The Lancet Regional Health - Americas 3 (2021) 100053

Contents lists available at ScienceDirect

The Lancet Regional Health - Americas

journal homepage: www.elsevier.com/locate/lana

Research paper

Data-driven risk stratification for preterm birth in Brazil: a population-based study to develop of a machine learning risk assessment approach

Thiago Augusto Hernandes Rocha [1,*], Erika Bárbara Abreu Fonseca de Thomaz [2], Dante Grapiuna de Almeida [3], Núbia Cristina da Silva [4], Rejane Christine de Sousa Queiroz [5], Luciano Andrade [6], Luiz Augusto Facchini [7], Marcos Luiggi Lemos Sartori [8], Dalton Breno Costa [9], Marcos Adriano Garcia Campos [10], Antônio Augusto Moura da Silva [11], Catherine Staton [12], João Ricardo Nickenig Vissoci [13]

[1] *Duke Global Health Institute, Duke University, Durham, North Carolina, United States of America*
[2] *Department of Public Health, Federal University of Maranhão, São Luís, Maranhão, Brazil*
[3] *MEDOMAI Technology Information Systems, Belo Horizonte, Minas Gerais, Brazil*
[4] *Duke Global Health Institute, Duke University, Durham, North Carolina, United States of America*
[5] *Department of Public Health, Federal University of Maranhão, São Luís, Maranhão, Brazil*
[6] *Department of Nursing, State University of the West of Parana, Foz do Iguaçu, Parana, Brazil*
[7] *Department of Social Medicine, Federal University of Pelotas, Pelotas, Rio Grande do Sul, Brazil*
[8] *Pontifical Catholic University of Rio Grande do Sul. Porto Alegre, Rio Grande do Sul, Brazil*
[9] *The Federal University of Health Sciences of Porto Alegre. Porto Alegre, Rio Grande do Sul, Brazil*
[10] *Department of Public Health, Federal University of Maranhão, São Luís, Maranhão, Brazil*
[11] *Department of Public Health, Federal University of Maranhão, São Luís, Maranhão, Brazil*
[12] *Duke Emergency Medicine, Duke University Medical Center, Durham, NC USA. Duke Global Health Institute, Duke University, Durham, North Carolina, United States of America*
[13] *Duke Emergency Medicine, Duke University Medical Center, Durham, NC USA. Duke Global Health Institute, Duke University, Durham, North Carolina, United States of America*

ARTICLE INFO

Article history:
Received 10 May 2021
Revised 1 August 2021
Accepted 3 August 2021
Available online 19 August 2021

Keywords:
Preterm Birth
Predictive Value of Tests
Primary Health Care
Appraisal, Health Risk, Machine Learning

ABSTRACT

Background: Preterm birth (PTB) is a growing health issue worldwide, currently considered the leading cause of newborn deaths. To address this challenge, the present work aims to develop an algorithm capable of accurately predicting the week of delivery supporting the identification of a PTB in Brazil.
Methods: This a population-based study analyzing data from 3,876,666 mothers with live births distributed across the 3,929 Brazilian municipalities. Using indicators comprising delivery characteristics, primary care work processes, and physical infrastructure, and sociodemographic data we applied a machine learning-based approach to estimate the week of delivery at the point of care level. We tested six algorithms: eXtreme Gradient Boosting, Elastic Net, Quantile Ordinal Regression - LASSO, Linear Regression, Ridge Regression and Decision Tree. We used the root-mean-square error (RMSE) as a precision.
Findings: All models obtained RMSE indexes close to each other. The lower levels of RMSE were obtained using the eXtreme Gradient Boosting approach which was able to estimate the week of delivery within a 2.09 window 95%IC (2.090–2.097). The five most important variables to predict the week of delivery were: number of previous deliveries through Cesarean-Section, number of prenatal consultations, age of the mother, existence of ultrasound exam available in the care network, and proportion of primary care teams in the municipality registering the oral care consultation.

* Corresponding author: Thiago Augusto Hernandes Rocha, Duke University
E-mail addresses: tar29@duke.edu (T.A.H. Rocha), erika.barbara@ufma.br (E.B.A.F. de Thomaz), dante@medomai.com.br (D.G. de Almeida), nubiacristina@gmail.com (N.C. da Silva), queiroz.rejane@gmail.com (R.C.d.S. Queiroz), luc.and1973@gmail.com (L. Andrade), luizfacchini@gmail.com (L.A. Facchini), marcos.sartori@acad.pucrs.br (M.L.L. Sartori), dalton.bc96@gmail.com (D.B. Costa), marcos.adrianogc@gmail.com (M.A.G. Campos), aamouradasilva@gmail.com (A.A.M. da Silva), catherine.staton@duke.edu (C. Staton), jnv4@duke.edu (J.R.N. Vissoci).

https://doi.org/10.1016/j.lana.2021.100053
2667-193X/© 2021 The Author(s). Published by Elsevier Ltd. This is an open access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/4.0/)

Interpretation: Using simple data describing the prenatal care offered, as well as minimal characteristics of the pregnant, our approach was capable of achieving a relevant predictive performance regarding the week of delivery.

Funding: Bill and Melinda Gates Foundation, and National Council for Scientific and Technological Development – Brazil, (Conselho Nacional de Desenvolvimento Científico e Tecnológico - CNPQ acronym in portuguese) Support of the research project named: Data-Driven Risk Stratification for Preterm Birth in Brazil: Development of a Machine Learning-Based Innovation for Health Care- Grant: OPP1202186

© 2021 The Author(s). Published by Elsevier Ltd.
This is an open access article under the CC BY-NC-ND license
(http://creativecommons.org/licenses/by-nc-nd/4.0/)

Research in context

Evidence before this study

Prediction models may contribute to personalized risk-based management of women at high risk of Preterm birth (PTB). However, the results have not been accurate in most of the studies. Many of the current models have an insufficient detection rate for application in clinical or public health practices. Predictive models with better predictive capacity are needed to locate women at high risk of PTB.

Added value of this study

This is the first study to develop models encompassing data from an entire country at the primary health center level trying to estimate the risk of PTB based on variables from millions of records. Our model was capable of estimating the week of delivery within a 2 weeks window margin of error. The maternal characteristics that are most predictive were number of previous deliveries through C-Section, number of prenatal consultations, age of the mother, existence of ultrasound exam available in the care network, and proportion of primary care teams in the municipality registering the oral care consultation.

Implications of all the available evidence

The model developed has the capability to identify the week of delivery and the chance of PTB. Thus, it would be possible to give an early alert enabling health professionals to intervene with the available resources aiming to avoid the premature delivery. Early interventions could be designed to address risk factors related to PTB. For a policy-maker perspective the tool can help health managers to better manage resources aiming to address the pregnant woman presenting an elevated chance of PTB.

1. Introduction

In 2017, the preterm birth (PTB) rate in Brazil, 11 per 100 live births [1], was the highest in all the Latin America and the Caribbean [2]. As in the rest of the world, the rates of PTB in Brazil have been increasing in the past years, going from 6•5% in 2004 to 10•9% in 2017. This growth corresponds to a 68% increase in cases during the same time period [2–5]. The increase in the PTB rates has not been followed by a similar increase in the neonatal death rate suggesting that the increased PTB may be associated with better care [6]. This hypothesis can be supported by the fact that better care could be responsible to increase the survival chances of a PTB newborn [7–11]. Despite this hypothesis, the PTB is a global public health concern and a leading cause of child death worldwide.

The Universal Health Coverage guidelines suggests that an early detection and adequate risk stratification of PTB could be determinant to ensure quality and equity in preventive care [7–11]. Thus, early detection can be a useful approach to improve child and maternal health in low- and middle-income countries. Early identification of risk factors associated to PTB risk is complex task due to the multifactorial characteristics of the phenomena. Cohort studies and systematic reviews have shown that the prediction models for PTB have moderate discrimination [12,13], especially for nulliparous women [14]. Additionally, it is worth highlighting that the prediction models already developed faced a lack of external validation [14]. This task proves to be even harder in the public health setting where in-depth patient characteristics are not always available. The variability of factors associated with PTB stem from health conditions (e.g. placental dysfunction, maternal diabetes, vascular disease, obesity, infections and inflammations), pregnancy history (e. g. multiple pregnancies, previous PTB, vaginal bleeding), sociodemographic (e. g. low education and socioeconomic level, single marital status, maternal age), health habits (e.g. smoking, alcohol use, physical and emotional stress, work habits), and health care (e.g. quality of prenatal care, number of prenatal care visits) [3,12,14–19]. Such a varied set of predictors thwarts the development of accurate methods to risk stratify pregnant women in the primary health care setting.

