5-2023

Modeling Perinatal Outcomes in Disaggregated Asian American Subgroups

Suditi Shyamsunder
William & Mary

Modeling Perinatal Outcomes in Disaggregated Asian American Subgroups

A thesis submitted in partial fulfillment of the requirement
for the degree of Bachelors of Science in the Data Science Department from
William & Mary

by

Suditi Shyamsunder

Accepted for High Honors

Dana 2 W

Dr. Dana Willner, Chair

RDS

Dr. Ronald Smith

Dr. Dan Cristol

Adrian M. Bacong

Dr. Adrian Bacong

Williamsburg, VA
May 9, 2023

Modeling Perinatal Outcomes in Disaggregated Asian American Subgroups

By: Suditi Shyamsunder

Abstract

Asian mothers, as an aggregate, may be at increased risk for adverse perinatal outcomes, but heterogeneity between disaggregated Asian American subgroups is an understudied topic. The first objective of this study is to examine differences in perinatal outcomes between disaggregated Asian American subgroups and differences compared to Non-Hispanic Whites (NHWs). The second goal is to develop models to predict gestational diabetes, one perinatal outcome, in Asian Indian mothers to see how precision medicine may be able to advance pregnancy care. Using the National Vital Statistics System Natality Dataset (n=10,823,868), odds ratios (OR) were calculated with 95% confidence intervals (CI) for four perinatal outcomes (gestational diabetes, gestational hypertension, low birthweight, preterm birth) in six Asian subgroups (Indian, Chinese, Filipino, Japanese, Korean, Vietnamese) compared to NHWs and by nativity. The models adjusted for mother's age, educational attainment, pre-pregnancy hypertension, pre-pregnancy diabetes, and pre-pregnancy BMI. Additionally, three types of models were built to predict gestational diabetes in Asian American mothers (logistic regression, random forest, and gradient boosted). The calculated odds ratios showed that Asian Americans

generally had an increased risk for gestational diabetes and low birth weight, and a decreased risk for gestational hypertension and preterm birth; however, results varied between disaggregated subgroups. Additionally, foreign born Asians generally had an increased risk of gestational diabetes compared to US born Asians, and a decreased risk for the other three perinatal outcomes, but variation between subgroups persisted. In addition, accuracy and recall scores varied substantially between models. Additionally, undersampling techniques also impacted the success of the various models. Ultimately, the results show that Asian Americans face different risks in perinatal outcomes compared to NHWs, there is heterogeneity in results between disaggregated Asian subgroups, and machine learning can be used to make personalized risk predictions.

Introduction

Maternal outcomes in the United States (US) are ranked among the lowest of all developed countries. In particular, for every 100,000 live births in the US, there are 24 that result in maternal death, and this number has been rising since 2015. This number places the US in a position where the maternal mortality rate is over three times the rate in most other high-income countries (Hoyert, 2022).

Importantly, this risk of adverse maternal outcomes vastly differs according to social factors and race. Factors such as unmarried status, US born status, lower education, and rural residence are associated with a 50-114% higher risk of maternal mortality. Non-Hispanic Black women are at a 2.4 times higher risk of maternal mortality than White women (Singh, 2021). Additionally, Asian/Pacific Islander women are at a particularly high risk for gestational diabetes compared to other racial groups and are more likely to have macrosomic infants. Asian women also face an increased risk of severe perineal laceration and postpartum hemorrhage (Bryant et al., 2010). Race and societal inequalities seem to translate to maternal health disparities in the US. However, the way these disparities impact disaggregated Asian American populations in particular is an understudied but equally important topic to consider.

Asian Americans make up about 5% of the US population and are one of the fastest growing immigrant groups. As of 2022, there are currently about 20 million Asian Americans residing in the US, a number predicted to rise to 35.8 million by 2060 (Budiman & Ruiz, n.d.). However, Asian Americans make up a disproportionately small fraction of overall study participants, and are not as well studied in medical research such as maternal health research as other racial groups are (Y. Liu et al., 2019). Further, research has shown the need to not solely increase Asian American study participants, but also to study them disaggregated at the subgroup level (Yom & Lor, 2022). As will be further shown in this study, when Asian Americans are treated as an aggregated, monolithic group, the study results can hide further disparities within the heterogenous Asian American community.

Further, as Asians are an immigrant group, some of the disaggregated Asian mothers in the US are actually foreign born. Moreover, it has been found that nativity status also plays a role in health and maternal outcomes. Research shows that preterm birth, hypertensive disorders, low birth weight, and NICU admission are more likely to occur for U.S.-born women than their foreign-born counterparts (Adegoke, 2021). Another study that looked into how maternal nativity impacts birth outcomes in Asian immigrants found that mothers from different Asian subgroups experienced heterogeneity in risk when disaggregated further by nativity. For example, U.S.-born Chinese mothers and Japanese mothers had increased preterm births compared to foreign born counterparts (Qin & Gould, 2010). Understanding how nativity impacts Asian mothers has the potential to uncover further information about healthcare for this demographic group.

The goal of this study is twofold: 1) to understand the individual risk that Asian American mothers face for four main perinatal outcomes by subgroup and nativity status, and 2) to build models to see if accurate prediction of perinatal outcomes is possible for this demographic. Perinatal outcomes are defined as maternal and neonatal outcomes in the weeks leading up to and after birth. The four outcomes included in this analysis are gestational diabetes, gestational hypertension, low birthweight, and preterm birth. While the prevalence of these perinatal outcomes in America has been fairly well studied, it is not known how they impact Asian Americans, especially by disaggregated subgroups and nativity; thus it is imperative that focus is directed towards them. Additionally, as precision medicine becomes increasingly utilized to improve patient care, predictive models may be used to help personalize care based on individual risk (Hulsen et al., 2019). Conventional epidemiological modeling relies on odds ratios that are derived from frequentist statistics, and a newer methodology involves predictive machine learning models to help understand relationships in the data. This study employs both of these techniques. The analysis worked to determine how perinatal outcomes vary in disaggregated Asian American subgroups compared to the reference group of NHWs, the effects of nativity on these outcomes, and how these outcomes can be modeled and predicted for the Asian American demographic.

Methods

The methodology for this project follows two main themes: Generalized Linear Modeling (GLM) using R to calculate odds ratios (OR), and modeling using Machine Learning methods with Python to make predictions. For both components, I examined the National Vital Statistics System Natality (NVSS) data from the years 2015-2019 (https://www.cdc.gov/nchs/nvss/births.htm). The inclusion criteria consisted of all births from Non-Hispanic White mothers and all births from Asian mothers. This consisted of 10,823,868 data points to analyze.

