

MULTIRACIAL IDENTITY RESPONSE AS A PREDICTOR OF PRETERM BIRTH  
AMONG NULLIPAROUS, SINGLETON BIRTHING PEOPLE IN THE US: AN  
APPLICATION OF MACHINE LEARNING ALGORITHMS

By

Heesu Kim

A THESIS

Submitted to  
Michigan State University  
in partial fulfillment of the requirements  
for the degree of

Epidemiology – Master of Science

2025

## ABSTRACT

**Background:** Preterm birth (PTB) is a significant cause of neurological or respiratory complications and infant death. Early identification of pregnant people at risk for PTB enables timely interventions and personalized pregnancy management to prevent potential complications. Over the past ten years, the multiracial population in the US has experienced significant growth. Multiracial disaggregation has been suggested as a factor that could help explain disparities in PTB rates, but it remains unclear whether classifying people into granular racial groups helps predict PTB.

**Objectives:** This study aims to build four predictive models for preterm birth and to investigate which of them are important predictors of PTB across 31 race/ethnicity groups that include multiracial identities among nulliparous, singleton birthing people.

**Methods:** We used population-based, cross-sectional data from U.S. birth records in 2019. Medical and socioeconomic factors potentially associated with PTB and race/ethnicity groups, including multiracial groups that are available within the first 16 weeks of pregnancy, were compared between nulliparous, singleton birthing people delivering preterm (<37 weeks of gestation) and term ( $\geq 37$  weeks of gestation). Logistic regression with all variables, logistic regression with selected main effect variables and two-way interaction variables, Decision Tree, and a Random Forest model were employed to build the prediction models. A Random Forest model from an oversampling dataset was utilized to assess the relative importance of risk factors.

**Results:** 97,555 individuals experienced PTB, and 24,041 were classified as multiracial among the analytic sample (N=1,032,465). The ranges of areas under the receiver-operating-characteristic curves(AUC) of all models with oversampling data were 57. The accuracy range of all models with an oversampling dataset was 62 to 65. The mean decrease in the accuracy of the importance plot indicated that some multiracial groups were important predictors of PTB compared with socioeconomic factors.

**Conclusions:** This study's results supported the idea that several granular multiracial groups could be considered meaningful predictors of PTB.

## TABLE OF CONTENTS

INTRODUCTION .....	1
METHODS .....	4
RESULTS .....	9
DISCUSSION .....	11
CONCLUSION.....	15
REFERENCES .....	16
APPENDIX A: TABLES.....	22
APPENDIX B: FIGURES .....	32

# INTRODUCTION

Preterm birth (PTB), which is defined as birth before the completion of 37 weeks of gestation (1,2), is the leading cause of infant morbidity and mortality in the world (28). In the US, preterm birth affects approximately 10% of live-born deliveries, and recent data from the National Vital Statistics System (NVSS) found an increase in preterm birth prevalence from 2016 to 2022, even after accounting for variation during COVID-19 (4). In addition, infants born prematurely face a heightened risk of immediate health issues, including neurodevelopmental disabilities and respiratory and gastrointestinal complications, as well as enduring challenges such as cardiovascular and metabolic disorders (5). Therefore, identifying risk factors for preterm birth is essential for effective early prevention. Early identification of pregnant people at risk for PTB enables timely interventions and personalized pregnancy management to prevent potential complications. Some proximal factors, including infection or inflammation, vascular disease, and uterine overdistension, are suspected to have a relationship with preterm birth (3).

With efforts to identify risk factors, PTB prediction has garnered significant attention in recent decades (15,16). According to the American College of Obstetricians and Gynecologists (ACOG) clinical management guidelines, it is hard to predict the spontaneous PTB of singleton infants in nulliparous people, and there remains controversy and uncertainty regarding what screening tests to use (20). For example, there appear to be no advantages in using screening tests such as short cervix or endovaginal ultrasonography (29, 30). Because of the lack of screening tests, maternal demographics like granular race group or socioeconomic factors can become strong predictors of PTB in clinical settings. Various risk stratification models have been developed based on demographic factors and medical obstetric history (34,35). Since these characteristics are readily available, they are easily applicable in clinical practice (11).

One of Healthy People 2030's objectives is to 'Eliminate health disparities, achieve

health equity, and attain health literacy to improve the health and well-being of all' (10). Race and ethnicity provide one dimension to evaluate health disparities in PTB. For example, African American people experience a rate of PTB that is more than 1.5 times that of PTB compared with non-Hispanic White women (9.5% vs 14.7%, respectively) (41). Additionally, AIAN people experienced higher rates of preterm birth (11.5% vs. 9.1%) and low birth weight (8.0% vs. 6.9%) compared to non-Hispanic white (NHW) infants (42). Although there is a growing body of research on racial and ethnic disparities in health outcomes, including PTB, most researchers have studied the health of minority monoracial (one race only) groups. A growing demand exists to comprehensively understand the health and health outcomes of the multiracial (two or more races) population.

Over the past ten years, the multiracial population in the US has experienced significant growth. According to the U.S. Census Bureau, the number of individuals identifying as multiracial increased by 276% between 2010 and 2020, rising from 9 million to 33.8 million people, and the White alone group decreased by 8.6% since 2010 (9). Multiracial disaggregation has been suggested as a factor that could help explain disparities in PTB rates (6,12, 22), but studies focusing on the role of multiracial identity in predicting preterm birth remain limited.

Starting in 2016, all 50 states, along with the District of Columbia, Puerto Rico, Guam, the Northern Mariana Islands, and the U.S. Virgin Islands, reported race data in alignment with the revised 1997 Office of Management and Budget (OMB) standards (27). These standards permit reporting of at least five race categories, either as single races (i.e., reported alone) or as combinations of multiple races (i.e., more than one race). Building upon this change, the 2003 revision of the U.S. Standard Certificate of Live Birth allowed the reporting of multiple races for each parent (26). The standards for collecting racial information were fully implemented in all U.S. states in 2016, creating an environment where multiracial research is more feasible using birth certificates. This study aims to build four predictive models for preterm birth and to investigate which of the 31 race/ethnicity groups are significant predictors of PTB among nulliparous,

singleton birthing people.