Most attempts to predict PTB have focused on maternal health conditions and health habits aligned with biomarkers capacities [20]. Thus, the ability to predict PTB from social determinants, health care characteristics and prenatal care could be foundational to the development of Universal Health Care in Low-and-middle income countries. The Low-and-middle income setting is characterized by a lack of sophisticated information as laboratory tests results, ultrasound data, and anonymized electronic health records [21]. Thus, a pilot test capable of assessing the feasibility of forecasting the chance of a PTB using minimal patient data may be suitable for the challenges faced in most parts of the Low-and-middle income countries.

Few studies have identified the importance of local-specific contextual determinants of PTB, such as the primary health care actions, sociodemographics variables, as well as care access [22–25]. Considering this context, the present work aims to fill this gap examining a pilot approach to test the feasibility of using parameters associated with primary health care, prenatal activities, and sociodemographic aspects to forecast the week of delivery, and thus the chance of a PTB birth.

We hypothesize that the use of current computational power and sophisticated modeling techniques could be used to develop a feasible risk stratification model of PTB at the Primary health care level. Therefore, the main aim of the study is to pilot the development of a Machine Learning approach to predict the week of delivery based on minimal patient data.

2. Methods

This is a secondary data analysis study using predictive algorithms seeking to estimate the week of delivery of pregnant women. To achieve this goal we conducted the study in two steps: (1) we linked three different national level data sources containing relevant predictors of PTB; (2) training and testing of machine learning–based predictive models to identify the best accurate model. The structure of the present manuscript followed the guidelines Developing and Reporting Machine Learning Predictive Models in Biomedical Research [26]. Ethical approval for this study was obtained from the Federal University of Maranhão IRB (Protocol # 3.172.930/2019).

2.1. Study setting and selection of participants

Centered in primary health care, the Brazilian Unified Public Health System (SUS - acronym in Portuguese) was designed to provide Universal Health coverage for all the population [27]. The alignment of SUS toward the Universal Health coverage principles arises as a unique opportunity to develop a risk stratification algorithm for PTB using health systems metrics and minimal patient data.

2.2. Data sources and variables

Three national level data sets were linked aggregating data from four different groups of variables: prenatal care procedures performed in primary care facilities, physical infrastructure of primary care facilities, women health and delivery characteristics, and sociodemographic characteristics of Brazilian municipalities (Supplementary material 1). Our analysis unit were mothers with children born alive during the years of 2012, and 2014 in municipalities participating in the Access and Quality Improvement Program - (PMAQ - acronym in Portuguese).

The PMAQ is a monitoring survey conducted every two years in Brazil to evaluate the quality of all primary care teams in terms of human resources available, work process and physical structure available to offer care services [28,29]. The PMAQ data covered a total of 17,202 primary care teams, across 3,944 municipalities in 2012, and 29,778 teams in 5,040 municipalities for 2014.

Data about women's health and delivery characteristics were obtained through the Live Births Information System (SINASC - acronym in Portuguese) [1]. Through the SINASC we were able to get information about the duration of the pregnancy, the newborn weight, mother's age, institutional births, and type of delivery. The data from SINASC characterized the delivery of 3,876,666 mothers with live births distributed across the 3,929 Brazilian municipalities. Data was selected corresponding to the years of 2012 (1,786,825), and 2014 (2,089,841), to match the years of availability from the PMAQ database.

The last source of information provided data about sociodemographic aspects regarding Brazilian context [30]. The 2010 census is a full-length portrait of the country with the population profile and the characteristics of their household. Sociodemographic data was aggregated at the municipality level.

To link all three datasets, we had to adjust the data to the most granular level possible. Our unit of analysis was at the individual level, corresponding to each pregnant woman with a live birth. The minimum aggregation level possible to link the PMAQ data and the sociodemographic data was at the municipality level. Therefore, we used the aggregated (mean) value at the municipality level for the variables originated from the PMAQ. The mean of PMAQ variables were linked to the social determinants for each municipality. For instance, the primary care physical structure value for the hypothetical municipality X was the mean of the indicators of all primary care teams within this city. After the aforementioned aggregation the combined score of PMAQ and the municipality was linked to each live birth within that municipality. Thus, we could link each individual level data to an aggregated level of primary care services and sociodemographic data at each municipality. All individual data was anonymized. The categorization of the variables between predictors and outcome as used in the modelling are highlighted in Supplementary material 1.

3. Data analysis

3.1. Pre-processing

One combined dataset was built, aggregating the data from 2012 and 2014. The machine learning models were trained using the aggregated dataset with 3,876,666 live births. The box 1 highlights the steps performed.

The data wrangling process was performed using R Language for Statistical Computing software (3•63) [31] the scikit-learn Python library was used to train the models with graphic processing unit (GPU) support [32].

Data was pre-processed for analysis to address issues regarding outliers, missing values, dummification of categorical variables, near zero variance analysis, linear combos and scaling. Cases presenting values lower than Q1 or greater than Q3 by more than 15 times the interquartile range (IQR=Q3−Q1) were excluded from the analysis. Missing values were addressed using multiple imputation through chained equations (MICE package in R) [33]. The imputation technique was based on the linear regression for continuous variables, and for categorical data was used the proportional odds model [33]. The imputation process was done in two steps. One regarding the variables do the PMAQ and socio-determinants of health and a second one for the individual characteristics from the SINASC database. The imputation for the PMAQ dataset considered 20 iterations, as well the SINASC imputation was done using 10 iterations. Five datasets were created for each step of the imputation process. The first one was selected as reference to impute the missing values in the original datasets. The results achieved after the imputation presented a minimum impact over the measures of central tendency. Categorical variables were transformed into binary pairs (dummies variable conversion). Near zero variance was evaluated to identify the minimum level of variance in variables incorporated in each simulation. Variables with a variance close to zero were excluded from the analysis. The identification of highly correlated variables was performed to suppress pairs of variables with a correlation higher than 90%. Finally, the remaining variables were scaled and mean centralized. The steps highlighted in box 1 describe the parameters selected to each preprocessing step. After running the pre-processing steps from the initial 74 variables, 62 remained for the training and test phase.

Box 1 - Analytical steps adopted to stratify PTB risk.

Analytical steps	Parameters adopted
Outlier analysis exclusion	Only live births within the threshold defined here were included in the analysis: Mother age <= 50, Number of previous live birth <= 10, Number of previous stillbirth <= 5APGAR1 <=10APGAR5 <=10, Newborn weight between 200 and 6.000 grams, Number of previous pregnancy <= 15, Number of previous normal labor <= 15, Number of previous C-Section <=3, Week of deliver <= 43, Number of prenatal consultations <= 15, Month of the beginning of the prenatal care <=9
Missing values imputation using Multiple imputation chained equations approach.	All variables with a maximum of 20% of missing were imputed.[1] Continuous variables were imputed using linear regression, and categorical data was imputed using polytomous regression imputation. The missing distribution of the variables imputed can be observed in the supplementary document 1.
Conversion of categorical variables into dummies	All categorical variables were converted to dummy before the analysis of near zero variance and linear combo.
Analysis of variables with near zero variance and its exclusion	All variables presenting a near zero variance were excluded from the analysis. A variable presents a near zero variance if the percentage of cases presenting unique values is less than 20%.
Suppression of variables highly correlated, as well as linear combos (perfect separation problem)	We dropped from the analysis one variable of every variable pair with a correlation equal or higher 0.9.
Preprocessing of remaining variables to adjust problems regarding differences in scale or distribution	Considering the difference in distribution of the remaining variables we normalized the remaining variables before starting the training process.
Split of database among training and test datasets using a cross-validation approach	The split between training and test dataset were performed using a cross-validation approach with 5 repetitions and 2 folds.

[1]Adequacy of the number of prenatal consultation for the gestational age during the current pregnancy (monthly consultations until the 28th week of gestational age, biweekly consultations from 28 to 36 weeks, and weekly consultations until the baby is born), % of primary care team in the municipality registering the prenatal evolution in health information system, % of primary care team in the municipality using the prenatal monitoring health record, Average number of physicians available by primary care team, Average number of nurses available by primary care team, Average number of dental care offices by health facility, % of health facilities in the municipality functioning for at least 8 hours a day, % of health facilities in the municipality offering medication, % of health facilities in the municipality offering vaccination services, % of health facilities in the municipality with paper forms to monitors the prenatal care evolution, Average proportion of health facilities with at least one equipment from the list: sphygmomanometer, scale, stethoscope, gynecological focus, glucometer, gynecological table, sonar, table. Average proportion of health facilities with at least one medical supplies from the list: glucose meter tape, gynecological speculum, endovaginal brush, Ayres spatula.

The flowchart presented in the Figure 1 details how the preprocessing steps affected the number of live births analyzed in the present study.