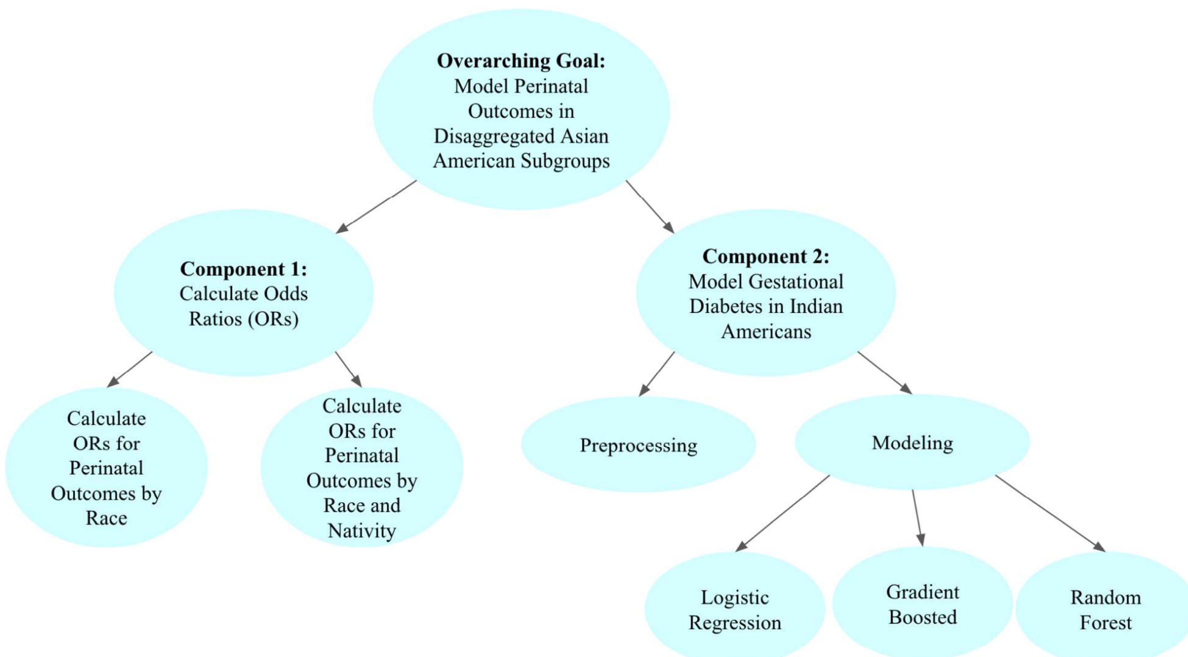

Figure 1: Methods Overview

The four outcomes included in our study are gestational diabetes, gestational hypertension, low birthweight, and preterm birth. Gestational diabetes is defined as elevated blood sugar or hyperglycemia during pregnancy. According to the American College of Obstetricians and Gynecologists (ACOG), this is measured by a 2 step procedure: A Glucose Challenge Test (GCT) with 50g glucose non-fasting, and if the value > 7.8mmol/l, it is followed by 3-hour Oral Glucose Tolerance Test (OGTT) to confirm (Rani & Begum, 2016). These results are recorded on the birth record. Gestational hypertension is high blood pressure during pregnancy, which most guidelines define as blood pressure ≥140/90 mm Hg and is recorded on the birth record (Garovic, 2021). Low birthweight was defined as the infant having a weight below 2500 grams at birth, and preterm birth was defined as birth before 37 weeks of gestation (Hughes et al., 2017; Suman & Luther, 2023).

Odds Ratios

Before explaining how odds ratios were calculated, here is a brief overview of how to interpret odds ratios. Odds ratios of less than 1 indicate a decreased probability or risk compared to the reference group. Odds ratios of greater than 1 indicate an increased risk compared to the reference group. Odds ratios that are close to 1 indicate an equal risk to the reference group. To provide some additional context, an odds ratio of 3 indicates a probability 3 times greater than what it is being compared to. For example, there is an odds ratio of 3 when looking at the probability of rolling an even number on a standard six-sided die (2 ,4, 6) compared to the reference group of rolling a 2.

ORs for Perinatal Outcomes by Race

In order to determine whether odds ratios differed between Asian American subgroups based on race and nativity, I created four generalized linear models in R, one for each of the four perinatal outcomes. Race served as the predictor variable and the other variables included in the model (mother's age, mother's education, prepregnancy hypertension, prepregnancy diabetes, and prepregnancy bmi) were included as potential confounders. I then calculated the odds ratio for each racial subgroup by exponentiating the model coefficient for that group (i.e. $e^{coefficient}$).

ORs for Perinatal Outcomes by Race and Nativity

In order to explore ORs for perinatal outcomes by race and nativity, I created a dataframe for each racial group by using pandas and then four generalized linear models (one with each perinatal outcome) in R for each race category. I then calculated ORs as explained above for each nativity value, which represents the odds ratio for the foreign born racial group versus the corresponding US born racial group. The fully adjusted model for gestational diabetes and gestational hypertension accounts for mother's age, educational attainment, pre-pregnancy hypertension, pre-pregnancy diabetes, and pre-pregnancy BMI. The fully adjusted model for low birthweight and preterm birth additionally accounts for gestational diabetes, gestational hypertension, and hypertension eclampsia.

Machine Learning

This component of my honors thesis is conducted as a pilot study to show how perinatal outcomes can be predicted in Asian American subgroups. For this reason, only one outcome and one subpopulation was chosen as an example of the kind of work that can be done. Specifically, of all the perinatal outcomes and Asian subgroups, I chose to model gestational diabetes in the Asian Indian population because as is shown later in the results, this is the highest odds ratio that was calculated. This shows that there is a particularly high risk for this demographic to develop this condition.

Additionally, the code developed to preprocess and create the machine learning models was initially based on a smaller subset of data (n=3,663) that was about 1/100th the size of the total data (n=368,638). I verified that this subset was representative of the larger population by comparing the overall characteristics of the datasets, and these tables are shown in Appendix A.1, A.2, and A.3.

Preprocessing

Before the modeling component of the thesis could begin, further preprocessing was required. All data pre-processing steps were carried out in Python using the pandas and numpy libraries. More specifically, there were certain variables in the original dataset that were repetitive or completely unnecessary when it came to modeling. I dropped those columns that were deemed unhelpful to the model. I also had to do additional recoding. Some missing values were encoded as 99s or other numerical values, so I analyzed each column name and process of encoding in order to change the relevant values to nan. Next, I dropped any rows where the target variable (gestational diabetes) was missing. I also dropped rows with greater than 5 missing values because after further examination, I saw that when there were at least that many values missing in a row, almost every value tended to be missing.

The next step required before modeling was imputation. I employed k-nearest neighbors (KNN) imputation to fill in missing values. Next, I one-hot encoded the categorical variables and ensured that each categorical variable had n-1 columns where n is the number of categories. One hot encoding is a process that encodes categorical variables into columns of 0s and 1s in preparation for modeling (Hancock & Khoshgoftaar, 2020). Additionally, I condensed certain variables such as mother's education into fewer categories because some of the original categories only occurred at a very low frequency.

Modeling

All of the machine learning based modeling was performed using Python's scikit-learn module. The three types of models used were logistic regression, random forest, and gradient-boosted classification. All models were evaluated using precision, recall, and accuracy, with a particular emphasis on recall as a metric for success. Due to an imbalance in the dataset, i.e., there were far more data points from people without gestational diabetes than with, I tested three undersampling strategies: random undersampling of the data prior to splitting into test and train sets, random undersampling of just the training data after splitting, and undersampling the training data using NearMiss-1 as implemented in the Python imbalance package (J. Zhang & Mani, 2003). Pre-split undersampling was to 504 observations per class for the small dataset, and 51224 for the large dataset. Based on the 30% test-train split used for model training and validation, post-split undersampling was to 228 observations per class for the small data set, and 24086 for the large dataset.

I then performed logistic regression using Python's scikit-learn module. Logistic regression is a special case of a linear regression model where the target variable is categorical rather than numerical (Geron, 2019). Cross-validation with 10 folds was used to optimize the value of the hyperparameter c. I also tested both L1 and L2 regularization, which required the use of the saga solver. The default value of c=1 was optimal according to the metrics I evaluated, and ultimately, L1 regularization was used to evaluate feature importance.

Next, using scikit-learn, I created random forest models. Random forest classification works by fitting the data to numerous decision trees and then applying soft voting to average class probabilities to make a final prediction. I also explored feature importance with this classifier. The feature importances were evaluated by looking at the mean decreases in gini impurity. This metric counts the number of times a particular feature is used to split a node, while considering the total number of samples it splits (*Feature Importances with a Forest of Trees*, n.d.). The features with the greatest gini impurity values are considered the most important.

Finally, I also employed a gradient boosted classifier in scikit-learn in order to see how its predictions performed compared to the other model types. The gradient boosted classifier is an ensemble learning method where multiple weaker models are combined to create a stronger final model that learns and improves with each iteration (Natekin & Knoll, 2013). Additionally, the default value of the learning_rate parameter (0.1) was determined to be optimal according to the metrics I evaluated (*Sklearn.Ensemble.RandomForestClassifier*, n.d.).