# METHODS

## 1 Target Population and Study Sample

This study used US nationwide live birth certificates from 2019 provided by the NVSS. Observations were included if they were singleton births to nulliparous birthing people aged 18-44. The 18–44 age range is used to reflect the natural reproductive age while excluding extreme age groups (teenagers and women aged 45 and older) who have a significantly higher risk of PTB. Observations were excluded if the number of preterm births in each multiracial group was less than 10 or if variables of interest exhibited missingness. The analytic sample included 1,032,465 births (Figure 1).

## 2 Variables

The outcome variable was preterm birth (PTB), defined as less than 37 weeks of gestation, estimated as weeks from an obstetric estimate of conception to the delivery date. The analyses included 28 variables, 18 of which represented race and ethnicity, and 10 represented socioeconomic factors or medical conditions that could be risk factors for PTB and are available to clinicians for decision-making in the 16th week of pregnancy. We use 16 weeks as a reference point because preterm birth prediction models are often built using data collected before the first trimester ends (45). The 18 racial/ethnicity groups were included among 31 race/ethnicity groups because of the exclusion criteria. Racial /ethnic identity were self-reported in the birth certificate dataset. The 10 socioeconomic factors and medical conditions were included because prior research suggests their relationship with PTB and they were included in the dataset (2,19,29,37) (Table 1).

## 3 Analysis Plan

### 3.1 Descriptive study

Table 2 presents frequency counts and percentages for all categorical variables.

### 3.2 Analytic study

To determine which model best fits the data, we created a training dataset comprised of a random sample of 80 percent (825,972) of the observed births. The remaining 20 percent

of observations were used as the test dataset. We split the data for model evaluation.

The oversampling method was used to create models. Oversampling is a technique used to equalize the sizes of two groups (those with and without PTB) to help train machine learning models more evenly. There was a significant data imbalance because PTB occurs in approximately 10% of births. Without the oversampling method, it is hard to train PTB's patterns and decrease the performance of learning algorithms. Therefore, we utilized an oversampling method to mitigate the imbalance and improve model performance.

For the first objective, we created four machine learning models. The four statistical algorithms compared were: 1) logistic regression model with all variables (logistic regression model 1), 2) a logistic regression model with interaction terms (logistic regression model 2), 3) a Decision Tree model (DT), and 4) a Random Forest model (RF). The relative performance of the four models in the test dataset was assessed by examining overall accuracy, sensitivity, specificity, area under the operating characteristic curve (AUC), and balanced accuracy. Accuracy is the number of correct predictions over the total number of predictions. Sensitivity is the percentage of true positive in a disease. Specificity is the percentage of true negative in non-diseased individuals, and balanced accuracy is sensitivity plus specificity divided by 2. Area under the operating characteristic curve can be calculated by using the true positive rate (TPR) and false positive rate (FPR) at every possible threshold.

### **3.2.1 Logistic regression**

Logistic regression is a widely used statistical method in clinical research to examine the relationships between patient characteristics and binary outcomes. These models are a specific type of generalized linear model, estimated using maximum likelihood. In generalized linear models, the expected outcome is modeled as a function of a linear combination of predictor variables, with logistic regression using the logit function (14). The logistic regression results provide odds ratios that describe the associations between the dependent and independent variables. Additionally, the model generates an estimated probability for the outcome, which can be applied for classification and prediction. We



used the resulting model for the prediction of PTB.

This study included two types of logistic regression models for prediction. Logistic regression model 1 included all 28 variables listed in Table 1 as main effects. The second logistic regression included all variables and two-way interactions. A univariable analysis was conducted to assess the statistical relationship between each individual, main effects variable and PTB, excluding race and ethnicity groups. Variables exhibiting a significant relationship with the outcome ( $\alpha \leq 0.05$ ) were selected for the final model. Because all variables' p-values were less than the alpha, we included them all. Spearman's correlations were chosen to evaluate collinearity between variables. Two variables were considered correlated when the Spearman's coefficient was higher than 0.7. No variables exhibited correlated relationships. The final logistic model was determined through the Akaike Information Criterion (AIC) and a backward stepwise elimination approach, where variables with a p-value greater than 0.05 were excluded from the model. All variables remained. Finally, two-way interactions between variables with p-values  $\leq 0.05$ , except for the race and ethnicity variables, were evaluated. Of the 45 two-way interaction variables assessed, 23 two-way interactions were added to the final logistic regression model.

### **3.2.2 Decision Tree and Random Forest**

The third prediction model was a Decision Tree. A Decision Tree is a model developed by asking and answering questions about independent variables, where a node represents each independent variable. After answering the different questions at each node, the algorithm goes to the destination, a leaf, which returns the predicted results. Such a tree is fitted by splitting nodes to minimize a particular loss function (17). Hyperparameter tuning was used to increase the performance of the Decision Tree (38). Three parameters were adjusted in the Decision Tree analysis in this study: complexity parameter, minimum observation for splitting a node, and maximum depth of the tree. Through cross-validation and grid search, the best combination of complexity parameters, minimum observation for splitting a node, and maximum depth of the tree were selected. The optimal hyperparameters were chosen based on the AUC metric, which is a helpful

measure for evaluating the performance of classification models (13). AUC reflects the balance between sensitivity (true positive rate) and specificity (false positive rate) rather than just accuracy.

Unfortunately, a Decision Tree can have high variance, indicating that its predictions are highly sensitive to fluctuations in the training data and thus overfitting. This overfitting results in the model capturing noise or random fluctuations in the training data rather than the underlying trend, which diminishes its predictive ability. Bootstrap aggregating, or bagging, can solve this issue. A bagging method trains multiple trees on different bootstrap samples of the same data and then averages these models. This technique smooths out the prediction and reduces the variance. We chose a Random Forest model as the bagging model, in which Decision Trees are trained with the vital restriction that in each step with a new split, only a few randomly selected features become available candidates for the split. By aggregating less correlated trees, Random Forest led to a possible improvement in variance (17).