3.2. Model building, validation and calibration

Internal validation of the predictive models was conducted using a participant-level based cross-validation approach with two folds and five repetitions for data splitting, following the recommendation of Dietterich [34]. We used a cross-validation approach as this method allows to bring down to the models a high variability to achieve satisfactory results during the training process. Our option for using the cross-validation approach is related to its primary purpose as measuring (generalization) performance of a model. Our idea was to design a model that could be replicated in other contexts. Furthermore, for large sample sizes, the variance issues become less important and the computational part is more of an issue. With our database of 3,8 million records some models required a lot of computational effort to run. Comparative studies have concluded that both cross-validation and bootstrap based approaches if performed well show convergent results [35,36]. The cross-validation process divides the dataset in folds. The K folds refers to the number of groups that a given data sample is to be split into each training round. For each cycle of training, a portion of the data is sampled to be the test set, and the remaining data is used to train the model. For each iteration, the data selected to be the test set is resampled assuring that any observation will be part, at least once, of the test set and of the training set. This procedure increases the chance of the model to be generalizable in a satisfactory manner when applied to a previously unforeseen data. The scheme used for the cross-validation is detailed in Figure 2.

We tested six different machine learning algorithms: eXtreme Gradient Boosting, Elastic Net, Quantile Ordinal Regression - LASSO, Linear Regression, Ridge Regression, and Decision Tree. Our option for these six algorithms was done base on the idea to use different strategies in terms of functioning. Thus, we selected an example according to the following methods:

- Ensemble - eXtreme Gradient Boosting
- Regularization Algorithms - Elastic Net, Ridge Regression, LASSO
- Regression Algorithms - Linear Regression, and
- Decision Tree Algorithms – decision tree.

The predictive models were trained to estimate the week of delivery. As a measure of precision associated with each technique, we consider the normalized root-mean-squared error (RMSE). The RMSE expresses the difference between the actual value of the outcome and that one obtained through the forecasting approach. Lower values of RMSE indicated more precise predicted values. The smaller RMSE gap resulting from training versus test comparison characterizes the best predictive approach.

The calibration of the best performing model was assessed through a graphic approach to examine in which range our prediction is being more or less accurate. The calibration compares the actual data versus the predicted one. The calibration plot is composed of two axes, one of them the actual week of delivery and the second the predicted week of delivery. An accurate predictive model should present a clustered distribution of points around the diagonal axis of the graphic, starting from the point 0,•0.

Role of the Funding Source: The manuscript was funded by the Bill and Melinda Gates foundation and by the National Council for Scientific and Technological Development – Brazil. The funders have no role in terms of the data analysis, interpretation or writing to the present manuscript.

4. Results

4.1. Description of the data

Of the 3,876,666 live births registered between the two years, we have 456,808 PTB being 219,658 (12%) in 2012, and 237,150 (11%) in 2014. The participants of the study were mostly young mothers (average year of delivery for both years of approximately 26 years for both years), single (720,082 for 2012, and 842,586 for 2014), and undergraduate (1,020,148 for 2012, and 1,232,757 for 2014).

The type of delivery varied substantially among the years considered. For 2014, 1,221,418 (68•36%) of deliveries were through C-Section, while in 2012 this number was 1,032,105 (57•76%) of all

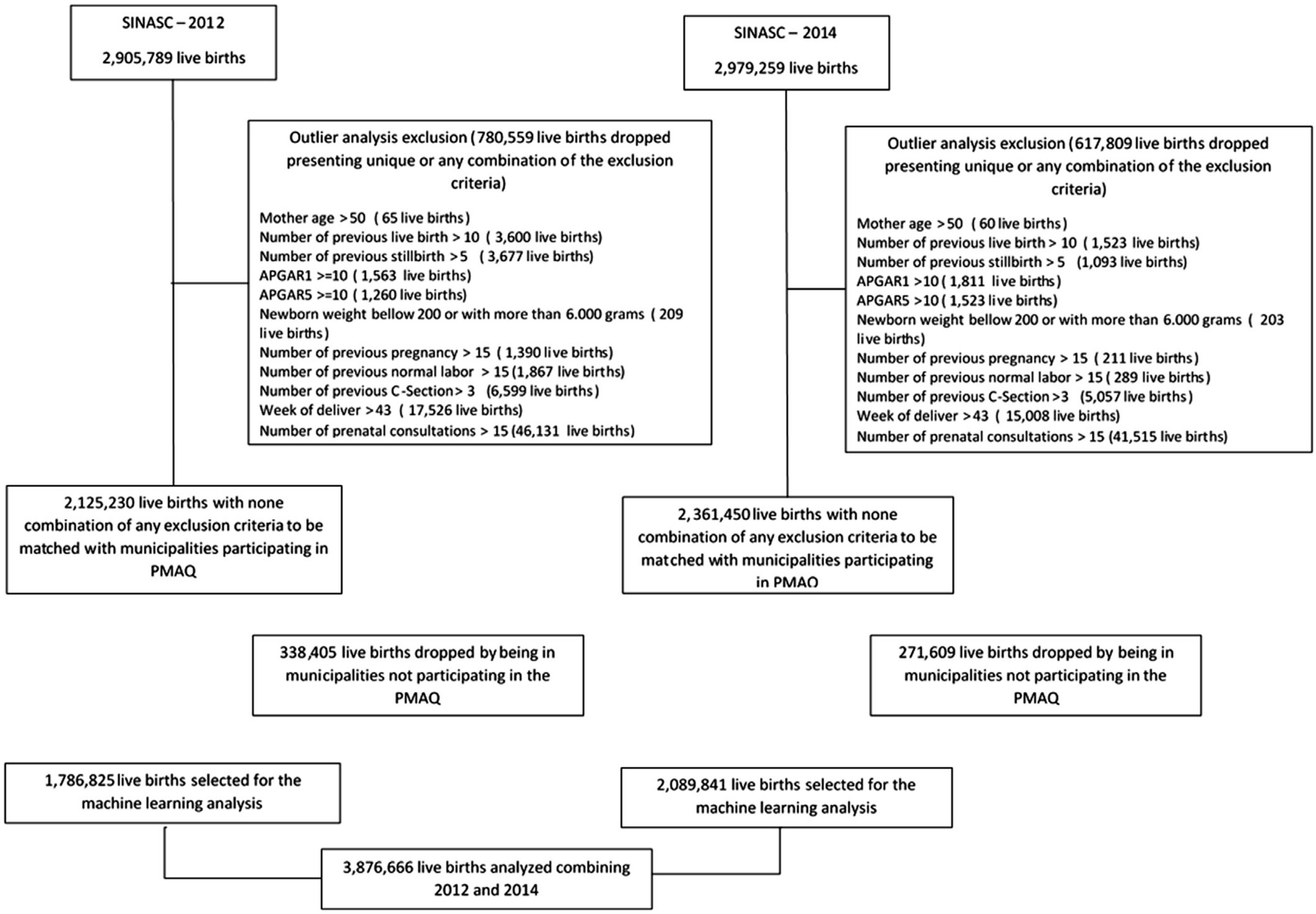

Figure 1. Flowchart of the live birth enrollment in the study.

births. The C-section rates in Brazil are higher than the values presented by other countries, due to local characteristics regarding the care process [40]. The average number of prenatal consultations for both years were very similar: 7•66 (±2•63) for 2012, and 7•76 (±2•63) for 2014, following the recommendation of the Ministry of Health. The month in which the prenatal monitoring started also presented a comparable value between both years analyzed.

The characterization of the primary care network pointed out a high availability in general of essential services dedicated to support prenatal care. The values regarding the physical structure and the primary care work process highlights an elevated offer of prenatal consultations, lab test requisition, and the register of the evolution concerning the prenatal in the information systems. In terms of physical structure, the items presenting the lower levels of availability are related to vaccines and medication for hypertension and diabetes. The coverage of primary care services is high, reaching a municipality average of 73•93% of coverage (±28•71) for 2012 and 78.97(±26.20) for 2014. Thus, the proportion of the population living with the minimum wage was 31•81% (±18•76). The average income per capita was R$ 27,259.38 (±R$ 20,061•90). As these statistics came from the 2010 Census there is no variation considering a yearly basis estimate. The complete table describing the variables analyzed can be found in the supplementary material 2.(Table 2)

4.2. Model validity

Model performance metrics are presented in Table 3. All models obtained RMSE indexes close to each other when comparing the results of the analysis. The absolute RMSE for the algorithms tested show an error of approximately 2 weeks in the forecasting of the week delivery. Models trained and tested using the eXtreme Gradient Boosting approach showed the lowest RMSE values.

Examining the variables in this model was possible to identify the relative importance described in **Figure 3**. The five most important variables to predict the week of delivery were: number of previous deliveries through C-Section, number of prenatal consultations performed, age of the mother, existence of ultrasound exam available in the care network, and % of primary care teams in the municipality registering the oral care consultation.

The model calibration presented a dispersal pattern in terms of comparison between the actual versus the predicted week of delivery. The best performing algorithm exhibited a general trend around the diagonal between the values, but with a significant amount of dispersion regarding the values associated to PTB. The prediction error histogram plots the distribution of the error associated with each observation. The shape of the curve highlights that the predicted week of delivery tends to be higher than the actual one (**Figure 4**).