Results/Discussion

Odds Ratios

Data Characteristics

Birth data from a total of 11,006,202 mothers (10,134,723 Non-Hispanic White & 871,479 Asian American) in the US between 2015-2019 were included in this study. As shown in Table 1, some demographic characteristics did vary slightly between each racial group. For example, the average age of the NHW group was younger than that of the Asian subgroups. The NHW average age was 29.3, but of the Asian subgroups, foreign born Indians had the lowest average age of 31.1, which is slightly greater than for NHWs. However, overall, each racial group had a relatively similar demographic makeup. Table 2 displays additional information about the births in the dataset and specifically focuses on perinatal and neonatal outcomes. These first two tables show an overview of the data by race that is used in the remainder of the analysis.

Variables	Non Hispanic White (n=10134723)	Indian (n=370791)	Chinese (n=293324)	Filipino (n=151366)	Japanese (n=32221)	Korean (n=72711)	Vietnamese (n=102432)
Maternal age (y)	29.3	31.23	32.4	32.12	34.41	33.61	31.92
Maternal height (in)	64.95	63.39	63.85	62.25	62.96	63.81	62.21
Pre-pregnancy	175.07	159.86	148.82	152.13	144.91	147.59	138.33

body weight (lb)							
Pre-pregnancy diabetes N (%)	73932 (0.73%)	4079 (1.10%)	1539 (0.52%)	1807 (1.19%)	173 (0.54%)	537 (0.74%)	917 (0.90%)
Pre-pregnancy Hypertension N (%)	179207 (1.77%)	3194 (0.86%)	1429 (0.49%)	3454 (2.28%)	285 (0.88%)	755 (1.04%)	631 (0.62%)
Maternal Education, N (%)							
8th grade or less	132494 (1.31%)	4666 (1.26%)	5017 (1.71%)	361 (0.00%)	37 (0.11%)	120 (0.17%)	2395 (2.34%)
9th through 12th grade with no diploma	594047 (5.86%)	10053 (2.71%)	10839 (3.70%)	3353 (2.22%)	414 (1.28%)	358 (0.49%)	6551 (6.40%)
High school graduate or GED completed	2124020 (20.96%)	27092 (7.31%)	27709 (9.45%)	16591 (10.96%)	2444 (7.59%)	3207 (4.41%)	22484 (21.95%)
Some college credit, but not a degree	1996927 (19.70%)	18294 (4.93%)	22857 (7.79%)	29103 (19.23%)	3042 (9.44%)	6602 (9.08%)	15948 (15.57%)
Associate degree (AA, AS)	963032 (9.50%)	12353 (3.33%)	16116 (5.49%)	16437 (10.86%)	3746 (11.63%)	4150 (5.71%)	8825 (8.62%)

Bachelor's degree (BA, AB, BS)	2619298 (25.84%)	131066 (35.35%)	96473 (32.89%)	64360 (42.52%)	14775 (45.86%)	32667 (44.93%)	27485 (26.83%)
Master's degree (MA, MS, MEng, MEd, MSW, MBA)	1213930 (11.98%)	124626 (33.61%)	70968 (24.19%)	10927 (7.22%)	4883 (15.15%)	14979 (20.60%)	7768 (7.58%)
Doctorate (PhD, EdD) or Professional Degree (MD, DDS, DVM, LLB, JD)	340129 (3.36%)	31608 (8.52%)	31453 (10.72%)	4781 (3.16%)	1921 (5.96%)	8654 (11.90%)	7003 (6.84%)
Foreign born, N (%)	697345 (6.88%)	331387 (89.4%)	252212 (86.0%)	108551 (71.7%)	23159 (71.9%)	53647 (73.8%)	80141 (78.2%)
Perinatal visits	7.15	7.11	6.99	6.95	7.09	7.12	6.85

Table 1: Maternal Characteristics By Race

Variables	**Non Hispanic White (n=10134723)**	**Indian (n=370791)**	**Chinese (n=293324)**	**Filipino (n=151366)**	**Japanese (n=32221)**	**Korean (n=72711)**	**Vietnamese (n=102432)**
Gestational weight gain (lb)	31.27	27.44	29.06	28.44	25.56	29	28.39

Gestational diabetes mellitus, N (%)	73932 (0.73%)	4079 (1.10%)	1539 (0.52%)	1807 (1.19%)	173 (0.54%)	537 (0.74%)	917 (0.90%)
Gestational hypertension, N (%)	707415 (6.98%)	14879 (4.01%)	7088 (2.42%)	10530 (6.96%)	1034 (3.21%)	2669 (3.67%)	3046 (2.97%)
Eclampsia, N (%)	24971 (0.25%)	511 (0.14%)	274 (0.09%)	976 (0.64%)	177 (0.55%)	130 (0.18%)	101 (0.10%)
Gestational age at delivery (week)	38.57	38.47	38.66	38.18	38.59	38.68	38.4
Preterm birth (<37 weeks), N (%)	919944 (9.08%)	33262 (8.97%)	19655 (6.70%)	17654 (11.66%)	2570 (7.98%)	5411 (7.44%)	8945 (8.73%)
Birth Weight: Very low (<1,500 g), N (%)	108797 (1.07%)	5074 (1.37%)	2018 (0.69%)	2354 (1.56%)	267 (0.83%)	591 (0.81%)	1001 (0.98%)
Birth Weight: Low (<2,500 g), N (%)	598409 (5.90%)	33782 (9.11%)	15607 (5.32%)	12830 (8.48%)	2453 (7.61%)	3973 (5.46%)	7170 (7.00%)
Birth Weight: Normal/High (>=2500 g), N (%)	9420383 (92.95%)	331706 (89.46%)	275560 (93.94%)	136126 (89.93%)	29489 (91.52%)	68098 (93.66%)	94224 (91.99%)

Delivery: Spontaneous, N(%)	6646690 (65.58%)	201941 (54.46%)	189439 (64.58%)	90329 (59.68%)	22254 (69.07%)	46483 (63.93%)	64809 (63.27%)
Delivery: Forceps, N(%)	61500 (0.61%)	3219 (0.87%)	1872 (0.64%)	1033 (0.68%)	359 (1.11%)	586 (0.81%)	637 (0.62%)
Delivery: Vacuum, N(%)	268817 (2.65%)	18540 (5.00%)	14538 (4.96%)	5664 (3.74%)	1258 (3.90%)	3333 (4.58%)	4828 (4.71%)
Delivery: Cesarean, N(%)	3116880 (30.75%)	144941 (39.09%)	86835 (29.60%)	53800 (35.54%)	8274 (25.68%)	21946 (30.18%)	32045 (31.28%)

Table 2: Perinatal and Neonatal Outcomes By Race

Odds Ratios by Disaggregated Asian Subgroup

The results indicate that for gestational diabetes, Asian Americans tend to have an increased risk compared to NHWs overall. All disaggregated groups had a higher risk than NHWs, but the degree to which the risk increased differed between groups. Japanese Americans for example, had an odds ratio close to 1 (OR: 1.06, 95% CI 1.02-1.11), but on the other hand, Asian Indians had an almost 3-fold increase in risk (OR: 2.75, 95% CI 2.72-2.77). Overall, this increased risk for gestational diabetes among Asian mothers is consistent with previous studies (L. Chen, 2019).

For gestational hypertension, Asian Americans tended to have decreased odds. With the exception of Filipino Americans (OR: 1.0, 95% CI 0.98-1.02), all other disaggregated Asian American groups had an odds ratio that dipped below 1 and confidence intervals that did not include 1. The most significant of these findings was for Chinese Americans who were ⅓ as likely to be diagnosed with gestational hypertension compared to NHWs (OR: 0.33, 95% CI 0.33-0.34). This is consistent with findings from a study examining disparities in maternal hypertension in the United States, which concluded that most Asian groups (including Chinese, Japanese, Vietnamese, Koreans, and Asian Indians) had a lower prevalence of maternal hypertension than NHWs (Singh, 2018). Another similar study corroborated these findings with results that show that Asian Americans are in general more likely to remain normotensive during pregnancy (Ghosh et al., 2014).