There were two steps for creating our final classification Random Forest model. First, we identified a specific number of trees with bootstrap samples and drew them from the training data. Second, the Decision Tree was trained on each bootstrap sample until a specific minimum node size that we specified was reached. This second step involved randomly selecting  $m$  variables that we specified from all the predictor variables, picking the best split among these variables, and splitting the node into two according to the best split. The predictions of all the trees in the Random Forest get to a prediction for a new observation, and a majority vote is utilized to decide on the final prediction (17).

Three parameters were adjusted in the Random Forest analysis in our study: the number of trees, the number of variables randomly drawn, and a specific minimum node size. As for the number of chosen variables, the default in a classification model is the square root of the number of variables, so we chose five as the number of variables. The best combination of the optimal number of trees and minimum node size was selected through cross-validation and grid search. The optimal hyperparameters were chosen based on AUC. We chose the mean decrease accuracy as the metric to gauge important figures of a

Random Forest. First, it is easy to interpret. For example, if the mean decrease accuracy value for a variable called A is 0.3, it means that when the information of variable A is removed (i.e., when it is shuffled), the model's accuracy decreases by 0.3 (43). The other reason was that the "permutation accuracy importance" measure is an advanced way to evaluate variable importance in Random Forests. It works by randomly shuffling the values of a predictor variable, disrupting its original relationship with the response variable. This disruption leads to a decrease in prediction accuracy, highlighting the importance of the variable in the model's performance (44).

#### **4 IRB and Statistical packages**

This research was exempt from IRB oversight because it used publicly available, deidentified data, and statistical analyses were conducted using R (v4.4.2 R Foundation for Statistical Computing).

# RESULTS

## 1 Descriptive study

A total of 1,032,465 birthing people were included in the analysis, of which 97,555 (9.45%) experienced PTB. Many birthing people were aged between 25 and 29, and they were usually born in the US. Birthing people were most often overweight and had private insurance. Birthing people also typically visit prenatal care for the first time within 2 months of the beginning of their pregnancy (Table 2).

## 2 Evaluating Preterm Birth Prediction: Methodologies and the Role of Multiracial Groups

### 2.1 Classification of preterm birth using four methods to identify the best model

Table 3 showed the PTB prediction models' accuracy, sensitivity, specificity, AUC, and balanced accuracy. Accuracy and AUC from the original dataset had the same value in the two logistic regression models and the Random Forest model. The Decision Tree's accuracy was 0.9, similar to other models. Four models using oversampled data showed better AUC and sensitivity than four models using the original data. The AUC of the four models using the oversampling dataset was 0.57. The accuracy using the oversampling dataset was the highest for the Random Forest (0.62, 0.63, 0.62, and 0.65, respectively).

### 2.2 Effects of multiracial groups on PTB using a Random Forest model

Figure 2 illustrates the ranked importance of variables in the Random Forest model derived from an oversampling dataset. Nine race and ethnicity groups had higher mean decrease accuracy than insurance type and cigarette smoking before the second trimester. Of those nine, five race and ethnicity groups were multiracial (e.g., non-Hispanic Black and White, non-Hispanic Asian and White, non-Hispanic Asian and NHOPI and White, non-Hispanic AIAN and White).

To explore the impact of race on preterm birth probabilities, we created three hypothetical observations using the Random Forest model trained on an oversampling dataset. The only difference between the three observations was race—non-Hispanic Asian, non-Hispanic Black and Asian, non-Hispanic Asian and NHOPI and White—

while all other variables remained constant. As shown in Table 4, the prediction for non-Hispanic Asian birthing people indicated no possibility of preterm birth. In other words, 100% term birth probability. In contrast, for the non-Hispanic Black and Asian birthing people, the model predicted a 45.1% probability of PTB despite identical values for all other covariates. Moreover, the model predicted a 27.4% PTB for non-Hispanic Asian birthing people and non-Hispanic NHOPI and White birthing people.

# DISCUSSION

In this study, we conducted a comprehensive analysis of the determinants of preterm birth (PTB) using a population-based, cross-sectional cohort of 1,032,465 people with a live birth in the U.S. in 2019, with 28 variables, including sociodemographic factors, maternal medical history, and detailed race and ethnicity categories. Using machine learning techniques, we established prediction models for PTB and investigated the importance of granular racial and ethnic groups.

Moreover, this study evaluated the performance of four machine learning models in predicting PTB; all except a Decision Tree model had the same accuracy and AUC when using the original dataset. However, when using the oversampling dataset, the Random Forest had the highest accuracy. Still, the accuracy or AUC's performance was poor across the four machine learning models. One similar study evaluated the clinical potential of spontaneous PTB prediction models, and the models' performances were similarly poor (AUC 0.51-0.56) for nulliparous women (39). Generally, a previous PTB history is considered the most critical risk factor for a future PTB (40). Indeed, Meertens et al. showed that having prior PTB data increases the performance of PTB prediction models. Therefore, the lack of previous PTB data in this study could be one reason why the predictive ability of our models is much lower, since the target population was nulliparous, singleton-birth people. So there is no data about previous PTB for our target population.

Finally, we identified the order of importance of variables based on results from a Random Forest with the oversampling dataset. Among the 28 variables analyzed, non-Hispanic Black and White, non-Hispanic Asian and White, non-Hispanic Asian and NHOPI and White, and non-Hispanic AIAN and White ranked higher than insurance types or cigarette smoking before 2<sup>nd</sup> trimester based on the mean decrease accuracy metric. The results of our analysis were similar to a previous study that used 2016-2019 Medical Expenditures Panel Survey (MEPS) data to predict foregone preventive dental care for adults and demonstrated that combining distinct multiracial groups into one

group resulted in the lowest model performance among stratified race and ethnicity groups (36). Moreover, we used the Random Forest model to create three virtual observations to estimate the probability of PTB, holding all other variables constant. We estimated that these three observations had substantial racial/ethnic differences in PTB probability. Such results would be obscured if aggregated into a single monoracial Asian group.

This study makes two contributions to the literature.