5. Discussion

This study aimed to test the feasibility to apply and evaluate the performance of a machine learning-based approach to risk stratify PTB. Our aim was to use Brazilian medical records, health system quality data, and social determinants of health to serve as a LMIC proof of concept. To our knowledge this is the first study to develop models encompassing data from an entire country at the primary health center level trying to estimate the risk of PTB based

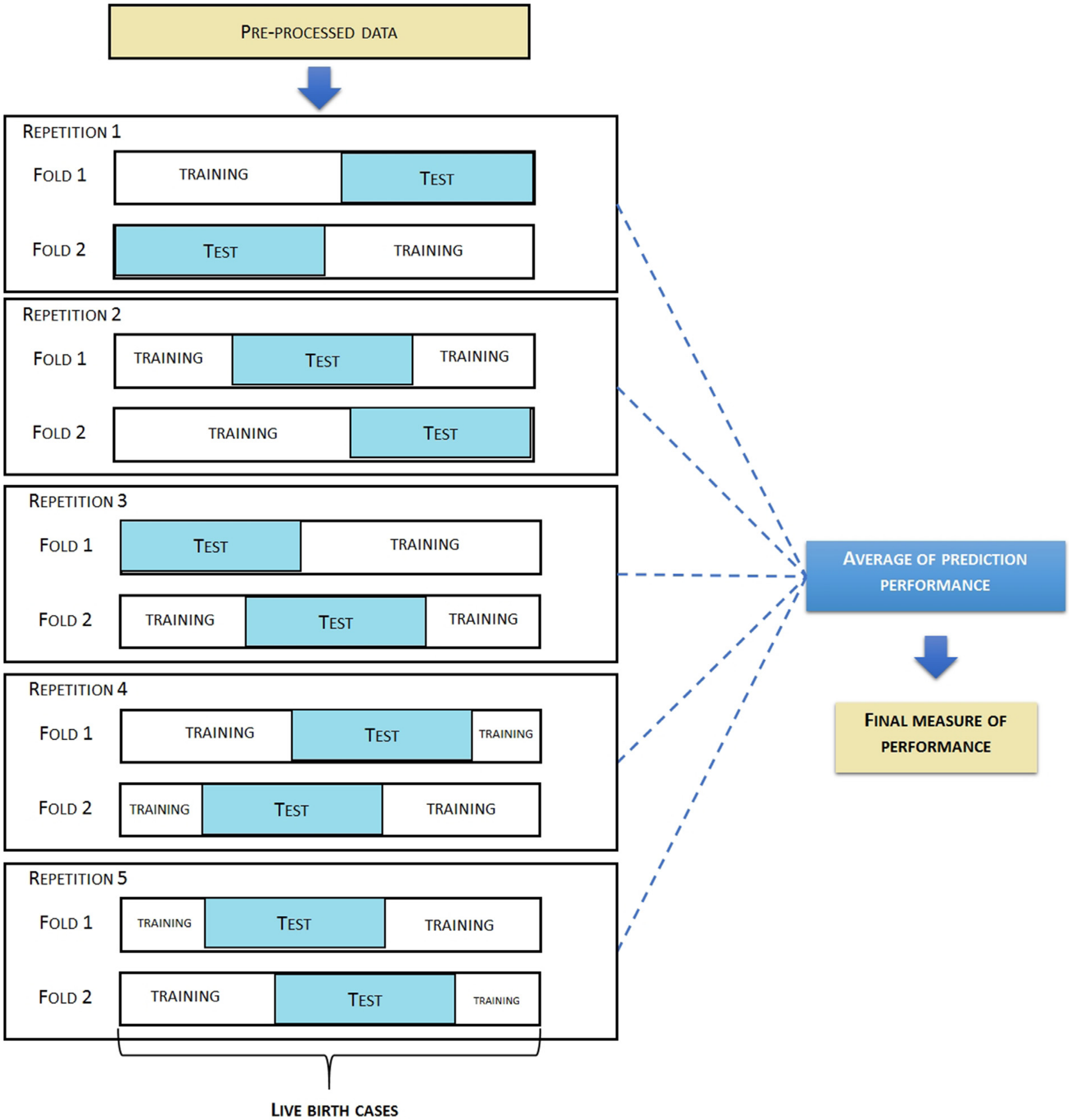

Figure 2. Cross-validation approach used.

on variables from millions of records. Using simple data describing the prenatal care offered, as well as minimal characteristics of the pregnant woman, our model was capable of estimating the week of delivery within a 2 weeks window margin of error.

We were able to identify other 11 papers using machine learning models to predict PTB [37–48]. Most of these studies were cohorts. The sample size varied from 300 [38] to 15,976,537 [45]. Two of them used big national data sets (more than 100,000 women), one in Iran [44] and one in the United States [45]. However, five studies used difficult to access variables, as ultrasound and biochemical markers [38,41,42,47,49]. This makes them difficult for replication in other lower resourced populations, so they are not useful in a LMIC context. None of the studies found were dedicated to create a forecasting approach for the primary health center level.

Prediction models may contribute to personalized risk-based management of women at high risk of spontaneous preterm delivery. However, the results have not been accurate in most of the models identified – the area under the receiver operating characteristic curve (AUC) was around 60-70% [39,40,42,45]. Many of the current models have an insufficient detection rate for application in clinical or public health practices. Predictive models with better predictive capacity are needed to locate women at high risk of preterm delivery.

The three best models we were able to identify were developed in: i) Korea, using artificial neural networks in a data set with 596 women. The accuracy was 0•9115 [41; ii) United Kingdom, using a sample of 300 women, with an AUC equal to 95% with 8% global error using the polynomial classifier [38]; and iii) Zambia, adopting super learner predictions in a sample of 468 women with an AUC of 97.96 for PTB prediction [37]. One important difference among these studies and ours refers to the outcome variable. All of these studies predict PTB as a categorical variable. In our study we could predict the most probable week of delivery. This strategy allows us

Table 2
Descriptive data of most important features.

Characteristic	**Distribution by PMAQ year** **2012**, N = 1,786,825[1]	**2014**, N = 2,089,841[1]	**p-value**[2]
Individual characteristics			
Week of delivery	38.38 (2.21)	38.45 (2.16)	<0.001
Preterm births			<0.001
Yes	219,658 (12%)	237,150 (11%)	
No	1,567,167 (88%)	1,852,691 (89%)	
Mother's age in years	26.24 (6.56)	26.31 (6.64)	<0.001
Number of prenatal consultation	7.66 (2.63)	7.76 (2.63)	<0.001
Marital status - with a companion			<0.001
Yes	408,859 (23%)	495,118 (24%)	
No	1,377,966 (77%)	1,594,723 (76%)	
Month of the pregnancy when the prenatal monitoring started	2.68 (1.55)	2.61 (1.48)	<0.001
Number of previous stillbirth deliveries	0.22 (0.53)	0.21 (0.53)	<0.001
Number of previous live births deliveries	0.93 (1.23)	0.92 (1.21)	<0.001
Number of previous C-section delivery	0.30 (0.59)	0.31 (0.60)	<0.001
Number of previous normal labor work	0.66 (1.21)	0.64 (1.19)	<0.001
Infant gender: Female			0.15
Yes	871,772 (49%)	1,018,075 (49%)	
No	915,053 (51%)	1,071,766 (51%)	
Primary Care health structure and work process			
Primary care team offering Penicillin G benzathine when necessary	0.60 (0.41)	0.64 (0.39)	<0.001
Primary care team using the prenatal monitoring health record	0.71 (0.41)	0.90 (0.16)	<0.001
Primary care team counting with the support for laboratory tests	0.89 (0.21)	0.90 (0.19)	<0.001
Primary care team offering prenatal consultation	0.80 (0.31)	0.75 (0.28)	<0.001
Primary care teams in the municipality counting with a referral network to support the delivery	0.75 (0.34)	0.74 (0.34)	<0.001
Primary care team registering the Papanicolaou tests performed	0.78 (0.27)	0.81 (0.22)	0.2
Primary care team Registering of the oral care consultation	0.58 (0.32)	0.66 (0.28)	<0.001
Primary care teams registering the health professional responsible for patient monitoring	0.93 (0.16)	0.95 (0.12)	<0.001
Primary care teams registering the vaccines performed	0.94 (0.15)	0.95 (0.12)	<0.001
Primary care team receiving the tests results before the delivery	0.76 (0.27)	0.71 (0.26)	<0.001
Primary care team Registering the prenatal evolution in health information system	0.93 (0.18)	0.94 (0.14)	<0.001
Primary care team offering the orientation for tetanus vaccination	0.99 (0.06)	1.00 (0.03)	<0.001
Primary care team offering ultrasound examination	0.63 (0.31)	0.86 (0.18)	<0.001
Socio determinants of health			

[1] Statistics presented: Mean (SD); n (%)
[2] Statistical tests performed: Wilcoxon rank-sum test; chi-square test of independencePTB: preterm birth. primary care: primary health care. N: number. SD: standard deviation.

