How well can we predict perinatal mortality and its associated risk factors using machine learning models - A case study on Malawi?

Student ID: 20325050

BUSI4374 UNUK

2022

Data Driven Dissertation

MSc Business Analytics

University of Nottingham

University of Nottingham

UK | CHINA | MALAYSIA

Table of Contents

Tables

1. Abstract

The danger of death is greatest during the newborn phase of life. Determining which babies are most vulnerable before they are born can have a significant impact in turning the fetal tides for the newcomers. Even though the situation for perinatal mortality has improved over the last two decades, Africa is still the region with highest mortality rates. This study will differentiate between pregnant women who are most likely to give birth to vulnerable babies in a poor southeastern country called Malawi. To find the early predictors of newborn mortality, sophisticated computational techniques such as predictive modelling can be handy. This study's goal is to compile, critically evaluate, and analyze perinatal mortality causes in Malawi using machine learning models. The output variable in this study is binary in nature with either normal or perinatal delivery. The data was unbalanced with more normal deliveries than perinatal ones. After cleaning and preprocessing the data, the unbalanced class issue was resolved through Synthetic Minority Oversampling Technique (SMOTE), and thereafter the models were trained. The algorithms proved to be handy in predicting perinatal mortality of the African country. Two ensemble models, Random Forest and Gradient Boost have shown higher accuracy and precision scores while predicting an adverse outcome of the pregnant women. Model based feature selection and Shapely Additive Explanations (SHAP) techniques identified most prominent risk factors affecting the mortality. The study included 56 features varying from the women's medical record to their socioeconomic condition. It is noticeable this study is already being conducted on women who had been diagnosed with iron deficiency and were being observed after treating them with iron infusion through oral and ferric carboxymaltose (FCM) methods. Given the history, medical reasons dominated the outcome of pregnancies. Respiratory rate, weight of mothers, and blood pressure levels drew clear lines in separating normal births from perinatal ones. Secondly, economic factors are also highly deciding. Those women who spent less money in hospital, whose families do not possess any land and cannot access safe drinking water and belong to daily wage earners were the most vulnerable in giving birth to unhealthy babies. Social factors such as illiteracy, early pregnancies, affiliation with certain tribes and religion also affect the birth outcome in Malawi. Those women who were given oral iron had relatively low birth casualties. Therefore, the study has shown that machine learning models are not just effective in rightly identifying mortality but also in identifying factors playing havoc for the poorly nourished women in Malawi.

2. Introduction

Perinatal mortality consists of both neonatal (child dying in the first four weeks of the birth) and stillbirth (born dead) mortality. Each year, 2.5 million neonatal deaths and 2.5 million still births happen globally. Every day, almost 7,000 infants die, making up about 50% of all fatalities among children under the age of five (WHO, 2022). In total, Neonatal deaths account for 47% of all children under the age of 5 fatalities. Therefore, perinatal mortality has emerged as an increasingly significant contributor to overall under-five mortality (also known as infant mortality) and is thus gaining attraction at a global and local level. This situation is further problematic in Africa with has the highest perinatal mortality rates in the world. Country cooperation strategy of Malawi with the World Health Organization shows that the country has a stillbirth and neonatal ratio of 23 and 19 per 1000 livebirths, respectively (WHO, 2022). Every year, almost 14600 babies are born as dead, i.e., stillbirths, whereas 13700 die in the first four weeks of the delivery (WHO, 2018) in Malawi. Given the grim situation in poor nations, there is a huge need for accurate prediction of this mortality and its associated risk factors in the developing countries.

Tackling this health issue is one of the top priorities in the world. United Nations has targeted to decrease the perinatal mortality to 12 per 1000 livebirth globally by 2030 in its Sustainable development goals (SDGs). International health organization are highly concerned about mitigating the risk of both mothers and babies. Major changes are being introduced in relation to the health of both mother and children all around the world to meet the objectives outlined in international declarations and country pledges. Governments and professional organizations must keep track of the overall effects of the changes implemented and compare them on a global scale.

Finding the causes of perinatal fatalities may help policymakers decide where to focus on in future interventions. Keeping track of obstetric history, socioeconomic conditions, physical and medical examinations are crucial for profiling every pregnant woman. Early arrangements for those with risk factors will decrease the death rates of infants. Newborns may require basic care for feeding support, infections, and respiratory complications as well as critical newborn care (drying, warming, prompt and exclusive nursing, and hygiene). More work is required to identify pregnant women who are at risk of preterm labor and encourage them to give birth in a medical setting that can provide additional care when required, such as assistance with adequate breastfeeding, ongoing skin-to-skin contact, antibiotics, and antenatal corticosteroids. Families, communities, and healthcare professionals must esteem newborns in order for them to receive the lifesaving care they require. Action throughout the spectrum of care is need of the hour if we are to reverse the trend of these avoidable fatalities.