Additionally, with the exception of Chinese Americans (OR: 0.91, 95% CI 0.90-0.93) and Korean Americans (OR: 0.96, 95% CI 0.93-0.99), Asians tended to have an increased odds of low birthweight. Asian Indians are at the highest risk (OR: 1.75, 95% CI 1.73-1.76). The overall trend for aggregate Asians is consistent with the literature as shown by Zang's study, which explored perinatal outcomes in Ontario, Canada; the results showed increased risk for low birthweight (Zeng et al., 2021). Additionally, the highest risk within the Asian Indian subgroup aligns with the results of another research study that found lower birth weight in South Asian babies that were born in the United Kingdom (Margetts et al., 2002).

For preterm birth, we found that the Asian American subgroups tended to have odds ratios around 1, which indicate a comparable risk compared to NHWs. Additionally, for preterm birth there were differences by disaggregated Asian group, and Chinese mothers had the lowest risk (OR: 0.75, 95% CI 0.74-0.77), while Filipinos had the highest (OR: 1.29, 95% CI 1.27-1.31). These results reinforce the need to disaggregate Asian American research into subgroups. As can be seen from the table above, the odds ratios for aggregate Asians do not show the full picture of the risks for Asian subgroups compared to the NHW reference group. The aggregate Asian OR for preterm birth was close to 1, indicating that the risk didn't differ much from that of NHWs. However, when looking at the ORs for some of the other Asian subgroups, the results are quite heterogeneous and show ORs from 0.75 to 1.29 as described above.

	Gestational Diabetes	**Gestational Hypertension**	**Low Birth Weight**	**Preterm Birth**
	OR (95% CI)	OR (95% CI)	OR (95% CI)	OR (95% CI)
Non-Hispanic White (n=10134723)	Ref.	Ref.	Ref.	Ref.
Indian (n=370791)	2.75 (2.72, 2.77)	0.56 (0.55, 0.57)	1.75 (1.73, 1.76)	1.06 (1.05, 1.08)
Chinese (n=293324)	1.86 (1.84, 1.89)	0.33 (0.33, 0.34)	0.91 (0.90, 0.93)	0.75 (0.74, 0.77)
Filipino (n=151366)	2.09 (2.06, 2.12)	1.00 (0.98, 1.02)	1.49 (1.46, 1.51)	1.29 (1.27, 1.31)
Japanese (n=32221)	1.06 (1.02, 1.11)	0.43 (0.41, 0.46)	1.24 (1.20, 1.29)	0.85 (0.82, 0.89)
Korean (n=72711)	1.42 (1.39, 1.46)	0.51 (0.50, 0.54)	0.96 (0.93, 0.99)	0.83 (0.81, 0.85)

Vietnamese (n=102432)	2.15 (2.11, 2.20)	0.41 (0.40, 0.43)	1.15 (1.13, 1.18)	0.95 (0.93, 0.97)
Aggregate Asians (n=1022845)	2.05 (2.02, 2.07)	0.56 (0.55, 0.57)	1.26 (1.24, 1.28)	0.96 (0.95, 0.97)

Table 3. Fully Adjusted Odds of Perinatal Outcomes by Race

Odds Ratios by Disaggregated Asian Subgroup and Nativity

This section of the study examines the odds ratios of the four main perinatal outcomes for each racial category by nativity. Compared to their US born counterparts, foreign born Asian Americans did have slightly different risks. For gestational diabetes, with the exception of Japanese Americans (OR: 0.58, 95% CI 0.53-0.63) and Korean Americans (OR: 1.00, 95% CI 0.94-1.06), foreign born Asians tended to have an increased risk. The highest risk group was foreign born Indians (OR: 1.72, 95% CI 1.66-1.78).

For gestational hypertension, all foreign born disaggregated Asian subgroups experienced a decreased risk. The lowest risk group was foreign born Japanese Americans (OR: 0.40, 95% CI 0.35-0.46). For low birth weight and preterm birth, foreign born Asian Americans tended to have a slightly lower risk. With the exception of Korean Americans (OR: 0.93, 95% CI 0.86-1.00), foreign born Asians had a decreased risk of low birthweight with the smallest risk belonging to Chinese Americans (OR: 0.73, 95% CI 0.71-0.77). Additionally, all foreign born disaggregated Asian subgroups had a decreased risk of preterm birth with the smallest risk belonging to foreign born Japanese Americans (OR: 0.64, 95% CI 0.59-0.70).

In general, the literature shows mixed results on how nativity impacts the perinatal health of mothers. The results of the study presented here show that with the exception of gestational diabetes, for each perinatal outcome, US Born Asians tended to experience a higher risk. This is consistent with the results of a study that examined acculturation and its impact on gestational diabetes. The results showed that in general, Asians do have an increased risk for this perinatal outcome but that higher acculturation may be a protective factor in this demographic group, and individuals who are born in the US tend to be more acculturated than their foreign-born counterparts (L. Chen, 2019). That study also discussed how other parts of the literature suggest the opposite impact of acculturation; some studies show that acculturation has led to decreased health and increased diabetes rates in Hispanic Americans (O'Brien et al., 2014) and Chinese individuals in Australia (Jin et al., 2017). It is possible that increased acculturation in the US culture may lead to a worsening diet and health that may be a factor in the increased risk for US born Asians shown in table 4.

	Gestational Diabetes	**Gestational Hypertension**	**Low Birth Weight**	**Preterm Birth**
	OR (95% CI)	OR (95% CI)	OR (95% CI)	OR (95% CI)
Foreign Born Indian (n=331387)	1.72 (1.66, 1.78)	0.80 (0.76, 0.84)	0.88 (0.85, 0.91)	0.90 (0.87, 0.93)
Foreign Born Chinese (n=252212)	1.13 (1.09, 1.17)	0.44 (0.42, 0.46)	0.73 (0.71, 0.77)	0.75 (0.72, 0.78)
Foreign Born Filipino (n=108551)	1.25 (1.20, 1.29)	0.89 (0.85, 0.93)	0.92 (0.88, 0.95)	0.96 (0.92, 0.99)

Foreign Born Japanese (n=23159)	0.58 (0.53, 0.63)	0.40 (0.35, 0.46)	0.81 (0.74, 0.88)	0.64 (0.59, 0.70)
Foreign Boren Korean (n=53647)	1.00 (0.94, 1.06)	0.87 (0.79, 0.95)	0.93 (0.86, 1.00)	0.95 (0.88, 1.01)
Foreign Born Vietnamese (n=80141)	1.34 (1.27, 1.41)	0.48 (0.45, 0.53)	0.85 (0.81, 0.90)	0.81 (0.77, 0.85)
Foreign Born Aggregate Asians (n=849097)	1.62 (1.61, 1.63)	0.54 (0.54, 0.55)	0.97 (0.96, 0.98)	0.82 (0.82, 0.83)

Table 4. Fully Adjusted Odds of Perinatal Outcomes by Race and Nativity

Machine Learning

Evaluating Model Success

The odds ratio results from the previous section demonstrated how various factors influence the heterogeneity of risk that diverse groups of mothers face during pregnancy. I next sought to determine how personalized risk may be predicted using machine learning models. Additionally, this process served as a methodological exploration to determine if machine learning methods could be applied to this dataset and produce meaningful predictive power that could be useful clinically.