Table 3
Report the predictive performance of the final model in terms of the validation metrics specified in the methods section.

Model	Number of repetitions *Number of folds	Average RMSE regarding the week of delivery (Standard Deviation)	95% Confidence Interval	Range of RMSE
eXtreme Gradient boosting	10	2•094 (0•005)	2.090 – 2.097	2.090 – 2.100
ElasticNet	10	2•159 (0•006)	2.154 – 2.163	2.150 – 2.170
Quantile Ordinal Regression - LASSO	10	2•182 (0•004)	2.179 – 2.184	2.180 – 2.190
Linear Regression	10	2•120 (0•000)	2.120 – 2.120	2.120 – 2.120
Ridge Regression	10	2•121 (0•003)	2•118 - 2•123	2.120 – 2.130
Decision Tree	10	3•001 (0•003)	2•998 - 3•003	3.000- 3.010

to stratify the risk of PTB in different ways. The 2 weeks window achieved by our prediction limits the possibility of clinical use of the tool developed. Despite this, our study demonstrates the feasibility of using machine learning approaches with minimal characteristics of the pregnant woman aiming to forecast the week of delivery.

In our study, the most important variables used for the prediction are easy to access, and the model has the potential for scalability for other low-and-middle income countries. The impact of the primary care in terms of the PTB could be slightly demonstrated with our approach. We believe that the use of Electronic Health Record data, pointing out the specific team responsible for the care of each pregnant woman could increase the impact of primary care over the PTB. Despite this, the lack of access to identified data for the current manuscript made it impossible to use more precise data from primary care teams. The incorporation of data regarding primary care level can change the patterns observed in the current model. Therefore, it is important to highlight that the weak effect observed from the variables associated to the primary care can be related to the methodological design adopted. Thus, we recommended that future studies should take into consideration of the effect of primary care variables.

The presence of oral care consultation between the most important variables suggests a high level of screening and treatment of periodontal disease during the prenatal care. Pregnant mothers with periodontitis have double risk of PTB as evidenced in an analytical case-control studies and prospective cohort studies meta-analysis [50]. The effect of periodontal pathogenic bacteria burden interfered positively with the occurrence of periodontal disease. More severe periodontal disease as well other systemic infections may unbalance the expression of regulatory inflammatory process cytokines and explain the occurrence of PT [51]. Mediated proinflammatory cytokines with the release of prostaglandin, increased uterine contractility, favoring premature rupture of fetal membranes [52].

Other predictive maternal characteristics for PTB were the number of previous deliveries through C-Section. A meta-analysis with population-based cohorts studies showed the previous C-section

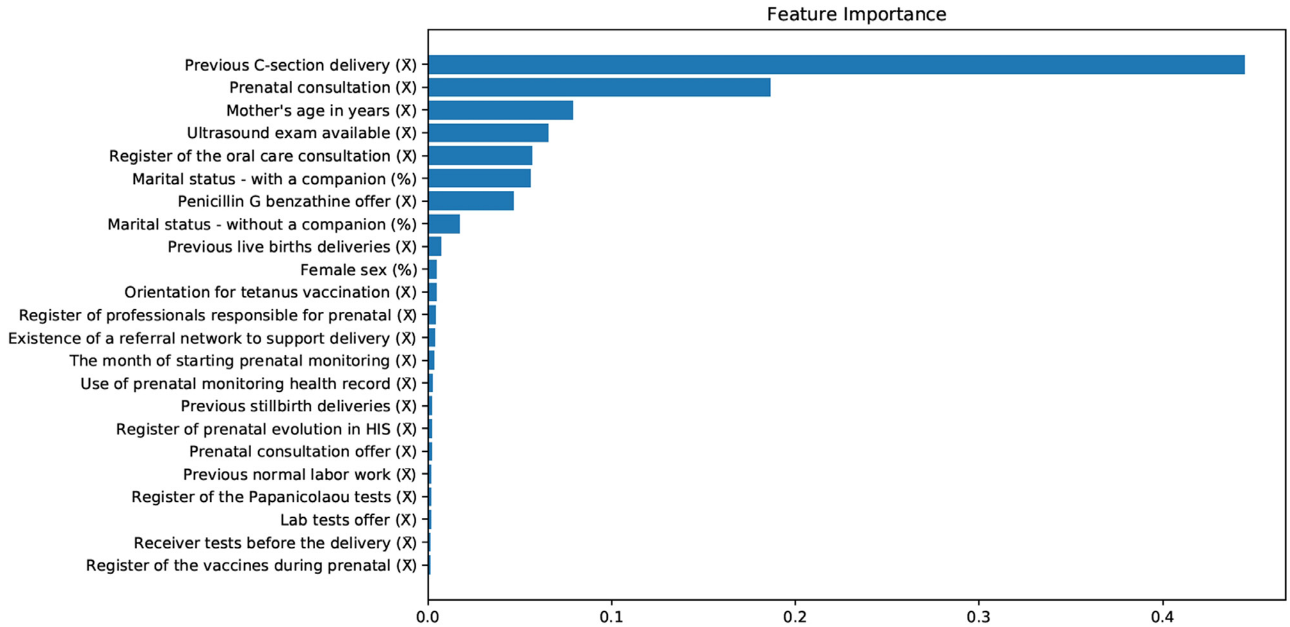

Figure 3. Feature importance.

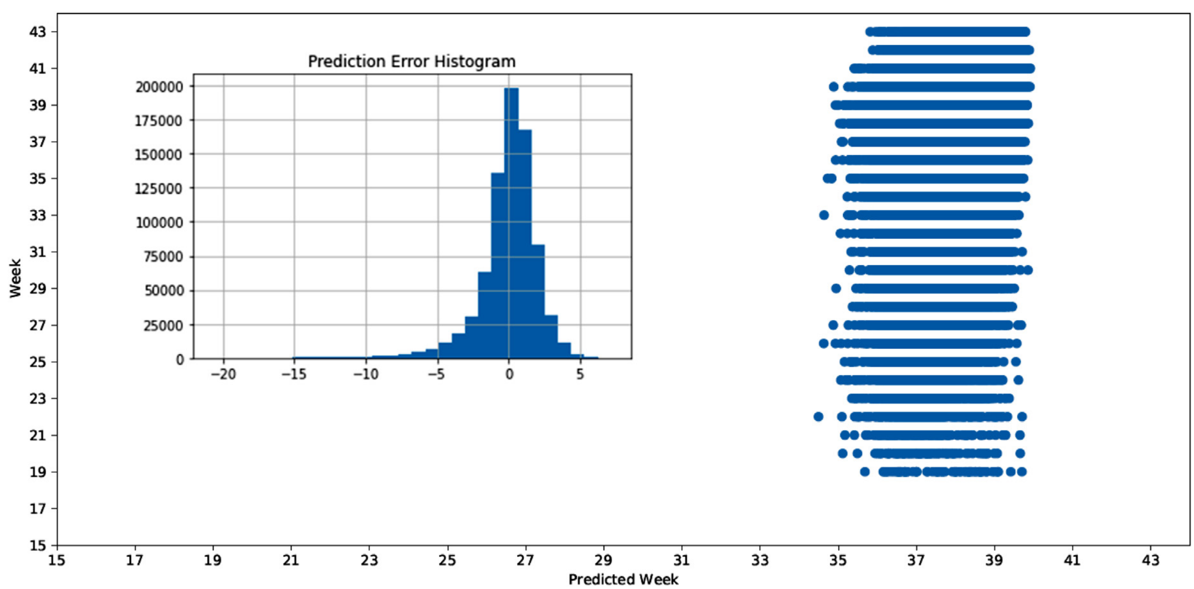

Figure 4. Model calibration and histogram of residuals.

could increase the risk of PTB in subsequent birth [53]. Some hypotheses claim that the uterine structure may be changed by previous C-section [54], for example cervical trauma in unintentional incision into the uterine cervix during the previous C-section could disrupt the cervical integrity, affect the function of the cervix and further increase the risk of PTB in future pregnancies [55]. The formation of uterine scar and intrauterine adhesions after previous C-section could reduce utero-placental function and disturb the position of blastocyst implantation that could further create inadequate conditions for fetal development, increasing the risk of PTB [56]. Another possibility, such maternal conditions (higher body mass index, advanced maternal age, diabetes mellitus, preeclampsia etc.), which were indications for the C-section in the first pregnancy can also be an important cause of PTB in the next pregnancies [57–60].