Since perinatal mortality has high dimensional data with a plethora of maternal factors, machine learning technique is suitable to predict pregnancy outcome. The study will not just be the first to highlight risk factors In Malawi but also to use sophisticated machine learning techniques in order to reach that prediction. Past studies have shown that traditional logistic regression has been less effective in the similar classification of health problems (Mfateneza, et al., 2022). The predictive algorithms used in this study have potential to differentiate future pregnancies based on the previous gestational experiences using diverse predictors. As a matter of the fact, these algorithms do not have form assumptions as opposed to other techniques, which then makes them superior to traditional modelling techniques. Predictive analysis also ranks feature in terms of their impact on the output variable. Feature importance and SHAP values rank input factors in order of their importance in differentiating outcomes. They will help us comprehend the situation as it stands, raise awareness of perinatal mortality in Malawi, and facilitate discussion and action among stakeholders. In order to discover and execute solutions for improved outcomes, data can be used to pinpoint the most crucial risk factors to address the gaps in care. Every nation's ability to build human capital and advance economically depends on giving children a safe and healthy start in life, thus providing for their needs is both a short-term and long-term investment. We can stop babies from being born too early and too small and make sure that small babies receive the vital life-saving care and nurturing they require by working together with government leaders, civil society organizations, health professionals, families, communities, and other partners to enact change. It is therefore becoming more and more clear that data backed insights are instrumental.

3. Literature Review

The study looked into different definitions of perinatal mortality and choose one used by the World Health organizations. Features from past studies are also studies. In order to understand cultural aspects, the study looked closely at the similar studies in neighboring countries like Tanzania. Countries from other regions with similar socioeconomic profiles were compared to get established risk factors. Neonatal and perinatal mortality causes were studied separately to understand the aggregation for perinatal rates. Past studies were also studied to rank features in terms of their impact are also considered to come up with important features in Malawi. Lastly, a review of traditional regression modelling has been compared to machine learning techniques for the prediction of perinatal mortality in the the same settings.

Perinatal mortality does not have a common definition. The World Health Organization states perinatal mortality as the aggregate of number of stillbirths and neonatal deaths per 1,000 live births. The American Academy of Pediatrics (AAP) Committee on Fetus and Newborn developed standard terminology for fetal, infant, and perinatal deaths based on the World Health Organization (WHO) standards and the National Center for Health Statistics (NCHS) of the Centers for Disease Control and Prevention (CDC) guidelines (Zacharias, 2020). The NCHS further classifies fetal deaths into "early" (20 to 27 weeks gestation) and "late" (28 weeks gestation) categories for statistical purposes. According to the WHO, a stillbirth occurs at or after 28 weeks of pregnancy. Miscarriages are pregnancies that end in fetal death before 20 weeks (Zacharias, 2020). WHO defines neonatal morality as a newborn dying in the first 28 completed days of life per 1000 livebirths in a given period. Neonatal deaths are further classified between late neonatal deaths, which occur after the 7th day but before the 28th completed day of life, and early neonatal deaths, which happen during the first 7 days of life (WHO). This study considers the WHO definition of perinatal mortality.

There have been country profiling studies undertaken by some international development agencies, which provide guidance ~~and they guide~~ in identifying some of the established risk factors associated with this mortality. Every Preemie Scale, a United States Agency for International Development (USAID) led project on child and mother health, works for the betterment of 25 developing countries in Africa and Asia. The project presents countrywide profile. A country profile of Malawi presents some grim facts (USAID, 2015). The report mainly focuses on preterm babies (those who are born before 37 weeks of gestation) and those with low birth weights. Prematurity is the main cause of disability and poor health later in life for children under the age of five all over the world. South Asia and sub-Saharan Africa together account for more than 60% of preterm births worldwide. More than a million of the fifteen million premature babies who are born each year pass away from problems associated to preterm delivery. Every Preemie Scale presents following facts about the preterm babies in Malawi: “120,000 babies are born too soon each year and 4,800 children under five die due to direct preterm complications” (USAID, 2015). Preterm birth rate in Malawi is reported to be 18 percent and low birth rate is estimated to be 13 percent (USAID, 2015). United Nation’s Children Fund (UNICEF) reported in its annual report that “Approximately 38 babies will die each day before reaching their first month; 40 stillbirths occur every day” (UNICEF, 2015). The same UNICEF report calculated that the leading causes of neonatal morality in Malawi to be prematurity (33 percent), birth asphyxia and birth trauma (25 percent), Sepsis (18.6 percent) and congenital anomalies (10 percent).