First, several studies demonstrate the importance of race disaggregation for understanding many health conditions, such as respiratory diseases or type 2 diabetes (7,8,21). Some studies also identify variations in PTB regarding multiracial granular categories and show the need for detailed racial/ethnic subcategories (6,12,23). In this study, we determined that multiracial groups are essential predictors of PTB. We showed that when we use granular race/ethnicity groups to predict PTB, the prediction probability changes. Thus, it may be helpful to predict PTB by granular groups.

Second, researchers often aggregate multiracial subgroups together due to small sample sizes (24). Two previous studies examining PTB race aggregation used rate comparison methods and conventional statistics (6,23). However, we used an oversampling approach to balance preterm birth and term birth to avoid the sample size requirements of conventional statistics. Moreover, the Random Forest method can handle high-dimensional data with many variables combined in non-linear fashions to predict outcomes or detect new patterns (25). This method can better identify the importance of granular racial groups, previously masked due to small sample sizes or data complexity.

This study has certain limitations.

First, we did not conduct subgroup analyses of PTB. Preterm birth can be classified based on its etiology or gestational age. In terms of etiology, PTB can be indicated, which results from medical intervention due to maternal or fetal complications such as severe preeclampsia or non-reassuring fetal heart rate, or spontaneous, which occurs due to spontaneous preterm labor or preterm premature rupture of membranes (PPROM). Based on gestational age, PTB is further divided into early PTB (birth before 32+0 weeks of

gestation) and late PTB (birth between 32+1 and 36+6 weeks of gestation) (31,32). The risk factors of each type of PTB, by etiology or gestational age, could be different since the pathophysiology of PTB can differ. If we divide PTB into its subgroups to create a prediction model, it is possible that the model's performance will improve.

Second, we could not include adequate measurement of relevant socioeconomic or medical conditions. We included only the variables that could both be obtained within the first 16 weeks of pregnancy (to support clinical relevance of findings) and that are included on birth certificates. Other potentially relevant, yet unmeasured, covariates include alcohol consumption and previous medical history. Similar to the first limitation, adding more variables to the models may further improve the model's predictive power of PTB for nulliparous, singleton birthing people.

Third, measurement errors may occur. Measurement error is one of the main factors that undermine data validity. A previous study found that the sensitivity and positive predictive value (PPV) of birth certificate data varied widely across items, ranging from 0% to 100% (46), suggesting the presence of inaccuracies that could introduce measurement error. We did not explore the impact the measurement error could have introduced into this study, but acknowledge its potential role in biasing reported results.

Fourth, additional machine learning algorithms are available, including diverse algorithms, such as the artificial neural network, the support vector machine, the neural network algorithm, extreme gradient boosting, and multi-layer perceptron (MLP) (11,33). Other machine learning methods might perform better with this data and could be explored in future research.

Lastly, we need to consider what the race/ethnicity category on the birth certificate actually represents. Roth discussed the multiple dimensions of race, including racial identity, self-classification, observed race, reflected race, phenotype, and racial ancestry (47). However, we do not know which of these dimensions people rely on when responding to birth certificate questionnaires. There are also some misclassification cases between birth certificate data and hospital data (48). Therefore, there is a need to be cautious when conducting research that takes race/ethnicity groups into account, as



indicated on birth certificates.

## CONCLUSION

This study developed four prediction models for preterm birth (PTB) and demonstrated that specific multiracial and ethnic groups might serve as meaningful predictors of PTB risk among nulliparous, singleton-birthing people. The findings suggest that considering granular racial/ethnic identity could enhance the accuracy of PTB risk prediction.

However, these four models are still hard to apply in a clinical setting. Future improvements in these prediction models, incorporating more diverse variables and advanced techniques, could better detect possible PTB in clinical settings, enabling more personalized and timely interventions for at-risk individuals.

## REFERENCES

1. Spong C. Y. (2013). Defining "term" pregnancy: recommendations from the Defining "Term" Pregnancy Workgroup. *JAMA*, 309(23), 2445–2446.  
<https://doi.org/10.1001/jama.2013.6235>
2. Purisch, S. E., & Gyamfi-Bannerman, C. (2017). Epidemiology of preterm birth. *Seminars in perinatology*, 41(7), 387–391.  
<https://doi.org/10.1053/j.semperi.2017.07.009>
3. Goldenberg, R. L., Culhane, J. F., Iams, J. D., & Romero, R. (2008). Epidemiology and causes of preterm birth. *Lancet (London, England)*, 371(9606), 75–84.  
[https://doi.org/10.1016/S0140-6736\(08\)60074-4](https://doi.org/10.1016/S0140-6736(08)60074-4)
4. Martin, J. A., & Osterman, M. J. K. (2024). Shifts in the Distribution of Births by Gestational Age: United States, 2014–2022. *National vital statistics reports : from the Centers for Disease Control and Prevention, National Center for Health Statistics, National Vital Statistics System*, 73(1), 1–11.
5. Saigal, S., & Doyle, L. W. (2008). An overview of mortality and sequelae of preterm birth from infancy to adulthood. *Lancet (London, England)*, 371(9608), 261–269.  
[https://doi.org/10.1016/S0140-6736\(08\)60136-1](https://doi.org/10.1016/S0140-6736(08)60136-1)
6. Brown, C. C., Moore, J. E., & Tilford, J. M. (2023). Rates Of Preterm Birth And Low Birthweight: An Analysis Of Racial And Ethnic Populations. *Health affairs (Project Hope)*, 42(2), 261–267. <https://doi.org/10.1377/hlthaff.2022.00656>
7. Pleis, J. R., & Barnes, P. M. (2008). A comparison of respiratory conditions between multiple race adults and their single race counterparts: an analysis based on American Indian/Alaska Native and white adults. *Ethnicity & health*, 13(5), 399–415.  
<https://doi.org/10.1080/13557850801994839>
8. Springer, Y. P., Filardo, T. D., Woodruff, R. S., & Self, J. L. (2024). Racial and Ethnic Disaggregation of Tuberculosis Incidence and Risk Factors Among American Indian and Alaska Native Persons-United States, 2001–2020. *American journal of public health*, 114(2), 226–236. <https://doi.org/10.2105/AJPH.2023.307498>
9. U.S. Census Bureau. (2021, August 12). *Improved race, ethnicity measures show U.S. is more multiracial*. Census.gov.  
<https://www.census.gov/library/stories/2021/08/improved-race-ethnicity-measures-reveal-united-states-population-much-more-multiracial.html>
10. Huang, D. T., Uribe, A., & Talih, M. (2024). Measuring progress toward target attainment and the elimination of health disparities in *Healthy People 2030*. *Vital and*

