at least 500 g or 25 cm in height per total births in the population of a given area [3]. The Sustainable Development Goals (SDGs) aim to reduce the global neonatal mortality rate to at least as low as 12 deaths per 1000 live births by 2030, however it does not specifically address to fetal mortality rate.

Fetal deaths are considered potentially preventable but it is important to identify the determinants of fetal deaths. Highly cited risk factors associated with mothers include obesity, alcohol and tobacco use, HIV seropositivity, Specific Hypertensive Disorders of Pregnancy (HDP), gestational diabetes mellitus, and placental and amniotic complications, which can directly influence congenital malformation, growth restriction and fetal death [6–8]. Recent studies indicate that pregnant women infected with SARS-CoV-2 may increase the risk of premature delivery and fetal death [9, 10]. Additionally, social factors such as maternal age, low income, inadequate schooling, and prenatal care also contribute to higher risk fetal death [11]. Notwithstanding these factors, significant and preventable factors that contribute to high rates of fetal and early neonatal mortality relate to poor quality prenatal care service, late diagnosis of complications during pregnancy, difficulty accessing care for pregnant women, and inadequate obstetric management [12]. The risks associated with these factors are exacerbated where there in multi-fetal gestation. Such instances are associated with additional prenatal risks including higher risk of preterm labor, preterm premature rupture of the membranes, intrauterine growth restriction, intrauterine fetal demise, gestational diabetes, and pre-eclampsia [13]. In such cases, planning prenatal care is crucial to estimate benefits and minimize adverse outcomes including fetal or multi-fetal death [14].

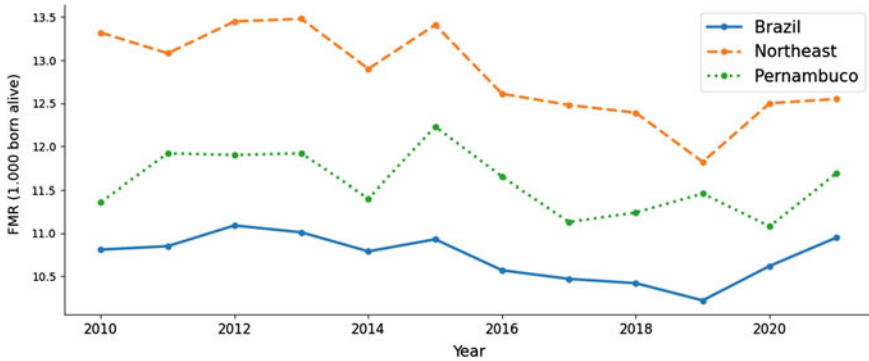


Fig. 1 Fetal death rate per 1000 live births in Brazil, Northeast Region, and state of Pernambuco. Data is available on the DATASUS website provided by the Brazilian Ministry of Health

In 2010, the Ministry of Health (MoH) mandated fetal and infant death monitoring and investigation as part of the Brazilian unified health system, the Sistema Único de Saúde (SUS). The data generated from this strategy enables researchers and policymakers to accurately gauge the scale of fatalities and categorize their root causes and contextual circumstances. This, in turn, facilitates the development of effective recommendations for targeted interventions aimed at preventing avoidable deaths [15, 16]. In 2010, the FMR for Brazil was estimated at 10.81 fetal deaths per 1000 live births; this decreased to 10.62 in 2020. Figure 1 depicts the evolution of the fetal death rate per 1000 live births in Brazil, in the Northeast region, and in the state of Pernambuco from 2010 to 2020. As can be seen from 1, in 2020, the Northeast region presented the second highest FMR in the country with 12.50 deaths per 1000 live births; the index for the state of Pernambuco was 11.08 [17].

In Pernambuco, one of the initiatives to reduce prenatal and neonatal still child-birth is the Programa Mãe Coruja Pernambucano (PMCP). Launched in 2007, the PMCP aims to provide comprehensive care to pregnant women and children up to 5 years of age. The PMCP is active in more than 105 municipalities in Pernambuco, mainly in vulnerable areas. Through the creation of a support network, the program ensures that mothers and their children receive the necessary care, including health services, education, social assistance, and family support. As a result, the program has significantly contributed to the reduction of maternal and infant mortality rates as well as improving social indicators and the quality of life of many families in Pernambuco [2, 18].

Despite the availability of the increased data and the PMCP, Fig. 1 suggests that challenges remain in the detection and prediction of adverse outcomes during pregnancy. Against this backdrop, machine learning models, due to their high predictive potential, have been widely proposed as solutions to support early diagnosis and monitoring during pregnancy and postpartum [19]. Extant research has used machine learning models to predict preterm birth, birth weight, mortality, hypertensive disorders, and postpartum depression, among other factors [20, 21]. Machine learning has

also been used to predict vaginal births after cesareans, understanding the characteristics of past and current pregnancies, and consequently assisting in the mode and management of labor [19]. Also, recent studies point to the use of machine learning to identify risks of fetal death and perinatal mortality [20, 22].

Developing predictive models and identifying factors associated with fetal death can aid in reducing its occurrence and improving healthcare services for affected parents and families. The primary aim of this study is to assess the effectiveness of predictive machine learning models based on data obtained from pregnant women who are receiving care at the PMCP. These initial findings represent a segment of an ongoing research project that seeks to establish decision support tools for healthcare professionals using predictive machine learning models in collaboration with the PMCP. In this work, we present the most significant clinical and socio-demographic attributes that contribute to the learning process of these models, thus enabling the selection of the most relevant features for further analysis. This study forms the foundation for future investigations aimed at developing practical tools to improve maternal health outcomes.

2 Related Works

Extant literature suggests that machine learning has significant potential for predicting fetal, neonatal, perinatal, and infant mortality [23]. In their review of the literature, Silva et al. [23] reviewed 18a publications from 2012 to 2021, however two publications by Shukla et al. [24] and Malacova et al. [22] focused on predicting fetal deaths.

Based on data from the NICHD Global Network for Women's and Children's Health Research Maternal and Newborn Health Registry, Shukla et al. [24] performed an analysis with data from women in the period of pregnancy up to the third day of delivery. The objective of the study was to predict the risk of fetal and neonatal mortality. For this, six machine learning models (Logistic Regression, Support Vector Machine, Logistic Elastic Net, Artificial Neural Networks (ANN), Gradient Boosted, and Random Forest) were used in two different scenarios for the prediction of fetal death, i.e., prenatal care variables up to the first prenatal visit (scenario 1) and prenatal care variables up to just before delivery (scenario 2). The dataset used was composed of 472,004 records labeled live and 15,322 records labeled stillborn records for scenario 1, and 485,966 records labeled live and 1360 records labeled stillborn for scenario 2. The results for the prediction of fetal death identified the Random Forest as the best model with an Area Under the Curve (AUC) of 63% for scenario 1 and a 71% AUC for the gradient boosted model in scenario 2. It was also possible to identify the most important attributes in the analysis, i.e., gestational age, hypertension, severe pre-eclampsia or eclampsia, and maternal age.