Stillbirths, a significant proportion of total perinatal mortality, garners less attention than neonatal deaths. They are not included in the UN's Global Burden of Disease measures, the UN's Millennium Development Goals, or the growing attention and funding given to maternal, neonatal, and child health (Lawn, et al., 2011)

The first step in addressing the issue of prenatal mortality is to identify and categorize the major causes of death in all circumstances. Past studies have warned about home-based deliveries and the absence of real data. A standardized classification system used by the World Health Organization (WHO), ICD-PM, is used to classify deaths that occur during the perinatal era (Allanson & Tunçalp, 2016). Almost all, precisely 98 percent, stillbirths occur in low- and middle-income nations with rates ranging from 0 per 1,000 live births in Finland to more than 40 per 1,000 live births in Nigeria and Pakistan (Lawn, et al., 2011). Around the world, rural families account for 67 percent of stillbirths, with rural sub-Saharan Africa and south Asia accounting for 55 percent. In these regions, caesarean sections and skilled birth attendance are significantly less common than in urban areas. Annually, there are an estimated 1.19 million intrapartum stillbirths. Antenatal stillbirths are linked to maternal illnesses and fetal growth limitation, while the majority of intrapartum stillbirths are related to obstetric crises. (Lawn, et al., 2011). Additionally, studies in Ghana have indicated that visits by medical personnel to homes with newly born babies might lower newborn morality by as much as 12% (Pitt, et al., 2015). Assessing child mortality from both sides of maternal health and child together helps in creating preventive measures (WHO, 2016).

Neonatal mortality causes are even wider than stillbirth reasons. A detailed report of WHO predicts that Poor mother health, insufficient prenatal care, wrong handling of difficulties during pregnancy and delivery, poor hygiene during delivery and the first crucial hours after birth, and a lack of newborn care are all factors that contribute to neonatal deaths and stillbirths (WHO, 2006). The report further indicates that Inappropriate cord care, letting the baby stay wet and cold, discarding colostrum, and feeding other foods are harmful practices that are deeply ingrained in the cultural fabric of societies. These factors, as well as a women's status in society, their nutritional status at the time of conception, early childbearing, too many closely spaced pregnancies, and ~~a~~ like factors are the primary reasons for these deaths (WHO, 2006)

Another important factor in perinatal mortality is low birth weight (LBW) in newborn babies. If the weight is less than 2500 grams at the time of birth, the baby is vulnerable to diseases and the chances of survival are very low. A study on Risk factors and adverse perinatal outcome associated with low birth weight on Northern Tanzania found the following factors: poor nutrition of mothers during and after pregnancy, anemia, chronic blood pressure, renal and heart diseases, alcohol and smoking, drug use during gestation, illiteracy, mother's occupation, short height, old maternal age, induced labor or elective caesarian section, and physical or emotional abuse~~s~~ **(Mitao & Philemon, 2016)**. Long-term effects of low birth weight include neurological sequelae, stroke, hypertension, type 2 diabetes, and recurring low birth in successive siblings. This study concluded that identifying low weight pregnancy outcome will help in making a preemptive policy to resolve high perinatal mortality issue in developing countries such as Malawi. Thus, mother's health factors should be taken care of in advance to lessen the risk of fetal and neonatal mortality. This dissertation findings will ultimately inform program implementers, policymakers and researchers addressing LBW-related mortality and morbidity.

According to data from African nations, newborn sex has also been linked to neonatal death **(Akalewold, et al., 2021)**. Unfortunately, there hasn't been much research done in ~~the~~ literature on gender disparities in birth weight. Thus, addressing health inequality requires a critical review of society. Documenting and adopting adequate measures will help in tracking progress on the issue **(Akalewold, et al., 2021)**. To guarantee that no one is left behind, SDG 17.18 advises making steps to expand the availability of data broken down by income, gender, age, race, ethnicity, migratory status, disability, and geographic location in developing countries.