*Health Statistics*, 2(211). <https://doi.org/10.15620/cdc/164019>

11. Lee, K. S., & Ahn, K. H. (2020). Application of Artificial Intelligence in Early Diagnosis of Spontaneous Preterm Labor and Birth. *Diagnostics (Basel, Switzerland)*, 10(9), 733. <https://doi.org/10.3390/diagnostics10090733>
12. Hamilton, B. E., & Ventura, S. J. (2007). Characteristics of births to single- and multiple-race women: California, Hawaii, Pennsylvania, Utah, and Washington, 2003. *National vital statistics reports : from the Centers for Disease Control and Prevention, National Center for Health Statistics, National Vital Statistics System*, 55(15), 1–20.
13. Fawcett, T. (2005). Using AUC and accuracy in evaluating learning algorithms. *IEEE Transactions on Knowledge and Data Engineering*, 17(3), 328–336. <https://doi.org/10.1109/TKDE.2005.50>
14. Hosmer, D. W., & Lemeshow, S. (2000). *Applied logistic regression* (2nd ed.). Wiley.
15. Narice, B. F., Labib, M., Wang, M., Byrne, V., Shepherd, J., Lang, Z. Q., & Anumba, D. O. (2024). Developing a logistic regression model to predict spontaneous preterm birth from maternal socio-demographic and obstetric history at initial pregnancy registration. *BMC pregnancy and childbirth*, 24(1), 688. <https://doi.org/10.1186/s12884-024-06892-3>
16. Mirzamoradi, M., Mokhtari Torshizi, H., Abaspour, M., Ebrahimi, A., & Ameri, A. (2024). A Neural Network-based Approach to Prediction of Preterm Birth using Non-invasive Tests. *Journal of biomedical physics & engineering*, 14(5), 503–508. <https://doi.org/10.31661/jbpe.v0i0.2201-1449>
17. Hastie, T., Tibshirani, R., & Friedman, J. (2009). Model inference and averaging. In *The elements of statistical learning: Data mining, inference, and prediction* (2nd ed.). Springer.
18. Hosmer, D. W., Lemeshow, S., & Sturdivant, R. X. (2013). *Applied logistic regression* (3rd ed.). Wiley & Sons, Inc.
19. Prediction and Prevention of Spontaneous Preterm Birth: ACOG Practice Bulletin, Number 234. (2021). *Obstetrics and gynecology*, 138(2), e65–e90. <https://doi.org/10.1097/AOG.0000000000004479>
20. Springer, Y. P., Filardo, T. D., Woodruff, R. S., & Self, J. L. (2024). Racial and Ethnic Disaggregation of Tuberculosis Incidence and Risk Factors Among American Indian and Alaska Native Persons-United States, 2001-2020. *American journal of public*

*health*, 114(2), 226–236. <https://doi.org/10.2105/AJPH.2023.307498>

21. Koyama, A. K., Bullard, K. M., Onufrak, S., Xu, F., Saelee, R., Miyamoto, Y., & Pavkov, M. E. (2023). Risk Factors Amenable to Primary Prevention of Type 2 Diabetes Among Disaggregated Racial and Ethnic Subgroups in the U.S. *Diabetes care*, 46(12), 2112–2119. <https://doi.org/10.2337/dci23-0056>

22. Blebu, B. E., Waters, O., Lucas, C. T., & Ro, A. (2022). Variations in Maternal Factors and Preterm Birth Risk among Non-Hispanic Black, White, and Mixed-Race Black/White Women in the United States, 2017. *Women's health issues : official publication of the Jacobs Institute of Women's Health*, 32(2), 140–146. <https://doi.org/10.1016/j.whi.2021.10.010>

23. Brown, C. C., & DuBois, D. (2024). Racial/Ethnic Disparities in Pregnancy-Associated Death: The Critical Importance of Disaggregation by Cause of Death and Race/Ethnicity. *American journal of public health*, 114(7), 666–668. <https://doi.org/10.2105/AJPH.2024.307700>

24. Miotto, R., Wang, F., Wang, S., Jiang, X., & Dudley, J. T. (2018). Deep learning for healthcare: review, opportunities and challenges. *Briefings in bioinformatics*, 19(6), 1236–1246. <https://doi.org/10.1093/bib/bbx044>

25. National Center for Health Statistics. (n.d.). 2003 revisions of the U.S. Standard Certificates of Live Birth and Death and fetal death report. Centers for Disease Control and Prevention. Retrieved from [https://www.cdc.gov/nchs/nvss/vital\\_certificate\\_revisions.htm](https://www.cdc.gov/nchs/nvss/vital_certificate_revisions.htm)

26. Office of Management and Budget. (1997). Revisions to the standards for the classification of federal data on race and ethnicity. *Federal Register*, 62(210), 58782–58790.