Malacova et al. [22] identified the factors that contribute to the prediction of fetal death and evaluated the performance of different machine learning models. Using data sourced from the Data Linkage Branch of the Western Australia Depart-

ment of Health, the dataset comprised 952,813 pregnancy records from 1980 to 2015. 947,025 of the records were labeled live and 5788 were labeled stillbirth. The grid search technique was used with the k-fold cross-validation technique (k-fold = 10) to configure five models—Regularized Logistic Regression, Decision Trees, Random Forest, Extreme Gradient Boosting (XGBoost), and a Multilayer Perceptron Neural Network (MLP). The AUC results of the models varied between 0.59 (CI95%; 0.58; 0.60) and 0.84 (CI95%; 0.83; 0.85). XGBoost and MLP exhibited the best performance. The most influential attributes in the prediction were pregnancy complications, congenital anomalies, maternal characteristics, and medical history.

The work by Ko et al. [14] performed a statistical analysis trends of multiple birth rates and fetal/neonatal/infant mortalities based on the number of gestations in Korea. The dataset used in the study comprised 41,214 fetal death records from the Korean Statistical Information Service. Logistic regression was used to identify the impact of gestational age on mortality in single or multiple pregnancies. Results showed higher fetal mortality rates for multiple pregnancies compared to single pregnancies and identified a higher risk of fetal death during the third trimester of a multiple pregnancy.

Koivu and Sairanen [20] proposed risk models to predict early and late term fetal deaths, as well as premature births, using two large United States (US) pregnancy databases sourced from the National Center of Health Statistics via their National Vital Statistics System (CDC) and the New York City Department of Health and Mental Hygiene (NYC). The CDC dataset comprised 11,901,611 records labeled normal pregnancies, 946,301 records labeled premature births, 7924 records labeled early stillbirths, and 8310 records labeled late stillbirths. The NYC dataset comprised 266,419 records labeled normal pregnancies, 19,203 records labeled premature births, 139 records labeled early stillbirths, and 110 records labeled late stillbirths. Classification models were developed using four different algorithms—logistic regression, gradient boosting decision trees, and two ANNs—a leaky-ReLU-based deep two-layer feed-forward neural network and deep feed-forward self-normalizing neural network based on the Scaled Exponential Linear Units (SELU) activation function. AUC was used to assess the effectiveness of the models. Performance ranged from 0.54 to 0.76; the SELU-based exhibited the best performance in predicting early stillbirth with an AUC of 0.76, while the leaky-ReLU-based ANN performed better for predicting late stillbirth with 0.63 AUC. The models were trained using various attributes, including social information, health, family history, and maternal habits. The results showed that the developed risk models were more effective in predicting early fetal deaths than late fetal deaths or premature births.

Our work contributes to the existing literature by examining data from the PMCP social project, which serves multiple cities across the State of Pernambuco. This approach offers a novel empirical context and perspective on the prediction of fetal death using machine learning. Therefore, investigating the clinical and socio-demographic data of Pernambuco is essential to mitigate this social problem in the future and contributing to Brazil's commitment to the SDGs. Our results provide insights into using machine learning with the PMCP dataset and evaluate the signif-

importance of the attributes used and identify the tree-based models that would be most effective in this scenario.

3 Background

3.1 Machine Learning Models

Machine learning is an area of artificial intelligence that encompasses methods that allow machines to train and learn from provided datasets. In this learning process, the model is allowed to learn to make decisions autonomously using sets of input and output data [25–27]. In this work, four tree-based machine learning models are used. The models evaluated for the prediction of fetal death are Decision Trees, Random Forest, AdaBoost, and XG Boost.

A decision tree model is a supervised machine learning algorithm that supports decision-making that can be used as a classification tree (to predict classes) or a regression tree (to predict numerical values). The structure of a decision tree is very similar to that of a flowchart, with steps that are easy to visualize and thus understand the conditions and probabilities that lead to results. The decision tree model consists of a root node (the most important node), internal nodes (nodes that are related to each other by a hierarchy), and leaf nodes (end results). The internal nodes split the dataset into smaller subsets based on the values of the selected feature. The internal nodes split the dataset into smaller subsets and each leaf node represents a numerical value for a regression problem [28, 29].

Random Forest is an ensemble model that combines multiple decision trees to improve prediction performance. It works by creating a set of decision trees using different subsets of the training data, and then averaging their predictions to make a final prediction. The random selection of features reduces the correlation between trees and results in a diverse set of trees with a lower probability of overfitting the data [30].

AdaBoost is a model that repeats the learning process and generates a final classifier that weighs the weak combinations of the model. This model is particularly effective at boosting the performance of weak classifiers and has the advantage of being able to be used on large datasets with many attributes [30].

The XGBoost model is a tree-based machine learning model that works by creating a set of decision trees iteratively, where each tree tries to correct the errors made by the previous trees. This technique has been effective in various machine learning tasks such as regression and classification [31].

3.2 Evaluation Metrics

For the evaluation of the model learning for predicting fetal death, quantitative metrics based on a confusion matrix were used. The confusion matrix presents the number of records classified correctly and incorrectly and is comprised of True Positive, False Positive, True Negative, and False Negative values [32].

Accuracy is widely used in extant research as a general measure of model performance [33]. This metric is based on the total ratio of samples correctly predicted by the classifier with the test data. In this scenario, the metric seeks to present the generalization capacity of the model. Accuracy is calculated by the equation:

$$\text{accuracy} = \frac{TP + TN}{TP + TN + FP + FN}. \quad (1)$$

Precision measures how many cases are classified by the model as positive and are truly positive in relation to all positive cases [34]. It is calculated using the following equation:

$$\text{precision} = \frac{TP}{TP + FP}. \quad (2)$$

Recall (also referred to as sensitivity) is the ratio of positive cases that were correctly classified by the model [35] and is defined as

$$\text{recall} = \frac{TP}{TP + FN}. \quad (3)$$

Specificity seeks to determine the proportion of actual negatives that were correctly predicted [35]. It is calculated using the following equation:

$$\text{specificity} = \frac{TN}{TN + FP}. \quad (4)$$

The f1-score is a metric that calculates the harmonic mean of two metrics (recall and precision) to calculate the total hit rate of the positive and negative classes performed by the model [36]. It is calculated as

$$\text{f1 - score} = 2 \times \frac{\text{precision} \times \text{recall}}{\text{precision} + \text{recall}}. \quad (5)$$

3.3 Hyper-parameter Optimization and Data Balancing

To improve the performance of the models, the grid search technique is used which seeks to define the best combination of hyper-parameters of a given model around

an analyzed problem based on a grid of initial parameters. Hyper-parameters are parameters used to configure the models such as the learning rate or the minimum number of samples that must exist in each leaf of a tree model. The execution of the technique results in a model that directly impacts its performance in data analysis [37].