Each model was evaluated using accuracy, precision, and recall. Accuracy is calculated by taking the ratio of the number of correctly classified data points to the total number of data points. By the nature of their calculations, precision and recall always have a tradeoff; precision tells you what proportion of samples classified as positive were done correctly, while recall shows what proportion of positive samples were classified as such in the model (Hicks et al., 2022). Generally, as recall increases, there are more false positives, which decreases precision. In the results below, I have included both accuracy and recall, and chosen to exclude precision. This is because in medical applications of artificial intelligence, recall is considered to be the most important metric (Hicks et al., 2022). In the case of the gestational diabetes models that were built, models with higher recall are able to classify a greater number of patients with gestational diabetes properly. This metric has increased significance because in the clinical setting, physicians are most interested in being able to determine which patients are most likely to develop gestational diabetes so that additional care can be taken to prevent or mitigate any subsequent risks. In general, the additional monitoring of a patient's health would not be detrimental to anyone in the false positive category.

Additionally, the results for the smaller dataset were also calculated as it was initially used to evaluate and tune the models. After the models were finished, they were tested on the large dataset. Notably, the results stayed pretty consistent between the small and large dataset; for this reason, the large dataset results are primarily highlighted for each model below.

Results Without Undersampling

Initially, when the models were built without undersampling, accuracy values tended to be quite high, but recall values were very low. With such imbalanced data that had far more patients without gestational diabetes than those that did, the model was able to achieve a high accuracy by classifying almost all of the data points as 0, i.e. no gestational diabetes. However, this in practice is a fairly useless model because the recall is so low; low recall in the clinical setting means that physicians are not able to predict which patients are at higher risk for gestational diabetes and therefore need additional screening and prevention assistance.

The table shown below demonstrates the high accuracy but low recall of the models before undersampling. Notably, the models all had an accuracy of about 0.86, which is among the highest seen in any of the results of this study. However, the highest recall score observed was 0.04 by the random forest classifier, which is very low. This means that only 4% of individuals with gestational diabetes were detected by the model. The random forest model was not very successful at predicting the outcome on the test data, but still slightly more successful than the others. It is possible that this was the case because tree-based algorithms tend to perform particularly well with imbalanced data (Feki, 2022; C. Chen, 2004). Notably the model performed perfectly on the training data, i.e. an accuracy and recall of 100%, indicating overfitting.

Model Type	**Accuracy**	**Recall**
Logistic Regression (L1)	0.86	0.01
Logistic Regression (L2)	0.86	0.01
Random Forests	0.86	0.04

Gradient Boosted	0.86	0.02

Table 5: Accuracy and Recall Values Without Undersampling

Results With Undersampling After Train Test Split

Once the imbalance in the dataset became clear, results were evaluated after performing undersampling to rebalance the data. The tables below show the accuracy and recall values of the models before and after undersampling of two kinds: random undersampling and NearMiss undersampling.

Compared to no undersampling, the models with undersampling had reduced accuracy, but both of the undersampling methods drastically increased the recall success of the models. In terms of accuracy, the random undersampling technique outperformed NearMiss, but the opposite was true for recall. Since recall is most important in this study, the NearMiss undersampling technique seems to have outperformed random undersampling. The NearMiss results had recall values at or above 0.7. These results may be related to the way the NearMiss algorithm works. NearMiss does not use random undersampling, and instead, it uses an approach similar to k-Nearest Neighbors. NearMiss-1 preferentially selects the N (default N =3) points from the majority class that are closest to the points from the minority class for undersampling (*Under-Sampling*, n.d.; J. Zhang & Mani, 2003). Since it is a generally more sophisticated algorithm, it makes sense that it outperformed random undersampling in terms of recall.

Model Type	Accuracy	Recall
Logistic Regression (L1)	0.49	0.55
Logistic Regression (L2)	0.49	0.55
Random Forests	0.52	0.49
Gradient Boosted	0.50	0.53

Table 6: Accuracy and Recall Values With Random Undersampling

Model Type	Accuracy	Recall
Logistic Regression (L1)	0.47	0.70
Logistic Regression (L2)	0.47	0.70
Random Forests	0.35	0.79
Gradient Boosted	0.38	0.78

Table 7: Accuracy and Recall Values With NearMiss Undersampling

Results With Pre-Split Undersampling

After employing the train test split method after undersampling the data, accuracies and recall scores were the highest overall across models. In terms of accuracy, the gradient boosted model scored highest with an accuracy of 0.67. On the other hand, both the logistic regression models scored lowest with an accuracy of 0.64.

In terms of recall, the gradient boosted model once again performed the best with a value of 0.69 for the large dataset. The logistic regression model performed the worst with a recall of 0.63. Interestingly, in terms of both accuracy and recall, the gradient boosted model performed among the worst in terms of the small dataset but the best with the large dataset.

Model Type	**Accuracy**	**Recall**
Logistic Regression (L1)	0.64	0.63
Logistic Regression (L2)	0.64	0.63
Random Forests	0.66	0.67
Gradient Boosted	0.67	0.69

Table 8: Accuracy and Recall Values With Pre-Split Undersampling

While these values tend to be higher than in the section prior when train test split preceded undersampling, they may be biased due to data leakage (Yagis et al., 2021; Samala, 2020). A comparison between the accuracy and recall scores in this section of the results as compared to the one prior is a strong example of data leakage. Data leakage occurs when information is available to the model during training from the test or validation data that it may not have when actually put into practice (*8. Common Pitfalls and Recommended Practices — Version 0.10.1*, n.d.). Specifically in this case, if we undersample prior to splitting the data, the test data will be balanced, which isn't representative of data that the model would encounter when actually used for prediction. This leads to results that are deceptively good and show much higher accuracy and recall than might occur when making predictions on new test cases (Santos et al., 2018).

Comparing Results By Undersampling Method

The figures below compare the accuracies and recall scores of the models employing the various undersampling techniques described above. Overall, the original data with no undersampling performed the best in terms of accuracy, but this is deceptive because the recall scores were very low. Additionally, the pre-split undersampling seemed to perform well in terms of accuracy as well, but due to data leakage, these results may be overly optimistic. When undersampling with NearMiss, relatively high recall was able to be achieved, which is the most important in this study.