27. Walani S. R. (2020). Global burden of preterm birth. *International journal of gynaecology and obstetrics: the official organ of the International Federation of Gynaecology and Obstetrics*, 150(1), 31–33. <https://doi.org/10.1002/ijgo.13195>

28. Kleinrouweler, C. E., Cheong-See, F. M., Collins, G. S., Kwee, A., Thangaratnam, S., Khan, K. S., Mol, B. W., Pajkrt, E., Moons, K. G., & Schuit, E. (2016). Prognostic models in obstetrics: available, but far from applicable. *American journal of obstetrics and gynecology*, 214(1), 79–90.e36. <https://doi.org/10.1016/j.ajog.2015.06.013>

29. Esplin, M. S., Elovitz, M. A., Iams, J. D., Parker, C. B., Wapner, R. J., Grobman, W. A., Simhan, H. N., Wing, D. A., Haas, D. M., Silver, R. M., Hoffman, M. K., Peaceman, A. M., Caritis, S. N., Parry, S., Wadhwa, P., Foroud, T., Mercer, B. M., Hunter, S. M., Saade, G. R., Reddy, U. M., ... nuMoM2b Network (2017). Predictive Accuracy of Serial

Transvaginal Cervical Lengths and Quantitative Vaginal Fetal Fibronectin Levels for Spontaneous Preterm Birth Among Nulliparous Women. *JAMA*, 317(10), 1047–1056. <https://doi.org/10.1001/jama.2017.1373>

30. Orzechowski, K. M., Boelig, R., Nicholas, S. S., Baxter, J., & Berghella, V. (2015). Is universal cervical length screening indicated in women with prior term birth?. *American journal of obstetrics and gynecology*, 212(2), 234.e1–234.e2345. <https://doi.org/10.1016/j.ajog.2014.08.029>

31. Brown, H. K., Speechley, K. N., Macnab, J., Natale, R., & Campbell, M. K. (2014). Neonatal morbidity associated with late preterm and early term birth: the roles of gestational age and biological determinants of preterm birth. *International journal of epidemiology*, 43(3), 802–814. <https://doi.org/10.1093/ije/dyt251>

32. Hendler, I., Goldenberg, R. L., Mercer, B. M., Iams, J. D., Meis, P. J., Moawad, A. H., MacPherson, C. A., Caritis, S. N., Miodovnik, M., Menard, K. M., Thurnau, G. R., & Sorokin, Y. (2005). The Preterm Prediction Study: association between maternal body mass index and spontaneous and indicated preterm birth. *American journal of obstetrics and gynecology*, 192(3), 882–886. <https://doi.org/10.1016/j.ajog.2004.09.021>

33. Wong, K., Tessema, G. A., Chai, K., & Pereira, G. (2022). Development of prognostic model for preterm birth using machine learning in a population-based cohort of Western Australia births between 1980 and 2015. *Scientific reports*, 12(1), 19153. <https://doi.org/10.1038/s41598-022-23782-w>

34. Koivu, A., & Sairanen, M. (2020). Predicting risk of stillbirth and preterm pregnancies with machine learning. *Health information science and systems*, 8(1), 14. <https://doi.org/10.1007/s13755-020-00105-9>

35. Liu, Y., Liu, J., & Shen, H. (2024). Machine learning model-based preterm birth prediction and clinical nomogram: A big retrospective cohort study. *International journal of gynaecology and obstetrics: the official organ of the International Federation of Gynaecology and Obstetrics*, 10.1002/ijgo.16036. Advance online publication. <https://doi.org/10.1002/ijgo.16036>

36. Schuch, H. S., Furtado, M., Silva, G. F. D. S., Kawachi, I., Chiavegatto Filho, A. D. P., & Elani, H. W. (2023). Fairness of Machine Learning Algorithms for Predicting Foregone Preventive Dental Care for Adults. *JAMA network open*, 6(11), e2341625. <https://doi.org/10.1001/jamanetworkopen.2023.41625>

37. Wang, R., Shi, Q., Jia, B., Zhang, W., Zhang, H., Shan, Y., Qiao, L., Chen, G., & Chen, C. (2022). Association of Preterm Singleton Birth With Fertility Treatment in the US. *JAMA network open*, 5(2), e2147782. <https://doi.org/10.1001/jamanetworkopen.2021.47782>

38. Sam'an, M., Farikhin, & Munsarif, M. (2025). An improved decision tree model through hyperparameter optimization using a modified gray wolf optimization for diabetes classification. *Computer methods in biomechanics and biomedical engineering*, 1–17. Advance online publication. <https://doi.org/10.1080/10255842.2025.2460178>
39. Meertens, L. J. E., van Montfort, P., Scheepers, H. C. J., van Kuijk, S. M. J., Aardenburg, R., Langenveld, J., van Dooren, I. M. A., Zwaan, I. M., Spaanderman, M. E. A., & Smits, L. J. M. (2018). Prediction models for the risk of spontaneous preterm birth based on maternal characteristics: a systematic review and independent external validation. *Acta obstetrica et gynecologica Scandinavica*, 97(8), 907–920. <https://doi.org/10.1111/aogs.13358>
40. Koullali, B., Oudijk, M. A., Nijman, T. A., Mol, B. W., & Pajkrt, E. (2016). Risk assessment and management to prevent preterm birth. *Seminars in fetal & neonatal medicine*, 21(2), 80–88. <https://doi.org/10.1016/j.siny.2016.01.005>
41. Hamilton, B. E., Martin, J. A., & Osterman, M. J. K. (2024). Births: Provisional data for 2023. *Vital statistics rapid release* (Vol. 35). Centers for Disease Control and Prevention. <https://www.cdc.gov/nchs/data/vsrr/vsrr035.pdf>
42. Martin, J. A., Hamilton, B. E., Osterman, M. J. K., & Driscoll, A. K. (2019). Births: Final Data for 2018. *National vital statistics reports : from the Centers for Disease Control and Prevention, National Center for Health Statistics, National Vital Statistics System*, 68(13), 1–47.
43. Genuer, R., & Poggi, J. M. (2020). *Random forests with R*. Springer. <https://doi.org/10.1007/978-3-030-56485-8>
44. Strobl, C., Malley, J., & Tutz, G. (2009). An introduction to recursive partitioning: rationale, application, and characteristics of classification and regression trees, bagging, and random forests. *Psychological methods*, 14(4), 323–348. <https://doi.org/10.1037/a0016973>
45. Arabi Belaghi, R., Beyene, J., & McDonald, S. D. (2021). Prediction of preterm birth in nulliparous women using logistic regression and machine learning. *PloS one*, 16(6), e0252025. <https://doi.org/10.1371/journal.pone.0252025>
46. Josberger, R. E., Wu, M., & Nichols, E. L. (2019). Birth Certificate Validity and the Impact on Primary Cesarean Section Quality Measure in New York State. *Journal of community health*, 44(2), 222–229. <https://doi.org/10.1007/s10900-018-0577-y>
47. Roth, W. D. (2016). The multiple dimensions of race. *Ethnic and Racial Studies*, 39(8), 1310–1338. <https://doi.org/10.1080/01419870.2016.1140793>