Data imbalance is one of the obstacles that hinder learning in classification algorithms as it can lead to a learning bias. Where there is learning bias, the model will learn more about the majority class than the minority resulting in low-performance models due to the imbalance between classes [38, 39]. One of the ways to resolve this problem is to use the random undersampling technique, a heuristic method that randomly eliminates instances of the majority class until the quantity is reduced to the same quantity or the next minority class [40].

4 Materials and Methods

4.1 Dataset

This study utilized a dataset provided by the PMCP, covering the period from 2012 to 2022. Initially, the dataset contained 231,505 records and 71 attributes. It provides extensive information on various aspects of pregnant women's health including maternal history, comorbidities, socio-demographic factors, prenatal and postpartum care, residential and healthcare unit data, personal informative dates, and newborn information. These information was collected by a health specialist at the time of care for the pregnant woman. The dataset's multifaceted variables provide a comprehensive view of the health status and background of pregnant women receiving care from the PMCP thus providing a valuable resource for developing predictive models aimed at enhancing prenatal, postpartum, and maternal health outcomes.

To better understand the dataset, a dictionary¹ was created to describe the attributes based on the information provided by PMCP. The STILLBIRTH attribute was chosen as the target class and named TARGET; it is described with a value of 1 for fetal death and a value of 0 for survival.

4.2 Data Pre-processing

To enable the machine learning models to utilize the dataset provided by the PMCP, it is essential to undertake a set of pre-processing steps to clean and prepare the data for model training and testing. By performing these pre-processing steps, we can ensure that the dataset is suitable for use in training accurate and robust predictive models

¹ Available at: <https://www.dropbox.com/scl/fi/33utcbk29kcc11log4332/Dictionary-dataset-full-MorteFetal.docx?dl=0&rlkey=gbpuo7a4dgrl0r0tjqsjxvn>.

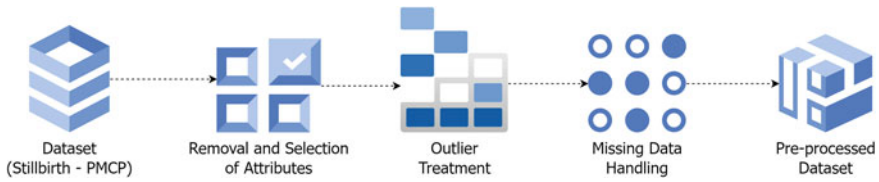


Fig. 2 Pre-processing steps performed on PMCP dataset

and consequently leading to better decision support for healthcare professionals. Figure 2 illustrates the steps involved in generating the pre-processed dataset used in this work.

During the attribute removal and selection stage, we began by excluding attributes related to residential data, service units, geographic environment codes, and other complementary information deemed irrelevant to the study. This step allowed us to streamline the dataset and remove extraneous variables that could potentially interfere with the accuracy and efficacy of the predictive models.

Subsequently, we removed attributes that contained more than 35% of missing values as well as those with low information content regarding the pregnant woman and the puerperium. This step allowed us to further refine the dataset by eliminating variables that could potentially introduce bias or noise into the predictive models.

Following these pre-processing steps, the resulting dataset was further reduced to 17 attributes containing information solely about the mother, current pregnancy, and family health history, as summarized in Table 1.

The next step in our analysis involved the assessment of the selected attributes for completeness and the treatment of outliers. We observed that the PREVIOUS_WEIGHT attribute contained several typing errors. This prompted us to define a maximum weight of 120 kg; any records that exceeded this value were marked as missing, to be treated in the subsequent pre-processing step. Similarly, we noted that the FIRST_PRENATAL attribute exhibited exceptionally high weekly values that did not accurately reflect the timing of the first prenatal care. Upon closer examination, we discovered that this attribute depended on the dates of pregnancy onset and first prenatal care thus inaccuracies in either date could affect the value in weeks of the first prenatal care. To address this issue, we established a maximum value of 35 weeks for the first prenatal care which corresponds to the eighth month of pregnancy. Any records found to be older than 35 weeks were also marked as missing and were designated to be handled in the subsequent step of pre-processing.

In the missing data handling step, we examined the 17 selected attributes and identified five attributes with missing data: PREVIOUS_WEIGHT, GESTATIONAL_RISK, SCHOOLING, AGE, and FIRST_PRENATAL. Of these, PREVIOUS_WEIGHT had the highest proportion of missing data (34.78%) which was close to the previously established threshold. AGE was the second most affected attribute (14.61% missing values) followed by GESTATIONAL_RISK (9.23%). The SCHOOLING attribute had the lowest proportion of missing data at 2.51%. To han-

Table 1 Dataset attributes

Attribute	Type	Description
PREVIOUS_WEIGHT	Numeric	Weight of the pregnant woman before pregnancy
GESTATIONAL_RISK	Numeric	High-risk pregnancy
SCHOOLING	Numeric	Pregnant woman's school level
HAS_HYPERTENSION	Categorical	Pregnant woman with hypertension
HAS_DIABETES	Categorical	Pregnant woman with diabetes
HAS_PELVIC_SURGERY	Categorical	Pregnant woman with pelvic surgery
HAS_URINARY_INFECTION	Categorical	Pregnant woman has urinary tract infection
HAS_CONGENITAL_MALFORMATION	Categorical	Family history of congenital malformation
HAS_FAMILY_TWINSHIP	Categorical	Family history of twins
AMOUNT_GESTATION	Numeric	Total number of pregnancies
AMOUNT_ABORTION	Numeric	Total number of abortions
AMOUNT_DELIVERIES	Numeric	Total number of deliveries
AMOUNT_CESAREAN	Numeric	Total number of cesarean deliveries
TARGET	Numeric	Birth or fetal death
AGE	Numeric	Pregnant's age
FIRST_PRENATAL	Numeric	First prenatal week
TIME_BETWEEN_PREGNANCIES	Numeric	Time in months between pregnancies

For handling the missing data, we adopted the median imputation technique which involves replacing missing values with the median value of the corresponding attribute [23]. By using this method, we were able to preserve the distribution and statistical properties of the data and ensure that the imputed values were consistent with the available data.

After completing the pre-processing steps, a new dataset was generated comprising 17 attributes and 231,505 records. Of these records, 224,076 related to live births and 7429 related to fetal deaths. The pre-processed dataset was then used to train and test the machine learning models for predicting pregnancy outcomes.

4.3 Experiment Design

Figure 3 outlines the methodology used to conduct our experiments. All tests were conducted using the Google Colab tool. As previously mentioned, the initial step was aimed at addressing the issue of data imbalance related to the target attribute. To solve this problem, we utilized the random undersampling approach to randomly select data from the majority class (live birth) and balance the dataset. After balancing the dataset, there were 7429 records for both live births and fetal deaths, 14,858 records in total.

Following the creation of the balanced dataset, we partitioned the dataset into two disjoint subsets: 70% of the data was allocated to the training set and the remaining 30% allocated to the test set. The test set was reserved exclusively for evaluating the performance of the models in the final stage, while the training set was used to train the models.