The increased risk of PTB among aged mothers is largely explained by early labor induction for medical conditions. Advanced maternal age is independently associated with spontaneous prematurity and mainly spontaneous rather than PTB iatrogenic [61,62]. A possible explanation is placental vascular pathology present in older women, as histological evidence of placental bleeding, loss of vessel integrity and lack of physiologic conversion of maternal spiral arteries. Preeclampsia and hypertensive disorders are an important cause of uteroplacental ischemia and PTB [63]. Another possible mechanism of PTB in aged mothers is the decrease of progesterone levels. The progesterone deficiency is associated with PTB and progesterone supplementation could minimize the PTB event [64].

To our knowledge, this is the first study including characteristics of the structure and work process regarding the primary care facilities performing prenatal care, as well as the socioeconomic context of the municipalities. Evidence points out that socioeconomic characteristics combined with individual-level predic-

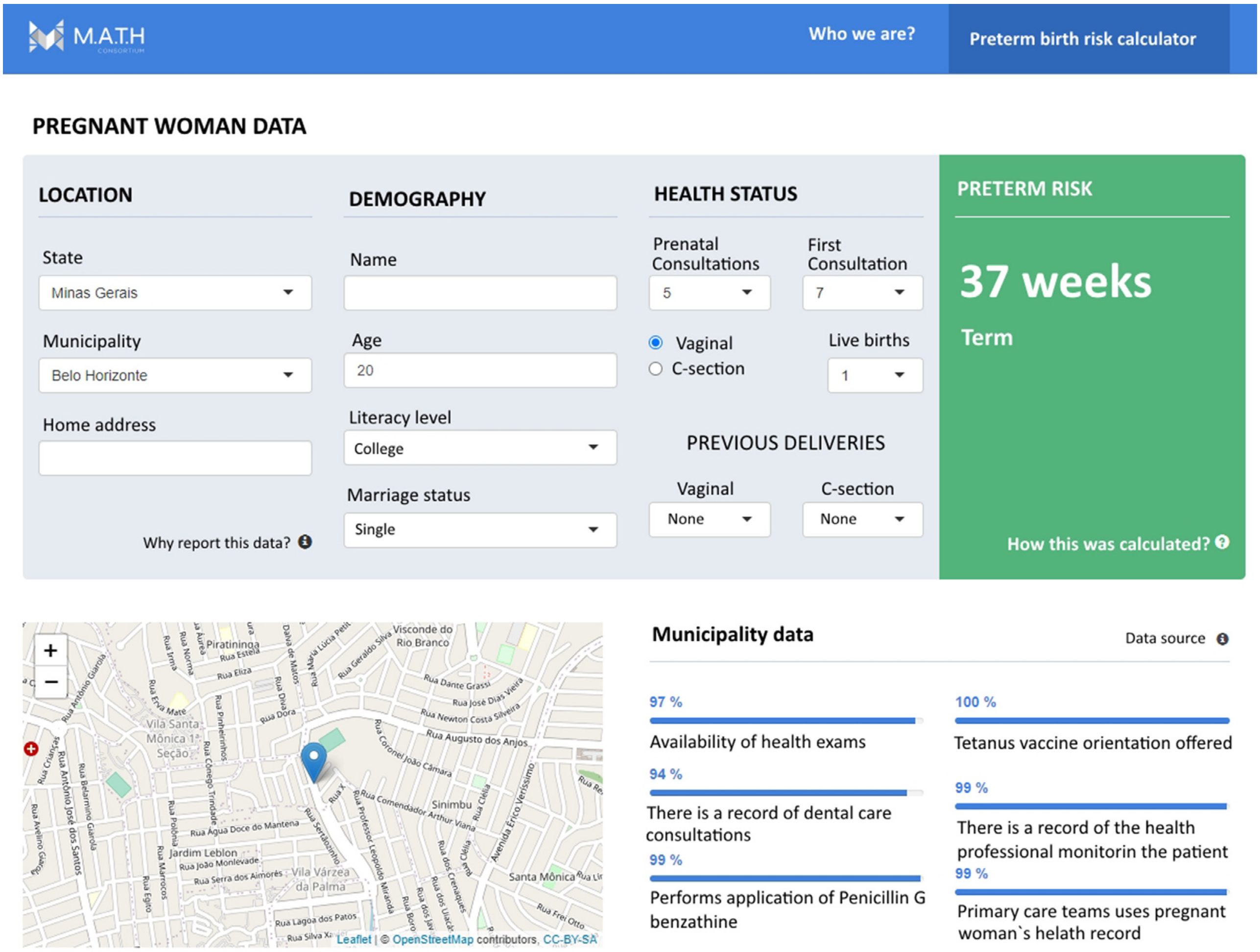

Figure 5. Example of automated risk calculator tool for PTB.

tors improve the overall prediction of PTB compared to individual-level predictors alone [43].

5.1. Application design and possible adoption context

Ideally the application of the algorithm applied would take place in primary care context. After the data gathering of the information of the woman an embedded version of the algorithm will flag the risk attributed to the specific patient being attended. Trying to improve the use of the solution, our idea is to avoid the development of standalone solutions once this approach contributes to increase the workload of the primary care professionals. Avoiding such a type of pitfall would be possible to integrate the use of the algorithm in the regular workflow of a prenatal consultation. After the registration of the data of the patient, the algorithm will flush a risk score of occurrences of a PTB, based on the week of delivery forecasted. Aware of the risk score the primary care professional would have the possibility to monitor closely each patient flagged as having an elevated risk of presenting a PTB. Figure 5 exemplifies how the risk calculator tool may be integrated to electronic health records already in use by health professionals.

A prediction algorithm can be adapted to work conjointly with a chatbot. A chatbot is a solution of communication capable of specific commands in an automatic way. Therefore, a professional or the patient itself can send a preformatted SMS message to a telephone number. The chatbot will receive the formatted message, preprocess the information contained, apply the algorithm of prediction and respond to the sent message with the forecast week of delivery and a possible risk score for PTB [65,66]. Thus, professionals offering care in remote or deprived regions may benefit from using machine learning models in clinical context.

The capability to forecast the week of delivery is not enough by itself to avoid a preterm delivery. Notwithstanding the early alert can create conditions for the health professionals to intervene with the available resources aiming to avoid the premature delivery. Early interventions could be designed to address risk factors related to PTB. The tool developed does not have by any means the intention to underestimate the medical judgement, in terms of best approaches to guarantee a healthy prenatal to the monitored patients. Our idea was to develop a complimentary tool capable of supporting the process of prenatal care.

5.2. Limitations

The data used by analysis presented some restrictions in terms of possibilities of disaggregation. The municipality data may have contributed to diminish the relevance of important variables, as the impact of primary care services. Due to a restriction to access Electronic Health Records, we could not precisely link each pregnant woman to the primary care team responsible for the prenatal care. The impossibility to link each patient to its primary care team of reference limited our capability to identify the impact of primary care over the risk of a preterm birth. We believe that this limitation contributed for the lack of effect related to the primary care work process. If we consider huge cities, São Paulo for example, with large inequities in access to health care and mainly in quality of care, the fact that all livebirths of the city had the same indicators regarding the primary care being offered may have impacted on the model obtained. We are currently working to overcome this limitation. One approach oriented in this sense is the use of geospatial techniques to develop catchment areas of the primary care teams and thus link sociodemographic data from a small region to a live birth [67]. Despite this advance, it is worth highlighting that the calibration results regarding the best performing model may have some issues associated with a regression to the mean. The feature importance demonstrated that the individual characteristics are more relevant to define which parameters are more important for the performance of the models assessed. Thus, the use of more granular data may increase the accuracy of future prediction algorithms. The clinical use of the present model is not recommended, once the 2 weeks margin of error is very broad to accurately support health interventions. Additionally, as the number of prenatal consultations was considered a significant predictor, it is worth highlighting that at the beginning of the prenatal care, the lack of precise information regarding this variable can restrict the model accuracy or even increase the margin of error. A narrower margin of error is desirable to create robust conditions to support medical procedures aiming to lower the chance of a PTB.

The calibration of the model highlighted that the week of delivery needs to be assessed with caution despite the RMSE values. Notwithstanding the advances obtained by our model a window of less than one week would be more robust to uphold actions dedicated to decrease the PTB risk. We have the expectation that the use of more granular data from the use of geospatial approach, as well as the incorporation of more individual data. The use of such information to design the model may help to address the challenge to narrow the window regarding the week of delivery.

Another limitation is the absence of important data regarding clinical conditions classically associated with PTB, like hypertension, diabetes, infections, vaginal bleeding, health habits (use of alcohol, cigarettes, and other drugs), and prior preterm birth. The inclusion of these variables could reduce the RMSE and improve the performance of our predictive model.

5.3. Implications for global health and future steps

The use of machine learning based technologies is increasing together with the broader availability of information. The present work highlighted a small glimpse of what can be done with the use of large datasets. It is necessary that countries across the globe foster initiatives like this one to address unmet objectives necessary to achieve Universal health coverage. The methodological steps presented here can be tailored to work in different settings and help to improve the quality of maternal and childcare, especially in deprived regions of the globe. After the COVID-19 pandemic some initiatives started aiming to better understand the PTB using innovative approaches, as the iPoP study [68].