48. Reid, C. N., Obure, R., Salemi, J. L., Ilonzo, C., Louis, J., Rubio, E., & Sappenfield, W. M. (2023). Race and Ethnicity Misclassification in Hospital Discharge Data and the Impact on Differences in Severe Maternal Morbidity Rates in Florida. *International journal of environmental research and public health*, 20(9), 5689.  
<https://doi.org/10.3390/ijerph20095689>



## APPENDIX A: TABLES

Table 1. Independent variables

<b>Number</b>	<b>Input variables</b>	<b>Conceptualization of the variables</b>
1	Hispanic	Yes, No
2	Non Hispanic American Indian/Alaska Native (AIAN)	Yes, No
3	Non Hispanic AIAN & Asian & White (AIAsW)	Yes, No
4	Non Hispanic AIAN & White (AIW)	Yes, No
5	Non Hispanic Asian	Yes, No
6	Non Hispanic Asian & Native Hawaiian and other Pacific Islander (NHOPI) (ANH)	Yes, No
7	Non Hispanic Asian & White (AW)	Yes, No
8	Non Hispanic Asian & NHOPI & White (AsNHW)	Yes, No
9	Non Hispanic Black	Yes, No
10	Non Hispanic Black & AIAN (BAI)	Yes, No
11	Non Hispanic Black & AIAN & White (BAIW)	Yes, No
12	Non Hispanic Black & Asian (BAs)	Yes, No
13	Non Hispanic Black & Asian & White (BAsW)	Yes, No

Table 1 (cont'd)

14	Non Hispanic Black & NHOPI (BNH)	Yes, No
15	Non Hispanic Black & White (BW)	Yes, No
16	Non Hispanic NHOPI	Yes, No
17	Non Hispanic NHOPI & White (NHW)	Yes, No
18	Non Hispanic White	Yes, No
19	Maternal age	(1) 18-19 years (2) 20-24 years (3) 25-29 years (4) 30-34 years (5) 35-39 years (6) 40-44 years
20	Nativity	(1) Born in the U.S. (2) Born outside the U.S
21	Education	(1) High school-level degree (2) More than a High school-level degree
22	Cigarettes smoked before 2 <sup>nd</sup> trimester	Yes, No
23	Body mass index (BMI)	(1) Underweight (2) Normal (3) Overweight
24	Pre-pregnancy diabetes	Yes, No
25	Pre-pregnancy hypertension	Yes, No

Table 1 (cont'd)

26	Insurance type	(1) Medicaid (2) Private insurance (3) Other
27	Month of 1 <sup>st</sup> prenatal visit	(1) 1-month Visit (2) 2-month Visit (3) 3-month Visit (4) 4-month Visit
28	Infertility treatment used	Yes, No

Table 2. Basic characteristics of the analytic population

<b>Variables</b>	<b>Term birth N= 934,910 ( ) = proportion of total (%)</b>	<b>Preterm birth N= 97,555 ( ) = proportion of total (%)</b>
<b>Age</b>		
<b>18 – 19 years</b>	79,138 (8.5)	10,180 (10.4)
<b>20 – 24 years</b>	266,841 (28.5)	28,656 (29.4)
<b>25 – 29 years</b>	280,532 (30)	26,073 (26.7)
<b>30 – 34 years</b>	221,138 (23.7)	21,318 (21.9)
<b>35 – 39 years</b>	75,123 (8)	9,242 (9.5)
<b>40 – 44 years</b>	12,138 (1.3)	2,086 (2.1)
<b>Nativity</b>		
<b>Born in the U.S.</b>	740,300 (79.2)	78,473 (80.4)
<b>Born outside the U.S.</b>	194,610 (20.8)	19,082 (19.6)
<b>High school-level degree</b>	295,986 (31.7)	37,228 (38.2)
<b>More than a High school-level degree</b>	638,924 (68.3)	60,327 (61.8)
<b>Cigarettes before 2<sup>nd</sup> trimester</b>		
<b>Yes</b>	54,359 (5.8)	6,914 (7.1)
<b>Body Mass Index</b>		
<b>Underweight</b>	35,063 (3.8)	4,018 (4.1)
<b>Normal</b>	434,138 (46.4)	40,256 (41.3)
<b>Overweight</b>	465,709 (49.8)	53,281 (54.6)
<b>Pre-pregnancy diabetes</b>		
<b>Yes</b>	6,075(0.65)	2,095 (2.1)

Table 2 (cont'd)

<b>Pre-pregnancy hypertension</b>		
<b>Yes</b>	14,818 (1.6)	3,869(4)
<b>Payor at Delivery</b>		
<b>Medicaid</b>	318,412 (34)	39,520 (40.5)
<b>Private insurance</b>	547,838 (59)	51,358 (52.7)
<b>Other</b>	68,660 (7)	6,677 (6.8)
<b>Prenatal Care Visit Timing</b>		
<b>No visit before 16 weeks</b>	108,137 (11.6)	15,303 (15.7)
<b>1-month Visit</b>	52,268 (5.6)	5,855 (6)
<b>2-month Visit</b>	405,069 (43.3)	37,080 (38)
<b>3-month Visit</b>	295,837 (31.6)	29,722 (30.5)
<b>4-month Visit</b>	73,599 (7.9)	9,595 (9.8)
<b>Infertility treatment used</b>		
<b>Yes</b>	20,020 (2.1)	2,882 (3)
<b>Race &amp; Ethnicity</b>		
<b>Hispanic</b>	205,153 (21.9)	22,015 (22.6)
<b>Non Hispanic AIAN</b>	5,514 (0.6)	649 (0.66)
<b>Non Hispanic AIAN &amp; Asian &amp; White (AIAsW)</b>	90 (0.01)	10 (0.01)