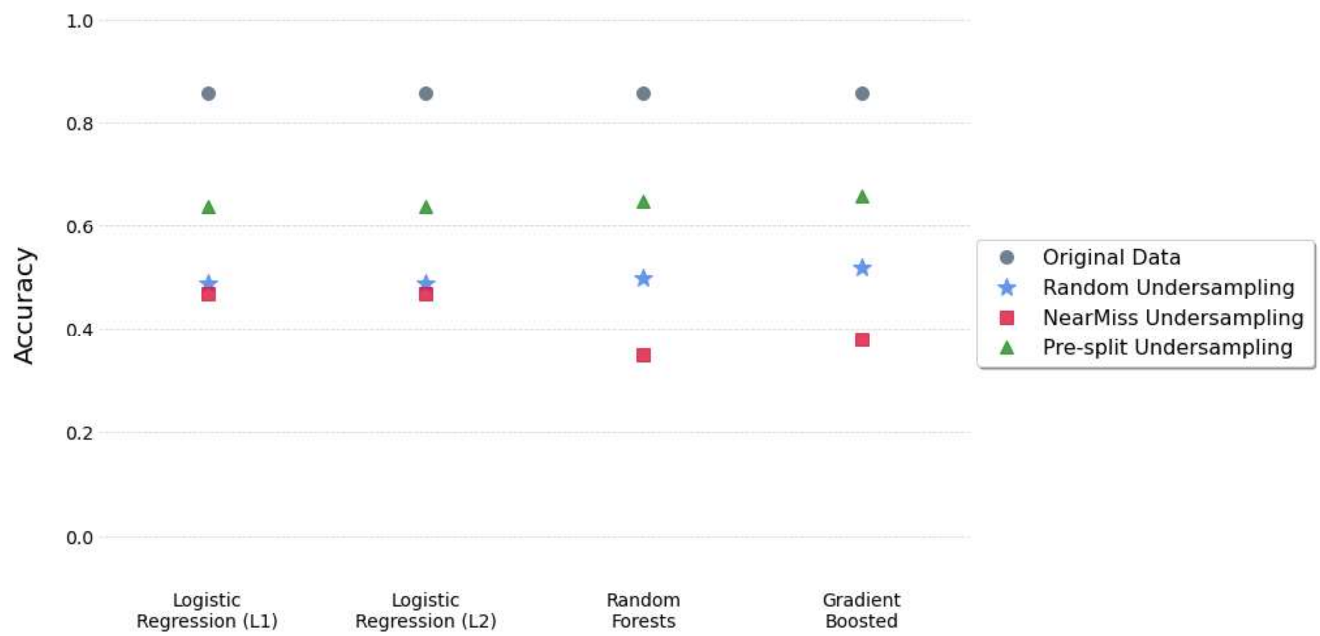

Figure 2: Accuracy Values By Undersampling Method

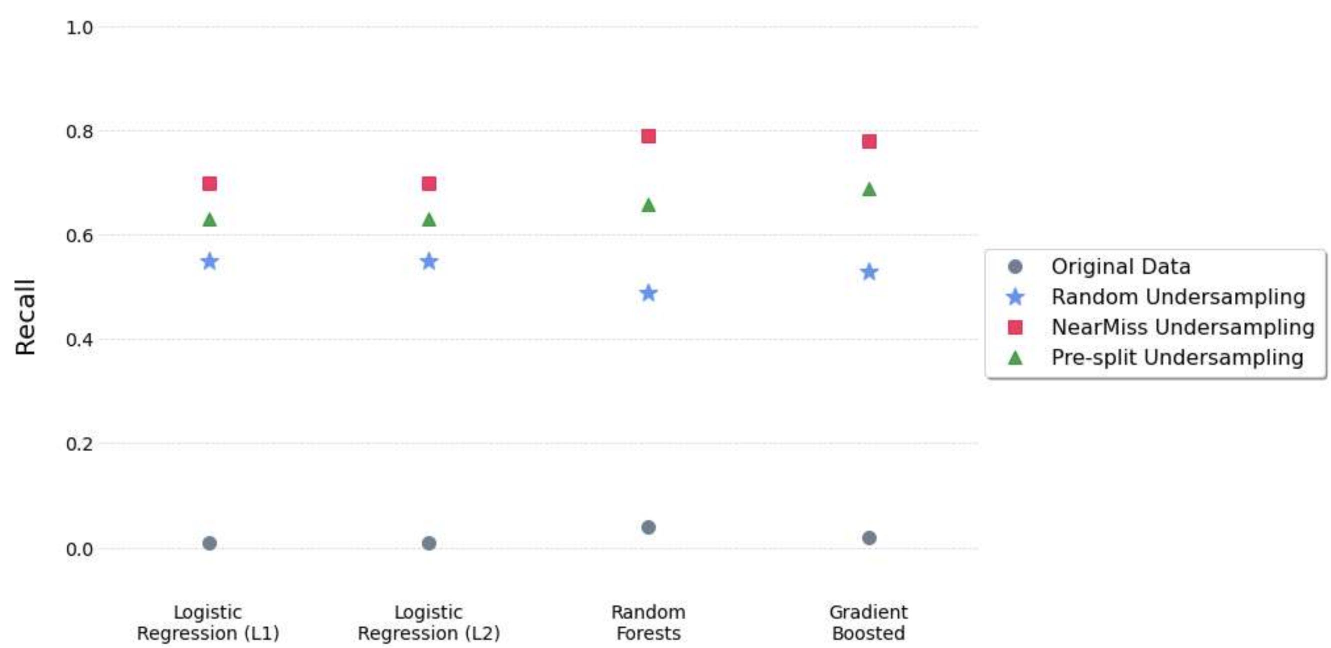

Figure 3: Recall Values By Undersampling Method

Evaluating Model Performance

Overall, it appears that the random forest and gradient boosted models outperformed the linear regression models in terms of accuracy and recall. There are many reasons why this may have been the case. First, logistic regression models use a linear equation with a logit link, i.e. a formalized regression model, and this is not the case with the random forest or gradient boosted classifier. Therefore, logistic regression makes more assumptions about the data. With real world data, these assumptions may not apply, making the model less accurate at prediction (Ranganathan et al., 2017).

Additionally, the gradient boosted and random forest classifier may have performed better because of their abilities to handle the type of data provided for the model. In general, ensemble models tend to be more effective at handling imbalanced data, especially when techniques such as oversampling or undersampling are used (L. Liu et al., 2022).

Feature Importances

Using both the logistic regression model with L1 regularization and the random forest classifier, I compared and contrasted feature importance values using the undersampled data. More specifically, the L1 regularization of the logistic regression model allows for coefficients to be set to zero. Those variables are effectively removed from the model and therefore deemed less important. The random forest classifier ranks feature importance by calculating a Gini impurity value of which a larger number indicates a more important variable.

With the random forest model, there were five variables that were deemed most important for prediction: BMI, weight gain, prepregnancy weight, delivery weight, and mother's age in descending order of importance. Notably, the first four of these variables are related to weight and body composition, which supports the idea that managing weight carefully during and even before pregnancy may be important in reducing gestational diabetes. When comparing this to the logistic regression with the L1 model, there were quite a few similarities. The most important predictors for the L1 model were determined on the basis of the magnitude of the coefficients. This metric does not always align perfectly, e.g. in cases of collinearity, but it can serve as a proxy for feature importance. The predictors deemed most important were weeks of gestation, cigarette use before pregnancy, weight gain, BMI, and prepregnancy weight in decreasing order of importance. These last three align with the random forest results, but the first two provide additional information about what may be important in these predictions. Additionally, connecting back to the GLMs created earlier in this study, this shows the importance of controlling for factors such as BMI/weight which are interrelated and may have some influence on odds of developing gestational diabetes.

Overall, there was a trend in which variables were considered particularly important in prediction for both of the models. Both models found metrics relating to weight and body composition to be particularly important. This aligns well with the current literature, which shows that body weight and larger composition is associated with increased risk of developing gestational diabetes (Rahnemaei et al., 2022). Additionally, current literature also supports the idea that a feature such as mother's age, which was deemed important by the random forest model may also play a role in gestational diabetes risk. In general, women of advanced maternal age (AMA) are considered to be at increased risk for numerous pregnancy outcomes such as gestational diabetes, miscarriage, chromosomal abnormalities, stillbirth, fetal growth restriction, preterm birth, pre-eclampsia, and cesarean section (Frick, 2021). Finally, cigarette use before pregnancy has also been considered to lead to increased risk of adverse outcomes whether before or during pregnancy (Yang et al., 2022).

Table 9. Most Important Features

Random Forest	**Logistic Regression - L1**
1. BMI	1. Weeks of Gestation
2. Weight gain	2. Cigarette Use Before Pregnancy
3. Prepregnancy Weight	3. Weight Gain
4. Delivery Weight	4. BMI
5. Mother's Age	5. Prepregnancy Weight

Limitations and Future Directions

There are some limitations of this study that can potentially be addressed by future work. One of the complications within our data from NVSS is that race isn't always as easily categorized as we assume for the purposes of this research. For example, we have not taken into account the health outcomes of individuals who affiliate with multiple racial groups. Additionally, in this study, we chose to ignore the fact that some mothers may have been duplicated in the data due to multiple births within the time frame of data collection. Additionally, while this work focuses on disaggregating Asians by subgroup, we were only able to categorize into six Asian subgroups, and this by no means truly captures the heterogeneity of Asia. Future studies could expand upon this work by disaggregating Asian groups further and exploring additional perinatal and maternal health outcomes.

Additionally, one complication in the study was the imbalance in the data. In future research, the study could be designed in a way that there would be a more even number of patients who have or don't have each perinatal outcome. For example, this might be done using a matched case-control study. Additionally, in order for models to be used in clinical settings, improved accuracy and recall may be needed. Further studies that work to improve the model success would be beneficial. Since some of the models perform stronger in some aspects than others, an ensemble-style model could also be examined.