The grid search technique with 10-k-fold and accuracy as score was employed to determine the optimal hyper-parameters for each of the models.

The hyper-parameters of the four models used in this work (Decision Tree, Random Forest, AdaBoost, and XGBoost) in the grid search can be viewed in Table 2.

After executing the grid search, we obtain the optimal hyper-parameters for each model. We then proceeded to the model evaluation phase, where the test data that was set aside previously was utilized. To quantitatively evaluate the models, we used the evaluation metrics mentioned earlier in Sect. 3: accuracy, precision, sensitivity,

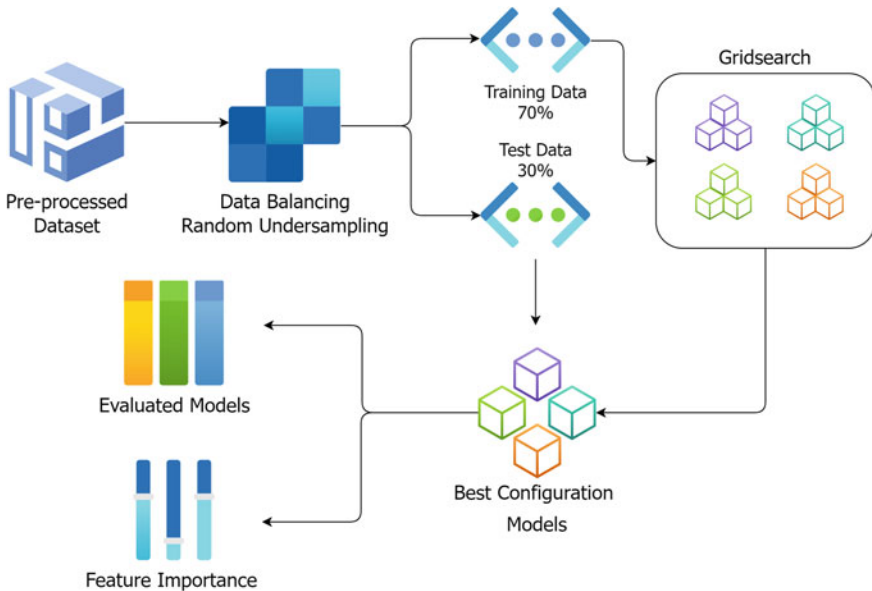


Fig. 3 Experiment design methodology

Table 2 Hyper-parameters used in the grid search

Model	Hyper-parameters	Values
Decision Tree	criterion splitter max_depth min_samples_leaf min_samples_split	['gini', 'entropy'] ['best', 'random'] [None, 1, 3, 5] [1, 3, 5, 7, 9, 11] [2, 5, 8]
Random Forest	n_estimators criterion max_depth min_samples_leaf min_samples_split bootstrap	[100, 500, 700, 900] ['entropy', 'gini'] [None, 1, 3, 5] [1, 3, 5, 7, 9, 11] [2, 5, 8] [True, False]
AdaBoost	n_estimators learning_rate	[50, 100, 200] [0.1, 1, 1.1]
XGBoost	max_depth learning_rate n_estimators gamma	[None, 1, 3, 5] [0.1, 1, 1.1] [50, 100, 200] [0, 0.2, 0.4, 0.8]

specificity, and f1-score. In addition, an analysis was performed to determine the attributes that have the most impact on the learning process of the tree models. This contributes to better understanding the importance of each attribute in the overall model performance.

5 Results and Discussions

5.1 Models' Performance

Table 3 displays the hyper-parameters selected by grid search as the optimal hyper-parameters for the models. Accuracy was utilized as the metric to evaluate the performance and models demonstrated accuracy ranging from 59.55 to 61.95%.

After applying the results chosen by the grid search, all models presented relatively close results as presented in Table 4. Similarly, all models presented similar performance in testing. The XGBoost presented the highest precision when compared to other models (64.02%) while the Decision Tree at 61.93% presented the lowest precision in this experiment.

Regarding sensitivity and specificity, Random Forest demonstrated disparity in these metrics. The model exhibited a sensitivity of 67.86% indicating that it can accurately predict the probable fetal death, the target class of this experiment. However, there was a slight decrease in specificity (59.62%). This suggests a possible challenge in predicting live births and may result in an increase in false positives.

Table 3 Grid search results of the models

Model	Best hyper-parameters	Accuracy (%)
Decision Tree	criterion: 'gini', splitter: 'best', max_depth: 5, min_samples_leaf: 11, min_samples_split: 2	59.55
Random Forest	n_estimators: 900, criterion: 'entropy', max_depth: None, min_samples_leaf: 11, min_samples_split: 2 Bootstrap: True	61.43%
AdaBoost	n_estimators: 50, learning_rate: 1	61.72
XGBoost	max_depth: 5 learning_rate: 0.1 n_estimators: 50, gamma: 0.8	61.95

Table 4 Model performance results

Model	Precision (%)	Sensitivity (%)	Specificity (%)	Accuracy (%)	f1-score (%)
Decision Tree	61.93	62.55	61.30	61.93	61.93
Random Forest	63.84	67.86	59.62	63.80	63.72
AdaBoost	62.74	62.55	62.94	62.74	62.74
XGBoost	64.02	66.05	61.94	64.02	64.00

Despite this disparity, the Random Forest model remained consistent with an overall accuracy of 63% and other evaluation metrics.

Of all the models used in this experiment, XGBoost stood out as with the best performance metrics when compared to the other models. This model achieved the best results in precision (64.02%), accuracy (64.02%), and f1-score (64%). It also ranked second best in sensitivity (66.05%) and specificity (61.94%) metrics. Among the models tested, Decision Tree had the lowest overall performance. Given that the other tested models are improved versions of trees, this most likely explains why the Decision Tree exhibited slightly lower metrics than the other tree-based models. As presented previously, Random Forest exhibited superior sensitivity compared to the other models, while XGBoost achieved the highest f1-score. In general, all models displayed consistent performance with similar performance.

Figure 4 presents a comparison of the metrics among all models. The shape of the radar graph can provide insights into the performance of different models. A model with consistently high performance across all metrics creates a regular, symmetrical

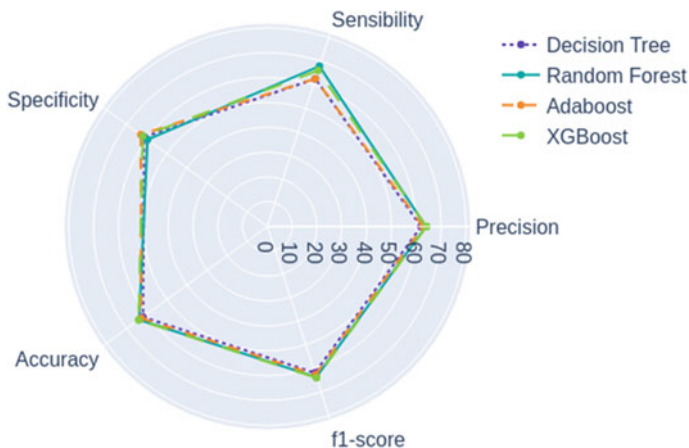


Fig. 4 Comparison in percentage between model metrics

shape, while a model with significant variations in performance creates an irregular, non-symmetrical shape. Patterns in the shape can also suggest features of strength or weakness for a model, such as a model that performs well in certain metrics but poorly in others.