The development of machine learning solutions is a continuous process. New data coming from different settings can improve the performance of the current solution. We believe that the combined use of geospatial approach and machine learning technology will leverage the performance levels achieved to forecast the week of delivery and PTB chance.

Contributors

TAHR: Conceptualization, Data curation, Formal analysis, Investigation, Methodology, Visualization, Writing, fund raising.

EBAFT: Conceptualization, Data curation, Formal analysis, Investigation, Methodology, Visualization, Writing

DGA: Conceptualization, Data curation, Formal analysis, Investigation, Methodology, Visualization

NCS: Conceptualization, Data curation, Formal analysis, Investigation, Methodology, Visualization

RCSQ: Conceptualization, Data curation, Formal analysis, Investigation, Methodology, Visualization

LA: Data curation, Formal analysis, Investigation, Methodology, Visualization

LAF: Data curation, Formal analysis, Investigation, Methodology, Visualization

MLLS: Data curation, Formal analysis, Investigation, Methodology, Visualization

DBC: Data curation, Formal analysis, Investigation, Methodology, Visualization

MAGC: Investigation, Methodology, Visualization, Writing

AAMS - Conceptualization, Investigation, Methodology, Visualization, Writing

CS - Conceptualization, Investigation, Methodology, Visualization, Writing

JRNV - Conceptualization, Data curation, Formal analysis, Investigation, Methodology, Visualization, Writing

Declaration of Interests

The authors declare no competing interests.

Acknowledgements

The authors would like to acknowledge the support of the Brazilian Ministry of Health and the Bill and Melinda Gates Foundation and the Brazilian National Council of Research (CNPQ).

Funding

The study was supported by the Bill and Melinda Gates Foundation (grant: OPP1202186) the Foundation of Research Support from Maranhão‘s State - FAPEMA (grants: RCUK-01538/19 and BEPP-03783/13), by the Brazilian National Council of Research - CNPQ (grants: 443834/2018-0, 306592/2018-5, 370275/2019-5, 371794/2019-6, 380456/2019-2, 80028/2019-0, and 381602/2018-4), and Coordination for the Improvement of Higher Education Personnel – CAPES (finance code 001).

Data sharing statement

The data and codes used to obtain the results presented in this manuscript can be found in the following links:

Raw datasets before imputation process: https://figshare.com/s/b74fc0342af7afbcf8a1

Datasets used for machine-learning modeling: https://figshare.com/s/0cb2656f2e41bc3b770e

Codes to estimate the models: https://figshare.com/s/bbd33153cf9291d682a2

Editor's Note

The Lancet Group takes a neutral position with respect to territorial claims in published maps and institutional affiliations.

Supplementary materials

Supplementary material associated with this article can be found, in the online version, at doi:10.1016/j.lana.2021.100053.