Conclusion

As medical data continues to be collected, the healthcare field is beginning to uncover the power of personalized medicine. There is a growing need to understand how an individual's risk of developing adverse outcomes differs from another individual's on the basis of various demographic characteristics.

In this study, I sought to uncover how the odds of developing adverse perinatal outcomes differs between disaggregated Asian American subgroups as well as on the basis of nativity. Results showed that risk does indeed vary between disaggregated subgroups, reinforcing the heterogeneity of Asian Americans and the need to study them further. The hope is that understanding relative odds of developing these conditions will help providers to better care for their patients.

Additionally, this pilot study aimed to determine how machine learning models can be used to provide precision medical care. One major issue in the models was imbalance of data, and in these predictions, it is likely that the data will always be imbalanced, so more complex models or deep learning may need to be explored further. Additionally, aside from predictive capability, feature selection could provide useful insights for clinical evaluation.

Overall, this study sought to further understand the ways that data science methodologies can transform maternal healthcare.

Acknowledgements

I would like to thank the committee members: Dr. Dana Willner (Chair), Dr. Ron Smith, Dr. Dan Cristol, and Dr. Adrian Bacong. Additionally, I would like to acknowledge the faculty with the Stanford Center for Asian Health Research and Education for their help and support. I would also like to thank Dyuthy Ramachandran, Chaeyoung Kim, and Clare Heinbaugh for their contributions. Finally, I would like to thank my friends and my family because I could not have done this without them.

Appendix

A.1: Small Versus Large Dataset Numerical Variables

Variable Name	Variable	Value for Small Dataset	Value for Large Dataset
mager	**Mother's Age**	31.15 (4.17)	31.23 (4.30)
previs_rec	**Number of Prenatal Visits**	6.95 (1.71)	6.98 (1.73)
wtgain	**Weight Gained During Pregnancy (lbs)**	29.59 (16.95)	29.65 (17.23)
dplural	**Plurality**	1.04 (0.19)	1.03 (0.19)
bmi	**BMI**	26.59 (12.80)	26.56 (12.71)
pwgt_r	**Pre-Pregnancy Weight**	140.17 (26.65)	139.83 (26.37)
dwgt_r	**Delivery Weight**	167.68 (26.92)	167.19 (27.06)
priorlive	**Number of Live Children from Previous Live Births**	0.64 (0.77)	0.66 (0.80)
priordead	**Number of Dead Children from Previous Live Births**	0.01 (0.01)	0.01 (0.14)
oegest_comb	**Number of Weeks of Gestation**	38.42 (2.05)	38.47 (1.99)
cig_0	**Prenatal Cigarettes**	0.02 (0.48)	0.02 (0.71)
cig_1	**First Trimester Cigarette Use**	0.01 (0.27)	0.01 (0.55)
cig_2	**Second Trimester Cigarette Use**	0.01 (0.18)	0.01 (0.47)
cig_3	**Third Trimester Cigarette Use**	0.01 (0.21)	0.01 (0.43)

A.2: Small Versus Large Dataset Dichotomous Variables

Variable Name	Variable	Value for Small Dataset	Value for Large Dataset
sex	**Sex of Infant (% Female)**	0.49	0.48
rf_pdiab	**Pre-pregnancy Diabetes (% yes)**	0.01	0.01
rf_gdiab	**Gestational Diabetes (% yes)**	0.14	0.14
rf_ehype	**Hypertension Eclampsia (% yes)**	0	0
rf_ppb	**Previous Preterm Birth (% yes)**	0.02	0.02
mm_mtr	**Maternal Transfusion (% yes)**	0	0
mm_plac	**Perineal Laceration (% yes)**	0.02	0.02
mm_rupt	**Ruptured Uterus (% yes)**	0	0
mm_uhyst	**Unplanned Hysterectomy (% yes)**	0	0
rf_phype	**Pre-pregnancy Hypertension (% yes)**	0.01	0.01
rf_ghype	**Gestational Hypertension (% yes)**	0.04	0.04
mm_aicu	**Admit to Intensive Care (% yes)**	0	0
mbstate_rec	**Mother's Nativity (% Born In US)**	0.1	0.1

A.3: Small Versus Large Dataset Categorical Variables

Variable Name	Variable	Value for Small Dataset	Value for Large Dataset
meduc	**Mother's Education**		
	8th grade or less	0.01	0.01
	9th through 12th grade with no diploma	0.03	0.03
	High school graduate or GED completed	0.07	0.08
	Some college credit, but not a degree	0.05	0.05
	Associate degree (AA, AS)	0.04	0.03
	Bachelor's degree (BA, AB, BS)	0.35	0.36
	Master's degree (MA, MS, MEng, MEd, MSW, MBA)	0.36	0.35
	Doctorate (PhD, EdD) or Professional Degree (MD, DDS, DVM, LLB, JD)	0.09	0.09
lbo_rec	**Live Birth Order**		
	1	0.49	0.48
	2	0.41	0.41
	3	0.09	0.08
	4+	0.02	0.03
tpo_rec	**Total Pregnancy Order**		
	1	0.4	0.4
	2	0.36	0.36
	3	0.15	0.14
	4+	0.1	0.1
me_rout	**Final Route/Delivery**		

	Spontaneous	0.54	0.55
	Forceps	0.01	0.01
	Vacuum	0.05	0.05
	Cesarean	0.4	0.39

Bibliography

3. Under-sampling—Version 0.11.0.dev0. (n.d.). Retrieved May 3, 2023, from https://imbalanced-learn.org/dev/under_sampling.html#controlled-under-sampling

8. Common pitfalls and recommended practices—Version 0.10.1. (n.d.). Retrieved May 3, 2023, from https://imbalanced-learn.org/stable/common_pitfalls.html

Adegoke, T. (2021). *Inequities in Adverse Maternal and Perinatal Outcomes: The Effect of Maternal Race and Nativity | SpringerLink*. https://link.springer.com/article/10.1007/s10995-021-03225-0

BRYANT, A. S., WORJOLOH, A., CAUGHEY, A. B., & WASHINGTON, A. E. (2010). Racial/Ethnic Disparities in Obstetrical Outcomes and Care: Prevalence and Determinants. *American Journal of Obstetrics and Gynecology*, *202*(4), 335–343. https://doi.org/10.1016/j.ajog.2009.10.864

Budiman, A., & Ruiz, N. G. (n.d.). Asian Americans are the fastest-growing racial or ethnic group in the U.S. *Pew Research Center*. Retrieved May 2, 2023, from https://www.pewresearch.org/short-reads/2021/04/09/asian-americans-are-the-fastest-growing-racial-or-ethnic-group-in-the-u-s/

Chen, C. (2004). *Using Random Forest to Learn Imbalanced Data*.

Chen, L. (2019). Influence of Acculturation on Risk for Gestational Diabetes Among Asian Women.

Preventing Chronic Disease, *16*. https://doi.org/10.5888/pcd16.190212

Feature importances with a forest of trees. (n.d.). Scikit-Learn. Retrieved May 3, 2023, from https://scikit-learn/stable/auto_examples/ensemble/plot_forest_importances.html

Feki, R. (2022, January 3). Imbalanced data: Best practices. *Medium*. https://rihab-feki.medium.com/imbalanced-data-best-practices-f3b6d0999f38

Frick, A. P. (2021). Advanced maternal age and adverse pregnancy outcomes. *Best Practice & Research. Clinical Obstetrics & Gynaecology*, *70*, 92–100. https://doi.org/10.1016/j.bpobgyn.2020.07.005

Garovic, V. (2021). *Hypertension in Pregnancy: Diagnosis, Blood Pressure Goals, and Pharmacotherapy: A Scientific Statement From the American Heart Association | Hypertension*. https://www.ahajournals.org/doi/10.1161/HYP.0000000000000208

Geron, A. (2019). *Hands-On Machine Learning with Scikit-Learn, Keras, and TensorFlow: Concepts, Tools, and Techniques to Build Intelligent Systems* (2nd ed.). O'Reilly Media, Inc.