5.2 Attributes' Importance

In this study, we also identified the attributes that most influenced the learning process of the models. The importance of attributes lies in their ability to capture relevant information about the data that is useful for prediction. Choosing the right attributes for a particular problem is critical to achieving good performance. Identifying the most relevant attributes often involves a combination of domain expertise and experimentation with different attribute subsets. Figure 5 presents the eight attributes that were most important in this process. These attributes are primarily related to the pregnant woman's socio-demographic information and medical history.

Specifically, data from the first prenatal visit, age, time between pregnancies, and pre-pregnancy weight had a significant influence on all models. On the other hand, the education level and number of abortions had a relatively lesser impact on all the models. Interestingly, the attributes of hypertension and gestational risk had no impact in the Decision Tree model but were found to be influential in the Random Forest, XGBoost, and AdaBoost models. These findings are consistent with a recent Systemic Literature Review (SLR) conducted by Silva Rocha et al. [23] which suggests attributes such as maternal age, mother's education, prenatal care, number of pregnancies, and number of cesarean deliveries were used in studies predicting fetal

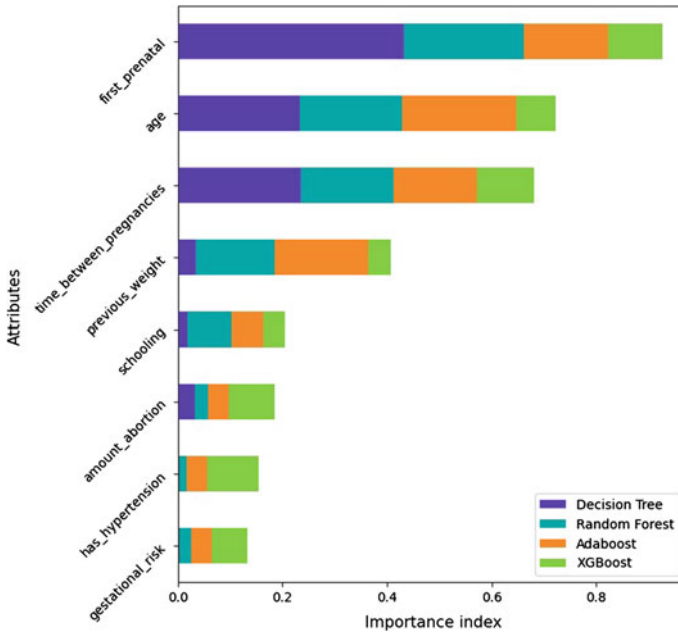


Fig. 5 Most important attributes in the model learning process

death. In Muin et al. [41], maternal demographic and obstetric characteristics were the ones that best explained prediction of women at risk for stillbirth.

To analyze the data distribution between live births and fetal deaths, we focused on the three most important attributes in the models’ learning. Figure 6 shows the data from the start of prenatal care, ranging from gestational week 1 to week 35. Notably, most of the records were concentrated in week 10, which is attributed to the use of median imputation to fill in the missing data. Our findings indicate that in cases where prenatal care began in weeks 5 to 9, the incidence of fetal deaths was considerably higher.

It is extremely critical that mothers have adequate and appropriate prenatal care to optimize the likelihood of a positive outcome in pregnancy. Studies associate inadequate prenatal care with an increased rate of fetal deaths [42, 43]. The lack of prenatal visits or visits without proper monitoring can increase fetal deaths. Conditions such as premature rupture of membranes, fetal growth restriction, and bleeding that can be detected with proper monitoring can prevent negative outcomes [43]. A systematic review conducted by Townsend et al. [44] revealed a total of 69 studies reporting on 64 different variables that were relevant to the development of stillbirth prediction models. Among these variables, the most frequently cited ones included maternal age, Body Mass Index (BMI), and previous history of stillbirth and diabetes. These results can provide important insights for healthcare providers in identifying high-

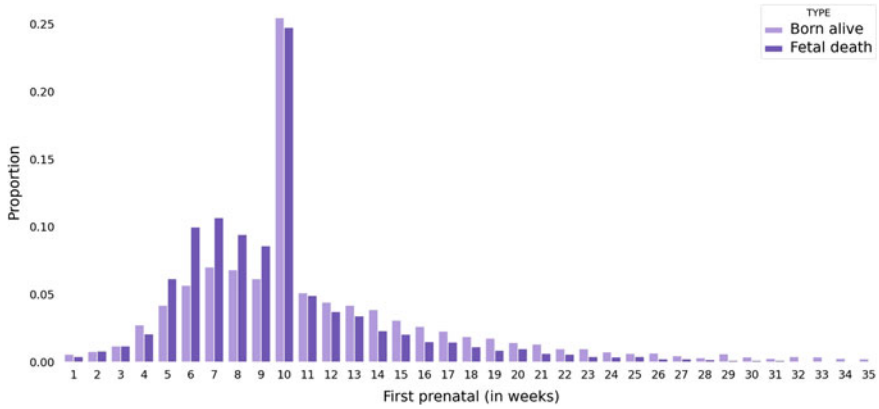


Fig. 6 Distribution of prenatal data between live births and fetal deaths

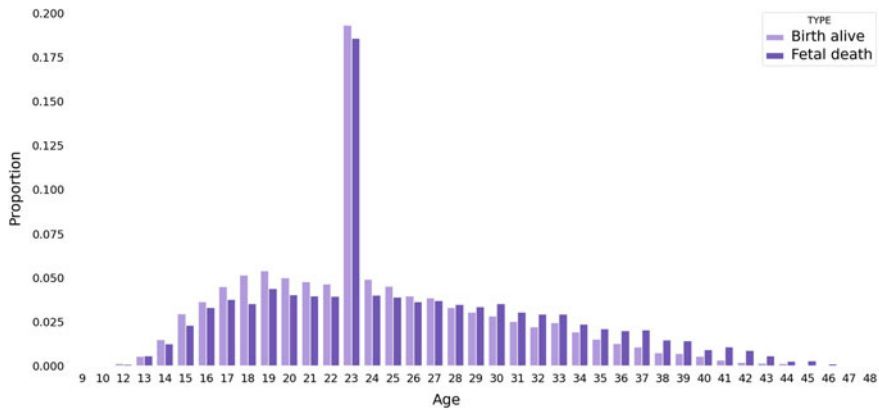


Fig. 7 Distribution of maternal age data between live births and fetal deaths

risk pregnancies and implementing targeted interventions to reduce the occurrence of fetal deaths.