References

[1] Departamento de Brasil http://www2.datasus.gov.br/DATASUS/index.php?area=02.
[2] . Born too soon. In: The Global Action Report on Preterm Birth. Geneva: World Heal Organ Publ; 2012. p. 1–126.
[3] Silveira MF, Santos IS, Barros AJD, Matijasevich A, Barros FC, Victora CG. Increase in preterm births in Brazil: review of population-based studies. Rev Saude Publica 2008. doi:10.1590/S0034-89102008000500023.
[4] Smid MC, Stringer EM, Stringer JSA. A Worldwide Epidemic: The Problem and Challenges of Preterm Birth in Low- and Middle-Income Countries. Am J Perinatol 2016;33:276–89.
[5] Purisch SE, Gyamfi-Bannerman C. Epidemiology of preterm birth. Semin. Perinatol. 2017. doi:10.1053/j.semperi.2017.07.009.
[6] Cristina da Silva N, Rocha TAH, Amaral PV, et al. Comprehending the lack of access to maternal and neonatal emergency care: Designing solutions based on a space-time approach. PLoS One 2020;15:e0235954.
[7] Huda TM, Tahsina T, El Arifeen S, Dibley MJ. The importance of intersectoral factors in promoting equity-oriented universal health coverage: A multilevel analysis of social determinants affecting neonatal infant and under-five mortality in Bangladesh. Glob Health Action 2016. doi:10.3402/gha.v9.29741.
[8] Reeves A, Gourtsoyannis Y, Basu S, McCoy D, McKee M, Stuckler D. Financing universal health coverage - Effects of alternative tax structures on public health systems: Cross-national modelling in 89 low-income and middle-income countries. Lancet 2015. doi:10.1016/S0140-6736(15)60574-8.
[9] WHO. Tracking Universal Health Coverage. 2015.
[10] Agustina R, Dartanto T, Sitompul R, et al. Universal health coverage in Indonesia: concept, progress, and challenges. Lancet 2019. doi:10.1016/S0140-6736(18)31647-7.
[11] Bloom G, Katsuma Y, Rao KD, Makimoto S, Yin JDC, Leung GM. Next steps towards universal health coverage call for global leadership. BMJ 2019. doi:10.1136/bmj.l2107.
[12] He J-R, Ramakrishnan R, Lai Y-M, et al. Predictions of Preterm Birth from Early Pregnancy Characteristics: Born in Guangzhou Cohort Study. J Clin Med 2018. doi:10.3390/jcm7080185.
[13] Farrant BM, White SW, Shepherd CCJ. Trends and predictors of extreme preterm birth: Western Australian population-based cohort study. PLoS One 2019. doi:10.1371/journal.pone.0214445.
[14] Meertens LJE, van Montfort P, Scheepers HCJ, et al. Prediction models for the risk of spontaneous preterm birth based on maternal characteristics: a systematic review and independent external validation. Acta Obstet. Gynecol. Scand. 2018. doi:10.1111/aogs.13358.
[15] Goldenberg RL, Culhane JF, Iams JD, Romero R. Epidemiology and causes of preterm birth. Lancet 2008. doi:10.1016/S0140-6736(08)60074-4.
[16] Gravett MG, Rubens CE, Nunes TM. Global report on preterm birth and stillbirth (2 of 7): discovery science. BMC Pregnancy Childbirth 2010;10:1–16.
[17] Thompson JMD, Irgens LM, Rasmussen S, Daltveit AK. Secular trends in socio-economic status and the implications for preterm birth. Paediatr Perinat Epidemiol 2006. doi:10.1111/j.1365-3016.2006.00711.x.
[18] Muglia LJ, Katz M. The enigma of spontaneous preterm birth. N. Engl. J. Med. 2010. doi:10.1056/NEJMra0904308.
[19] Rondó PHC, Ferreira RF, Nogueira F, Ribeiro MCN, Lobert H, Artes R. Maternal psychological stress and distress as predictors of low birth weight, prematurity and intrauterine growth retardation. Eur J Clin Nutr 2003. doi:10.1038/sj.ejcn.1601526.
[20] Newnham JP, Dickinson JE, Hart RJ, Pennell CE, Arrese CA, Keelan JA. Strategies to prevent preterm birth. Front. Immunol. 2014;5. doi:10.3389/fimmu.2014.00584.
[21] Katz J, Lee ACC, Kozuki N, et al. Mortality risk in preterm and small-for-gestational-age infants in low-income and middle-income countries: A pooled country analysis. Lancet 2013;382:417–25.
[22] de Oliveira LL, Gonçalves A de C, da Costa JSD, Bonilha AL, de L. Maternal and neonatal factors related to prematurity. Rev da Esc Enferm 2016. doi:10.1590/S0080-623420160000400002.
[23] Kenji R, Cuman N. Risk factors for prematurity : document search. Esc Anna Nery Rev Enferm; 2009.
[24] Barros FC, Bhutta ZA, Batra M, Hansen TN, Victora CG, Rubens CE. Global report on preterm birth and stillbirth (3 of 7): Evidence for effectiveness of interventions. BMC Pregnancy Childbirth 2010. doi:10.1186/1471-2393-S1-S3.
[25] Villar J, Carroli G, Khan-Neelofur D, Al Et. Patterns of routine antenatal care for low-risk pregnancy (Cochrane Review). (Date of most recent substantive update: 18 August 2001). Cochrane Database Syst Rev 2001. doi:10.1002/14651858.CD000934.
[26] Luo W, Phung D, Tran T, et al. Guidelines for developing and reporting machine learning predictive models in biomedical research: A multidisciplinary view. J Med Internet Res 2016. doi:10.2196/jmir.5870.
[27] Castro MC, Massuda A, Almeida G, et al. Brazil's unified health system: the first 30 years and prospects for the future. Lancet 2019. doi:10.1016/s0140-6736(19)31243-7.
[28] Macinko J, Harris MJ, Rocha MG. Brazil's national program for improving primary care access and quality (PMAQ) fulfilling the potential of the world's largest payment for performance system in primary care. J Ambul Care Manage 2017;40:S4–11.
[29] Rocha TAH, Thomaz EBAF, da Silva NC, et al. Oral primary care: An analysis of its impact on the incidence and mortality rates of oral cancer. BMC Cancer 2017;17. doi:10.1186/s12885-017-3700-z.
[30] CensoIBGE Censo 2010. Inst. Bras. Geogr. e Estatística. 2010. http://censo2010.ibge.gov.br/.
[31] Core Team R. R: A language and environment for statistical computing. R Found. Stat. Comput. 2019.
[32] Pedregosa F, Varoquaux G, Gramfort A, et al. Scikit-learn: Machine learning in Python. J Mach Learn Res 2011;12:2825–30.
[33] White IR, Royston P, Wood AM. Multiple imputation using chained equations: Issues and guidance for practice. Stat Med 2011. doi:10.1002/sim.4067.
[34] Dietterich TG. Approximate Statistical Tests for Comparing Supervised Classification Learning Algorithms. Neural Comput 1998. doi:10.1162/089976698300017197.
[35] Kim JH. Estimating classification error rate: Repeated cross-validation, repeated hold-out and bootstrap. Comput Stat Data Anal 2009. doi:10.1016/j.csda.2009.04.009.
[36] Jonathan O, Omoregbe N, Misra S. Empirical Comparison of Cross-Validation and Test Data on Internet Traffic Classification Methods. Journal of Physics: Conference Series. Institute of Physics Publishing 2019:12044.
[37] Rittenhouse KJ, Vwalika B, Keil A, et al. Improving preterm newborn identification in low-resource settings with machine learning. PLoS One 2019;14. doi:10.1371/journal.pone.0198919.
[38] Fergus P, Cheung P, Hussain A, Al-Jumeily D, Dobbins C, Iram S. Prediction of Preterm Deliveries from EHG Signals Using Machine Learning. PLoS One 2013;8. doi:10.1371/journal.pone.0077154.
[39] Tucker CM, Berrien K, Menard MK, et al. Predicting preterm birth among participants of North Carolina's pregnancy medical home program. Matern Child Heal J 2015;19:2438–53.
[40] Do Carmo Leal M, Esteves-Pereira AP, Nakamura-Pereira M, et al. Prevalence and risk factors related to preterm birth in Brazil. Reprod Health 2016. doi:10.1186/s12978-016-0230-0.
[41] Kim YS. Analysis of spontaneous preterm labor and birth and its major causes using artificial neural network. J Korean Med Sci 2019;34:1–10.
[42] Weber A, Darmstadt GL, Gruber S, et al. Application of machine-learning to predict early spontaneous preterm birth among nulliparous non-Hispanic black and white women. Ann Epidemiol 2018;28:783–9 e1.
[43] Adhikari K, Patten SB, Williamson T, et al. Does neighborhood socioeconomic status predict the risk of preterm birth? A community-based Canadian cohort study. BMJ Open 2019;9:1–10.
[44] Gao C, Osmundson S, Velez Edwards DR, Jackson GP, Malin BA, Chen Y. Deep learning predicts extreme preterm birth from electronic health records. J Biomed Inform 2019;100:103334.
[45] Khatibi T, Kheyrikoochaksarayee N, Sepehri MM. Analysis of big data for prediction of provider-initiated preterm birth and spontaneous premature deliveries and ranking the predictive features. Arch Gynecol Obstet 2019;300:1565–82.
[46] Koivu A, Sairanen M. Predicting risk of stillbirth and preterm pregnancies with machine learning. Heal Inf Sci Syst 2020;8. doi:10.1007/s13755-020-00105-9.
[47] Lacey L, Daulton E, Wicaksono A, Covington JA, Quenby S. Volatile organic compound analysis, a new tool in the quest for preterm birth prediction—an observational cohort study. Sci Rep 2020;10:1–9.
[48] Lee KS, Ahn KH. Application of artificial intelligence in early diagnosis of spontaneous preterm labor and birth. Diagnostics 2020;10:1–11.
[49] Koivu A, Sairanen M. Predicting risk of stillbirth and preterm pregnancies with machine learning. Heal Inf Sci Syst 2020;8. doi:10.1007/s13755-020-00105-9.
[50] Manrique-Corredor EJ, Orozco-Beltran D, Lopez-Pineda A, Quesada JA, Gil-Guillen VF, Carratala-Munuera C. Maternal periodontitis and preterm birth: Systematic review and meta-analysis. Community Dent Oral Epidemiol 2019. doi:10.1111/cdoe.12450.
[51] Costa EM, de Araujo Figueiredo CS, Martins RFM, et al. Periodontopathogenic microbiota, infectious mechanisms and preterm birth: analysis with structural equations (cohort—BRISA). Arch Gynecol Obstet 2019. doi:10.1007/s00404-019-05355-x.
[52] Vrachnis N, Vitoratos N, Iliodromiti Z, Sifakis S, Deligeoroglou E, Creatsas G. Intrauterine inflammation and preterm delivery. Annals of the New York Academy of Sciences 2010. doi:10.1111/j.1749-6632.2010.05684.x.
[53] Zhang Y, Zhou J, Ma Y, et al. Mode of delivery and preterm birth in subsequent births: A systematic review and meta-analysis. PLoS One 2019. doi:10.1371/journal.pone.0213784.
[54] Yasseen AS, Bassil K, Sprague A, Urquia M, Maguire JL. Late preterm birth and previous cesarean section: a population-based cohort study. J Matern Neonatal Med 2019. doi:10.1080/14767058.2018.1438397.
[55] Koyama S, Tomimatsu T, Kanagawa T, Sawada K, Tsutsui T, Kimura T. Cervical insufficiency following cesarean delivery after prolonged second stage of labor: Experiences of two cases. J Obstet Gynaecol Res 2010. doi:10.1111/j.1447-0756.2009.01152.x.
[56] Timor-Tritsch IE, Monteagudo A, Cali G, et al. Cesarean scar pregnancy is a precursor of morbidly adherent placenta. Ultrasound Obstet Gynecol 2014. doi:10.1002/uog.13426.
[57] Carnero AM, Mejía CR, García PJ. Rate of gestational weight gain, pre-pregnancy body mass index and preterm birth subtypes: A retrospective cohort study from Peru. BJOG An Int J Obstet Gynaecol 2012. doi:10.1111/j.1471-0528.2012.03345.x.
[58] Waldenström U, Cnattingius S, Vixner L, Norman M. Advanced maternal age increases the risk of very preterm birth, irrespective of parity: a population-based register study. BJOG An Int J Obstet Gynaecol 2017. doi:10.1111/1471-0528.14368.

[59] Dorfman H, Srinath M, Rockhill K, Hogue C. The Association Between Diabetes Mellitus Among American Indian/Alaska Native Populations with Preterm Birth in Eight US States from 2004–2011. Matern Child Health J 2015. doi:10.1007/s10995-015-1761-7.
[60] Connealy BD, Carreno CA, Kase BA, Hart LA, Blackwell SC, Sibai BM. A history of prior preeclampsia as a risk factor for preterm birth. Am J Perinatol 2014. doi:10.1055/s-0033-1353439.
[61] Fuchs F, Monet B, Ducruet T, Chaillet N, Audibert F. Effect of maternal age on the risk of preterm birth: A large cohort study. PLoS One 2018. doi:10.1371/journal.pone.0191002.
[62] Londero AP, Rossetti E, Pittini C, Cagnacci A, Driul L. Maternal age and the risk of adverse pregnancy outcomes: A retrospective cohort study. BMC Pregnancy Childbirth 2019. doi:10.1186/s12884-019-2400-x.
[63] Holzman C, Kelly R, Senagore P, et al. Placental vascular pathology findings and pathways to preterm delivery. Am J Epidemiol 2009. doi:10.1093/aje/kwp131.
[64] Norwitz ER, Caughey AB. Progesterone supplementation and the prevention of preterm birth. Rev Obstet Gynecol 2011.
[65] Linde DS, Korsholm M, Katanga J, Rasch V, Lundh A, Andersen MS. One-way SMS and healthcare outcomes in Africa: Systematic review of randomised trials with meta-analysis. PLoS One 2019;14:e0217485.
[66] Nadarzynski T, Miles O, Cowie A, Ridge D. Acceptability of artificial intelligence (AI)-led chatbot services in healthcare: A mixed-methods study. Digit Heal 2019;5:205520761987180.
[67] Rocha TAH, de Almeida DG, do Amaral PVM, et al. Proposta de metodologia para estimar a área de cobertura potencial por equipes de atenção primária. Rev Panam Salud Pública 2019;43:1.
[68] Stock SJ, Zoega H, Brockway M, et al. The international Perinatal Outcomes in the Pandemic (iPOP) study: protocol. Wellcome Open Res 2021. doi:10.12688/wellcomeopenres.16507.1.