Table 2 (cont'd)

<b>Non Hispanic AIAN &amp; White (AIW)</b>	<b>3,336 (0.36)</b>	<b>338 (0.3)</b>
<b>Non Hispanic Asian</b>	75,968 (8.1)	6,912 (7.1)
<b>Non Hispanic Asian &amp; NHOPI (ANH)</b>	460 (0.05)	60 (0.06)
<b>Non Hispanic Asian &amp; White (AW)</b>	5,163 (0.55)	495 (0.5)
<b>Non Hispanic Asian &amp; NHOPI &amp; White (AsNHW)</b>	524 (0.06)	77 (0.08)
<b>Non Hispanic Black</b>	110,928 (11.87)	17,699 (18.1)
<b>Non Hispanic Black &amp; AIAN (BAI)</b>	528 (0.06)	81 (0.08)
<b>Non Hispanic Black &amp; AIAN &amp; White (BAIW)</b>	638 (0.07)	66 (0.07)
<b>Non Hispanic Black &amp; Asian (BAs)</b>	624 (0.07)	75 (0.08)
<b>Non Hispanic Black &amp; Asian &amp; White (BAsW)</b>	230 (0.02)	29 (0.03)
<b>Non Hispanic Black &amp; NHOPI (BNH)</b>	128 (0.01)	17 (0.02)
<b>Non Hispanic Black &amp; White (BW)</b>	9,464 (1.01)	1,042 (1.1)
<b>Non Hispanic NHOPI</b>	1,813 (0.2)	256 (0.26)
<b>Non Hispanic NHOPI &amp; White (NHW)</b>	522 (0.06)	44 (0.05)
<b>Non Hispanic White</b>	513,827 (55)	47,690 (48.9)

Table 2 (cont'd)  
Values are n (% total)

Table 3. Performance of the models when performed on the training data, original and oversampling datasets

Model	Measures	Original data	Oversampling
Logistic regression model 1	Accuracy	0.91	0.62
	Sensitivity	0.0008	0.50
	Specificity	99.9	0.64
	AUC	0.5	0.57
	Balanced Accuracy	0.5	0.57
Logistic regression model 2	Accuracy	0.91	0.63
	Sensitivity	0.0002	0.49
	Specificity	99.9	0.65
	AUC	0.5	0.57
	Balanced Accuracy	0.5	0.57
Decision Tree model	Accuracy	0.9	0.62
	Sensitivity	0.002	0.5
	Specificity	99.9	0.63
	AUC	0.5	0.57
	Balanced Accuracy	0.5	0.57
Random Forest model	Accuracy	0.91	0.65
	Sensitivity	0	0.47
	Specificity	1	0.67
	AUC	0.5	0.57
	Balanced Accuracy	0.5	0.57



Table 4. Predicted probability of preterm birth for three virtual observations

Observation characteristics	Term birth probability (%)	Preterm birth probability (%)
<b>Non Hispanic Asian</b> - More than high school education - No cigarette smoking before pregnancy - Private insurance - Age: 30 years - Born outside the US - First prenatal visit: 3rd month - BMI: Normal - No diabetes or hypertension history - No infertility treatment	100%	0%
<b>Non Hispanic Black &amp; Asian</b> - More than high school education - No cigarette smoking before pregnancy - Private insurance - Age: 30 years - Born outside the US - First prenatal visit: 3rd month - BMI: Normal - No diabetes or hypertension history - No infertility treatment	54.9%	45.1%

Table 4 (cont'd)

<b>Non Hispanic Asian &amp; NHOPI &amp; White</b>	72.6%	27.4%
- More than high school education		
- No cigarette smoking before pregnancy		
- Private insurance		
- Age: 30 years		
- Born outside the US		
- First prenatal visit: 3rd month		
- BMI: Normal		
- No diabetes or hypertension history		
- No infertility treatment		

## APPENDIX B: FIGURES

Figure 1. A snapshot of flow chart for the derivation of the analytic sample

Flow Chart: Illustrates derivation of the analytic sample

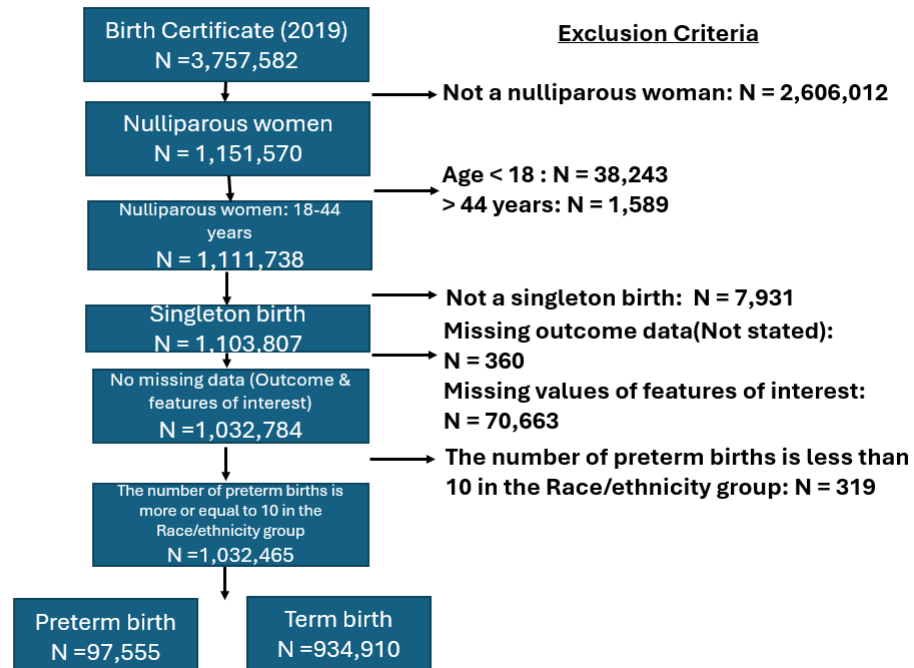


Figure 2. Importance figures of a Random Forest model using an oversampling dataset