Figure 7 displays the distribution of data relating to fetal deaths and live births based on maternal age. The highest concentration of records is observed at 23 years of age, again potentially attributed to the technique of imputing missing values through median substitution. Notably, from the age of 28, the frequency of fetal death data exceeds that of live births. A correlation is observed between maternal age and the disparity between the number of live birth and fetal death records, with an increasing maternal age demonstrating a wider discrepancy.

The risk of fetal death has been shown to increase with advancing maternal age (AMA), which may be attributed to the higher incidence of chronic diseases such as diabetes and hypertension in this population [45]. Several studies have reported that advanced maternal age, typically defined as 32 years or older, is a significant risk

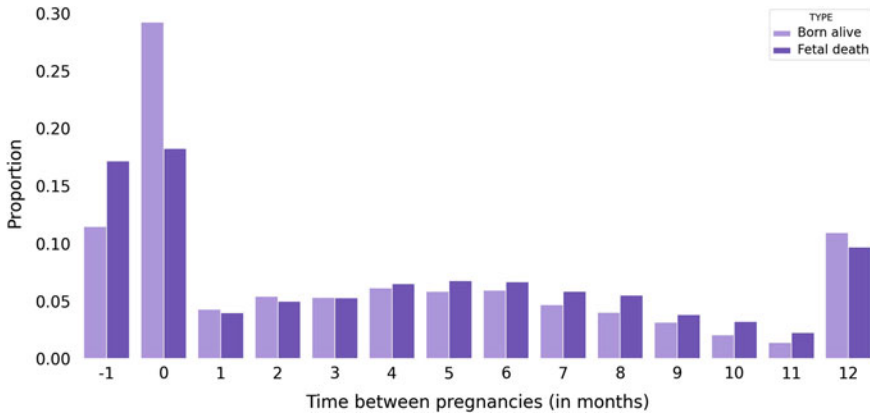


Fig. 8 Distribution of data by time (months) between pregnancies for live births and fetal deaths

factor for fetal death, with ectopic pregnancy being one of the primary contributors to this association. In this age group, the chances of spontaneous abortion are also elevated [46]. AMA has been found to be a significant predictive factor in several studies, including those using Decision Tree models, for predicting fetal death and prematurity [47–49]. Our study also identified AMA as an important attribute in predicting fetal death.

Despite the increased risk of fetal loss associated with advanced maternal age, the use of assisted reproductive technology (ART) has enabled older women to conceive. To ensure the best possible outcome, it is crucial for women to have accurate information about the potential risks and make informed decisions about their health and pregnancy. Studies suggest that appropriate prenatal monitoring and adoption of a healthy lifestyle can improve the health outcomes of older pregnant women [50].

Another important factor for predicting fetal death is the interval between pregnancies (Fig. 8). The distribution of this data ranges from -1 (records that we were unable to identify the time between pregnancies) up to 12 months (where the time between pregnancies is at least 12 months). The records classified at 0 months are those that had no interval between one pregnancy and another.

A significant difference was identified between the number of live births and fetal deaths classified as -1 . Unfortunately, it was not possible to determine the duration of the pregnancy for these records. Additionally, we observed a high number of live births within a 0-month interval indicating that some women were able to carry the pregnancy to full term and achieve a positive outcome despite a short time between pregnancies. Between month 4 and month 11 the proportion of fetal deaths exceeded the proportion of live births. According to WHO, the recommended time for having a new pregnancy safely is 24 months. A shorter time than this period increases the risk of fetal, perinatal, and infant death [51].

The Interpregnancy Interval (IPI) is a measure of time between a woman’s previous delivery and the next conception. IPI is calculated by subtracting the date of the

previous delivery from the mother's last menstrual period. Studies have shown that an IPI of less than 6 months is associated with an increased risk of adverse outcomes such as premature birth, low birth weight, and fetal death [52]. Further, short IPIs may be associated with women who had a pregnancy loss in the previous gestation. With a short time period between a previous pregnancy and a new one, the woman's body is more likely to enter a reproductive cycle poor in nutrients during the pre-conception period, a factor associated with fetal growth restriction and congenital anomalies [53].

6 Conclusion and Next Steps

In several states in Brazil, social programs have been initiated to focus on maternal, fetal, and child care, providing not only clinical health support for mothers and babies but also psychological support and a network of assistance. These initiatives aim to prevent fetal deaths and improve the well-being of mothers and their babies. The Programa Mãe Coruja Pernambucana is a crucial initiative that reaches out to hundreds of families across more than a hundred cities in the state of Pernambuco. By conducting studies on fetal death, we can further assist and strengthen such programs to combat this social problem and minimize adverse outcomes in the lives of pregnant women and babies.

In the present work, machine learning models were used to predict fetal death and can be considered a promising tool in monitoring the maternal health and supporting clinical decision-making. Specifically, we utilized four tree-based models in our analysis—Decision Tree, Random Forest, AdaBoost, and XGBoost. Of these models, XGBoost demonstrated the best performance in terms of evaluative metrics, consistently achieving values between 61 and 66%, while exhibiting good sensitivity.

We also evaluated the importance of the attributes used in the models' learning process. In our study, socio-demographic information about the mother and health history were essential in the learning process. Data such as the start of prenatal care, maternal age, and time between pregnancies were important factors in this study. Laboratory data was not used in this study. Instead, all information used in the models was based on the pregnant woman and her family's inherent information. This decision was made with the aim of simplifying the data and avoiding the need for costly laboratory tests during the learning process. The approach used in this study is therefore considered to be of low cost and practical.

We emphasize that this study presented some preliminary results using the PMCP database. The identification of fetal deaths in regions with lower levels of socio-economic development, such as the Brazilian Northeast, is of paramount importance due to the likelihood of under-reporting of these events, limited access to quality healthcare services, and elevated maternal and infant morbidity and mortality rates linked to social determinants [54].

The usage of machine-learning-based systems for diagnostic, prognostic, and health assessment may allow a better performance of professionals to take their

decisions. Our work aims to assist health professionals in predicting fetal death; we do not aim to diagnose but to use the predictive model as an auxiliary decision support tool. Despite the limitations posed by incomplete data and limited information in the database, we were able to achieve promising results in terms of evaluation metrics.

As part of our future work, we plan to refine our methodology by improving the selection of attributes and exploring different techniques for handling missing data. Another critical aspect that can be considered is the impact of social and behavioral variables. For instance, situations where women experience domestic violence, stress, unemployment, and deprivation can significantly affect their health and well-being, and could be taken into account within a population-based conceptual framework [41]. We recognize that there is still much room for improvement and future studies could benefit from a more comprehensive datasets. Nonetheless, our current findings provide an encouraging starting point for further research into the detection of fetal death using predictive modeling.