Ghosh, G., Grewal, J., Männistö, T., Mendola, P., Chen, Z., Xie, Y., & Laughon, S. K. (2014). Racial/ethnic differences in pregnancy-related hypertensive disease in nulliparous women. *Ethnicity & Disease*, *24*(3), 283–289.

Hancock, J. T., & Khoshgoftaar, T. M. (2020). Survey on categorical data for neural networks. *Journal of Big Data*, *7*(1), 28. https://doi.org/10.1186/s40537-020-00305-w

Hicks, S. A., Strümke, I., Thambawita, V., Hammou, M., Riegler, M. A., Halvorsen, P., & Parasa, S. (2022). On evaluation metrics for medical applications of artificial intelligence. *Scientific Reports*, *12*, 5979. https://doi.org/10.1038/s41598-022-09954-8

Hoyert. (2022, December 1). *The U.S. Maternal Mortality Crisis Continues to Worsen: An International Comparison*. https://doi.org/10.26099/8vem-fc65

Hughes, M. M., Black, R. E., & Katz, J. (2017). 2500-g Low Birth Weight Cutoff: History and Implications for Future Research and Policy. *Maternal and Child Health Journal*, *21*(2), 283–289. https://doi.org/10.1007/s10995-016-2131-9

Hulsen, T., Jamuar, S. S., Moody, A. R., Karnes, J. H., Varga, O., Hedensted, S., Spreafico, R., Hafler, D.

A., & McKinney, E. F. (2019). From Big Data to Precision Medicine. *Frontiers in Medicine*, *6*, 34. https://doi.org/10.3389/fmed.2019.00034

Jin, K., Gullick, J., Neubeck, L., Koo, F., & Ding, D. (2017). Acculturation is associated with higher prevalence of cardiovascular disease risk-factors among Chinese immigrants in Australia: Evidence from a large population-based cohort. *European Journal of Preventive Cardiology*, *24*(18), 2000–2008. https://doi.org/10.1177/2047487317736828

Liu, L., Wu, X., Li, S., Li, Y., Tan, S., & Bai, Y. (2022). Solving the class imbalance problem using ensemble algorithm: Application of screening for aortic dissection. *BMC Medical Informatics and Decision Making*, *22*(1), 82. https://doi.org/10.1186/s12911-022-01821-w

Liu, Y., Elliott, A., Strelnick, H., Aguilar-Gaxiola, S., & Cottler, L. B. (2019). Asian Americans are less willing than other racial groups to participate in health research. *Journal of Clinical and Translational Science*, *3*(2–3), 90–96. https://doi.org/10.1017/cts.2019.372

Margetts, B. M., Yusof, S. M., Dallal, Z. A., & Jackson, A. A. (2002). Persistence of lower birth weight in second generation South Asian babies born in the United Kingdom. *Journal of Epidemiology & Community Health*, *56*(9), 684–687. https://doi.org/10.1136/jech.56.9.684

Natekin, A., & Knoll, A. (2013). Gradient boosting machines, a tutorial. *Frontiers in Neurorobotics*, *7*, 21. https://doi.org/10.3389/fnbot.2013.00021

O'Brien, M. J., Alos, V. A., Davey, A., Bueno, A., & Whitaker, R. C. (2014). Acculturation and the prevalence of diabetes in US Latino Adults, National Health and Nutrition Examination Survey 2007-2010. *Preventing Chronic Disease*, *11*, E176. https://doi.org/10.5888/pcd11.140142

Qin, C., & Gould, J. B. (2010). Maternal Nativity Status and Birth Outcomes in Asian Immigrants. *Journal of Immigrant and Minority Health*, *12*(5), 798–805. https://doi.org/10.1007/s10903-008-9215-6

Rahnemaei, F. A., Abdi, F., Pakzad, R., Sharami, S. H., Mokhtari, F., & Kazemian, E. (2022). Association of body composition in early pregnancy with gestational diabetes mellitus: A meta-analysis. *PLoS ONE*, *17*(8), e0271068. https://doi.org/10.1371/journal.pone.0271068

Ranganathan, P., Pramesh, C. S., & Aggarwal, R. (2017). Common pitfalls in statistical analysis: Logistic regression. *Perspectives in Clinical Research*, *8*(3), 148–151. https://doi.org/10.4103/picr.PICR_87_17

Rani, P. R., & Begum, J. (2016). Screening and Diagnosis of Gestational Diabetes Mellitus, Where Do We Stand. *Journal of Clinical and Diagnostic Research : JCDR*, *10*(4), QE01–QE04. https://doi.org/10.7860/JCDR/2016/17588.7689

Samala, R. (2020). *Hazards of data leakage in machine learning: A study on classification of breast cancer using deep neural networks*. https://www.spiedigitallibrary.org/conference-proceedings-of-spie/11314/1131416/Hazards-of-data-leakage-in-machine-learning--a-study/10.1117/12.2549313.full?SSO=1

Santos, M. S., Soares, J. P., Abreu, P. H., Araujo, H., & Santos, J. (2018). Cross-Validation for Imbalanced Datasets: Avoiding Overoptimistic and Overfitting Approaches [Research Frontier]. *IEEE Computational Intelligence Magazine*, *13*(4), 59–76. https://doi.org/10.1109/MCI.2018.2866730

Singh, G. K. (2018). *Racial/Ethnic, Nativity, and Sociodemographic Disparities in Maternal Hypertension in the United States, 2014-2015—PMC*. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5985132/

Singh, G. K. (2021). Trends and Social Inequalities in Maternal Mortality in the United States, 1969-2018. *International Journal of Maternal and Child Health and AIDS*, *10*(1), 29–42. https://doi.org/10.21106/ijma.444

Sklearn.ensemble.RandomForestClassifier. (n.d.). Scikit-Learn. Retrieved May 3, 2023, from https://scikit-learn/stable/modules/generated/sklearn.ensemble.RandomForestClassifier.html

Suman, V., & Luther, E. E. (2023). Preterm Labor. In *StatPearls*. StatPearls Publishing. http://www.ncbi.nlm.nih.gov/books/NBK536939/

Yagis, E., Atnafu, S. W., García Seco de Herrera, A., Marzi, C., Scheda, R., Giannelli, M., Tessa, C., Citi, L., & Diciotti, S. (2021). Effect of data leakage in brain MRI classification using 2D convolutional neural networks. *Scientific Reports*, *11*(1), Article 1. https://doi.org/10.1038/s41598-021-01681-w

Yang, L., Wang, H., Yang, L., Zhao, M., Guo, Y., Bovet, P., & Xi, B. (2022). Maternal cigarette smoking before or during pregnancy increases the risk of birth congenital anomalies: A population-based retrospective cohort study of 12 million mother-infant pairs. *BMC Medicine*, *20*, 4. https://doi.org/10.1186/s12916-021-02196-x

Yom, S., & Lor, M. (2022). Advancing Health Disparities Research: The Need to Include Asian American Subgroup Populations. *Journal of Racial and Ethnic Health Disparities*, *9*(6), 2248–2282. https://doi.org/10.1007/s40615-021-01164-8

Zeng, N., Erwin, E., Wen, W., Corsi, D. J., Wen, S. W., & Guo, Y. (2021). Comparison of adverse perinatal outcomes between Asians and Caucasians: A population-based retrospective cohort study in Ontario. *BMC Pregnancy and Childbirth*, *21*(1), 9. https://doi.org/10.1186/s12884-020-03467-w

Zhang, J., & Mani, I. (2003). *KNN Approach to Unbalanced Data Distributions: A Case Study Involving Information Extraction | BibSonomy*. https://www.bibsonomy.org/bibtex/2cf4d2ac8bdac874b3d4841b4645a5a90/diana