Integrating machine learning solutions into clinical practice can be particularly beneficial in supporting obstetric counseling and prenatal care, especially in countries that face economic vulnerability and social fragility, improving maternal and fetal health outcomes. By leveraging advanced analytical tools and combining them with clinical expertise, we plan to develop more accurate and effective predictive models that can aid in the prediction and prevention of fetal death.

Acknowledgements This work was funded by Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq), Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES), Fundação de Amparo à Ciência e Tecnologia do Estado de Pernambuco (FACEPE), and Universidade de Pernambuco (UPE). Patricia Takako Endo is funded by CNPq via PQ 2 grant.

References

1. UNICEF: A neglected tragedy: the global burden of stillbirths. Report of the UN Inter-agency Group for Child Mortality Estimation;. www.unicef.org/reports/neglected-tragedy-global-burden-of-stillbirths-2020. Accessed 28 Apr. 2023
2. WHO: World Health Organization. Stillbirths. WHO, Geneva (2023). www.who.int/health-topics/stillbirth#tab=tab_1. Accessed 28 Apr. 2023
3. da Saúde, M.: Manual de vigilância do óbito infantil e fetal e do Comitê de Prevenção do Óbito Infantil e Fetal (2009). https://bvsm.s.saude.gov.br/bvsm/publicaç~oes/manual_obito_infantil_fetal_2ed.pdf. Accessed 28 Apr. 2023
4. Organization WH: SDG target 3.1: Maternal mortality. www.who.int/data/gho/data/themes/topics/sdg-target-3-1-maternal-mortalityTarget. Accessed 28 Apr. 2023
5. Organization WH: SDG target 3.2: Newborn and child mortality. www.who.int/data/gho/data/themes/topics/indicator-groups/indicator-group-details/GHO/sdg-target-3.2-newborn-and-child-mortality. Accessed 28 Apr. 2023
6. Khan, M.N., Rahman, M.M., Shariff, A.A., Rahman, M.M., Rahman, M.S., Rahman, M.A.: Maternal undernutrition and excessive body weight and risk of birth and health outcomes. *Archi. Public Health* **75**, 1–10 (2017)

7. Aminu, M., Bar-Zeev, S., van den Broek, N.: Cause of and factors associated with stillbirth: a systematic review of classification systems. *Acta obstetrica et gynecologica Scandinavica*. **96**(5), 519–28 (2017)
8. Quibel, T., Bultez, T., Nizard, J., Subtil, D., Huchon, C., Rozenberg, P.: Morts fœtales in utero. *Journal de Gynécologie Obstétrique et Biologie de la Reproduction*. **43**(10), 883–907 (2014)
9. Yan, J., Guo, J., Fan, C., Juan, J., Yu, X., Li, J., et al.: Coronavirus disease 2019 in pregnant women: a report based on 116 cases. *Am. J. Obstet. Gynecol.* **223**(1), 111-e1 (2020)
10. Lambelet, V., Vouga, M., Pomar, L., Favre, G., Gerbier, E., Panchaud, A., et al.: SARS-CoV-2 in the context of past coronaviruses epidemics: consideration for prenatal care. *Prenat. Diagn.* **40**(13), 1641–54 (2020)
11. Barbeiro, F.M.D.S., Fonseca, S.C., Tauffer, M.G., Ferreira, M.D.S.S., Silva, F.Pd., Ventura, P.M., et al.: Fetal deaths in Brazil: a systematic review. *Revista de Saúde Pública* **49**, 22 (2015)
12. de Santana, T.C.P., da Silva, L.M., da Silva, L.R.F.G., Rocha, L.M., Canhoto, C.T.S., da Silva, A.C.F.A., et al.: Dificuldades dos enfermeiros no atendimento ao pré-natal de risco habitual e seu impacto no indicador de morbimortalidade materno-neonatal. *Revista Eletrônica Acervo Saúde*. **20**, e711-1 (2019)
13. Norwitz, E.R., Edusa, V., Park, J.S.: Maternal physiology and complications of multiple pregnancy. In: *Seminars in Perinatology*, vol. 29, pp. 338–348. Elsevier (2005)
14. Ko, H.S., Wie, J.H., Choi, S.K., Park, I.Y., Park, Y.G., Shin, J.C.: Multiple birth rates of Korea and fetal/neonatal/infant mortality in multiple gestation. *PLoS One* **13**(8), e0202318 (2018)
15. Marques, L.J.P., Silva, Z.P.D., Alencar, G.P., Almeida, M.Fd.: Contribuições da investigação dos óbitos fetais para melhoria da definição da causa básica do óbito no Município de São Paulo, Brasil. *Cadernos de Saúde Pública* **37**, e00079120 (2021)
16. Oliveira, C.M.D., Bonfim, C.V.D., Guimarães, M.J.B., Frias, P.G., Medeiros, Z.M.: Mortalidade infantil: tendência temporal e contribuição da vigilância do óbito. *Acta paulista de enfermagem* **29**, 282–290 (2016)
17. da Saúde, M.: Sistema de Informações de Mortalidade—SIM. <http://tabnet.datasus.gov.br/cgi/tabcgi.exe?sim/cnv/fet10uf.def>. Accessed 28 Apr. 2023
18. da Saúde, M.: Institui a rede cegonha. https://bvsms.saude.gov.br/bvs/saudelegis/gm/2011/prt1459_24_06_2011.html. Accessed 28 Apr. 2023
19. Lipschuetz, M., Guedalia, J., Rottenstreich, A., Persky, M.N., Cohen, S.M., Kabiri, D., et al.: Prediction of vaginal birth after cesarean deliveries using machine learning. *Am. J. Obstet. Gynecol.* **222**(6), 613-e1 (2020)
20. Koivu, A., Sairanen, M.: Predicting risk of stillbirth and preterm pregnancies with machine learning. *Health Inform. Sci. Syst.* **8**, 1–12 (2020)
21. Sheikhtaheri, A., Zarkesh, M.R., Moradi, R., Kermani, F.: Prediction of neonatal deaths in NICUs: development and validation of machine learning models. *BMC Med. Inform. Decis. Making* **21**(1), 1–14 (2021)
22. Malacova, E., Tippaya, S., Bailey, H.D., Chai, K., Farrant, B.M., Gebremedhin, A.T., et al.: Stillbirth risk prediction using machine learning for a large cohort of births from Western Australia, 1980–2015. *Sci. Rep.* **10**(1), 1–8 (2020)
23. Silva Rocha, E.D., de Moraes Melo, F.L., de Mello, M.E.F., Figueiroa, B., Sampaio, V., Endo, P.T.: On usage of artificial intelligence for predicting mortality during and post-pregnancy: a systematic review of literature. *BMC Med. Inform. Decis. Making* **22**(1), 1–17 (2022)
24. Shukla, V.V., Eggleston, B., Ambalavanan, N., McClure, E.M., Mwenechanya, M., Chomba, E., et al.: Predictive modeling for perinatal mortality in resource-limited settings. *JAMA Netw. Open* **3**(11), e2026750 (2020)
25. Jakhar, D., Kaur, I.: Artificial intelligence, machine learning and deep learning: definitions and differences. *Clin. Exp. Dermatol.* **45**(1), 131–2 (2020)
26. Pesapane, F., Codari, M., Sardanelli, F.: Artificial intelligence in medical imaging: threat or opportunity? Radiologists again at the forefront of innovation in medicine. *Eur. Radiol. Exp.* **2**, 1–10 (2018)
27. Helm, J.M., Swiergosz, A.M., Haerberle, H.S., Karnuta, J.M., Schaffer, J.L., Krebs, V.E., et al.: Machine learning and artificial intelligence: definitions, applications, and future directions. *Curr. Rev. Musculoskelet. Med.* **13**, 69–76 (2020)

28. Dietterich, T.G.: An experimental comparison of three methods for constructing ensembles of decision trees: bagging, boosting, and randomization. *Mach. Learn.* **40**, 139–57 (2000)
29. Murdoch, W.J., Singh, C., Kumbier, K., Abbasi-Asl, R., Yu, B.: Interpretable machine learning: definitions, methods, and applications (2019). arXiv preprint [arXiv:1901.04592](https://arxiv.org/abs/1901.04592)
30. Breiman, L.: *Classification and Regression Trees*. Routledge (2017)
31. Mitchell, R., Frank, E.: Accelerating the XGBoost algorithm using GPU computing. *PeerJ Comput. Sci.* **3**, e127 (2017)
32. Susmaga, R.: Confusion matrix visualization. In: *Intelligent Information Processing and Web Mining*, pp. 107–116. Springer (2004)
33. Hossin, M., Sulaiman, M.N.: A review on evaluation metrics for data classification evaluations. *Int. J. Data Min. Knowl. Manag. Process.* **5**(2), 1 (2015)
34. Olson, D.L., Delen, D.: *Advanced Data Mining Techniques*. Springer Science & Business Media (2008)
35. Parikh, R., Mathai, A., Parikh, S., Sekhar, G.C., Thomas, R.: Understanding and using sensitivity, specificity and predictive values. *Indian J. Ophthalmol.* **56**(1), 45 (2008)
36. Chicco, D., Jurman, G.: The advantages of the Matthews correlation coefficient (MCC) over F1 score and accuracy in binary classification evaluation. *BMC Genomics* **21**, 1–13 (2020)
37. Fuadah, Y.N., Pramudito, M.A., Lim, K.M.: An optimal approach for heart sound classification using grid search in hyperparameter optimization of machine learning. *Bioengineering* **10**(1), 45 (2022)
38. Prati, R.C., Batista, G.E., Monard, M.C.: Class imbalances versus class overlapping: an analysis of a learning system behavior. In: *Mexican International Conference on Artificial Intelligence*, pp. 312–321. Springer (2004)
39. Batista, G.E., Prati, R.C., Monard, M.C.: A study of the behavior of several methods for balancing machine learning training data. *ACM SIGKDD Explor. Newsl.* **6**(1), 20–9 (2004)
40. He, H., Ma, Y.: *Imbalanced Learning: Foundations, Algorithms, and Applications*. Wiley (2013)
41. Muin, D.A., Windsperger, K., Attia, N., Kiss, H.: Predicting singleton antepartum stillbirth by the demographic fetal medicine foundation risk calculator—a retrospective case-control study. *PLoS One* **17**(1), e0260964 (2022)
42. Vintzileos, A.M., Ananth, C.V., Smulian, J.C., Scorza, W.E., Knuppel, R.A.: The impact of prenatal care on neonatal deaths in the presence and absence of antenatal high-risk conditions. *Am. J. Obstet. Gynecol.* **186**(5), 1011–6 (2002)
43. Heaman, M.I., Martens, P.J., Brownell, M.D., Chartier, M.J., Derksen, S.A., Helewa, M.E.: The association of inadequate and intensive prenatal care with maternal, fetal, and infant outcomes: a population-based study in Manitoba, Canada. *J. Obstet. Gynaecol. Canada* **41**(7), 947–59 (2019)
44. Townsend, R., Sileo, F., Allotey, J., Dodds, J., Heazell, A., Jorgensen, L., et al.: Prediction of stillbirth: an umbrella review of evaluation of prognostic variables. *BJOG: An Int. J. Obstet. Gynaecol.* **128**(2), 238–250 (2021)
45. Lean, S.C., Derricott, H., Jones, R.L., Heazell, A.E.: Advanced maternal age and adverse pregnancy outcomes: a systematic review and meta-analysis. *PLoS One* **12**(10), e0186287 (2017)
46. Attali, E., Yogev, Y.: The impact of advanced maternal age on pregnancy outcome. *Best Pract. Res. Clin. Obstet. Gynaecol.* **70**, 2–9 (2021)
47. Guarga Montori, M., Álvarez Martínez, A., Luna Álvarez, C., Abadía Cuchí, N., Mateo Alcalá, P., Ruiz-Martínez, S.: Advanced maternal age and adverse pregnancy outcomes: a cohort study. *Taiwanese J. Obstet. Gynecol.* **60**(1), 119–24 (2021)
48. Amini, P., Maroufzadeh, S., Samani, R.O., Hamidi, O., Sepidarkish, M.: Prevalence and determinants of preterm birth in Tehran, Iran: a comparison between logistic regression and decision tree methods. *Osong Public Health Res. Perspect.* **8**(3), 195 (2017)
49. Singh, A., Bhatia, M., Garg, A.: Prediction of abnormal pregnancy in pregnant women with advanced maternal age and pregestational diabetes using machine learning models. *IEEE* 262–267 (2022)

50. Correa-de Araujo, R., Yoon, S.S.: Clinical outcomes in high-risk pregnancies due to advanced maternal age. *J. Women's Health* **30**(2), 160–7 (2021)
51. Organization WH: Interpregnancy interval: effects on maternal and perinatal health. World Health Organization, Department of Reproductive Health and Research (2013). https://apps.who.int/iris/bitstream/handle/10665/73710/RHR_policybrief_birthspacing_eng.pdf
52. Gupta, P.M., Freedman, A.A., Kramer, M.R., Goldenberg, R.L., Willinger, M., Stoll, B.J., et al.: Interpregnancy interval and risk of stillbirth: a population-based case control study. *Ann. Epidemiol.* **35**, 35–41 (2019)
53. Regan, A.K., Gissler, M., Magnus, M.C., Håberg, S.E., Ball, S., Malacova, E., et al.: Association between interpregnancy interval and adverse birth outcomes in women with a previous stillbirth: an international cohort study. *The Lancet* **393**(10180), 1527–35 (2019)
54. Barros, P.D.S., Aquino, É.C.D., Souza, M.R.D.: Mortalidade fetal e os desafios para a atenção à saúde da mulher no Brasil. *Revista de Saúde Pública* **53** (2019)