The issue at hand has a lot of variables. Selecting a scientific way to rank these factors will be a distinct goal of this thesis. Previous studies have conducted feature importance using different machine learning techniques. Top five important features affecting neonatal mortality were found to be birthweight, gestational age, multiple births, prenatal steroids, maternal NTH/PIH and blood pressure by the study published in the Neonatology journal in the USA **(Mangold, et al.,**

2021). Another study published in the leading medical journal, BMC, find the following features to be highly important for infant mortality in Rwanda: Martial status, children ever born, birth order, wealth index, preceding birth order, maternal ~~education~~education, and source of drinking water **(Mfateneza, et al., 2022)**. JAMA, a leading peer-reviewed medical journal, published the paper on predictive modelling for prenatal mortality and found out that the main predictor of newborn death was birth weight **(Vivek V. Shukla, et al., 2020)**. Therefore, potential risk factors (independent variables) in this study for stillbirths and neonatal mortality are considered keeping in mind previous literature and relevance to outcomes. For instance, gestational age and prematurity were discovered to increase the risk of neonatal mortality in a study published by JAMA global **(Vivek V. Shukla, et al., 2020)**. This study also pointed that no prenatal care is found to be associated with all-cause stillbirths. Other variables identified using different datasets included prior stillbirth, parity of five or more, preterm and poverty **(Lawn, et al., 2011)**.

Even though studies to forecast child mortality have been conducted using traditional statistical methods, it has been demonstrated by several papers that machine learning approaches outperform the conventional regression techniques that were employed in similar settings (Mangold, et al., 2021). One more study employing comparable medical data and applying machine learning algorithms to categorization medical problems supported more precise decision-making (Raita, et al., 2019). They are thought to do better than others at differentiating predictions. The claim of machine learning algorithms for research like this is strengthened by a paper published in the leading publisher Biomedical Engineering: " ML and CI techniques are shown to improve the assessment using both indicators already in place and other clinical variables that are routinely measured" (Velido & Ribas, 2018).

Another combined study of the leading universities in USA - ~~especially~~ specifically the Department of Pediatrics of the ~~,~~ University of Texas at San Antonio - concluded that neonatal deaths can accurately be predicted using machine learning models. (Mangold, et al., 2021). This analysis shows the measures and predictors for newborn mortality that are most frequently utilized in AI prediction models. The study in America recommended that ~~the~~ future research should concentrate on external validation, calibration, and application deployment that can be easily accessed by healthcare professionals. However, there is conflicting evidence suggesting these strategies do not work well when dealing with data of babies under the age of five (Lamping, et al., 2018). Nevertheless, very few papers have also disputed the effectiveness of machine learning models on newborn infants receiving intensive care after birth (Goto, et al., 2019).

In developing nations, there has been a lot of research on neonatal mortality, but the majority of methods used up to now have relied on traditional regression analyses, which have a limited capacity for prediction. Although advanced machine learning (AML) techniques offer precise neonatal mortality prediction, no research employing ML techniques ~~has~~ have been done in Malawi. Therefore, this study has used machine learning techniques to forecast newborn mortality in Malawi. A BioMed Council (BMC) published study on infant mortality recommended using the same machine learning techniques to tackle similar issues such as survival very preterm, perinatal mortality, stunting and low birth weights in infants (Mfateneza, et al., 2022). In order to uncover historical trends that were missed by earlier studies, ML methods are used to reflect an original logical and analytical perspective.

Tackling this grim issue with high precision is important. Machine learning approach to solve the issue in low-income countries is backed by leading research journals. For instance, a study published by the Journal of American Medical Association (JAMA), a peer-reviewed medical journal, has recommended using machine learning with big data sets. The paper concluded the superior quality of supervised learning models outperform traditional models due to their capacity to define complicated relationships and discover unique interaction between variables (Vivek V. Shukla, et al., 2020). However, the similar study by Shukla claims that it is yet to test whether the same modelling will work efficiently with small data sets.

This study is a step ahead ~~than~~ in comparison to others in a way that they have used conventional statistical models. A paper published by peer-review medical journal, BMJ Open, opined that traditional regression is inferior to machine learning since it ignores many factors due to model assumptions (Mboya, et al., 2020). In contrast, non-parametric machine learning (ML) algorithms identify the most predictive combinations of parameters based on their frequency and strength of correlation. Additionally, few ML algorithms use any assumptions. This study will differ from past studies for three additional reasons. First of all, no one has determined the causes of perinatal mortality using machine learning models on Malawi data. Secondly, the majority of earlier studies concentrated on identifying factors that contribute to just infant mortality. Third, this study will add more variables recommended by similar studies (Vivek V. Shukla, et al., 2020).

Machine learning models can delve deep, assess perinatal risk factors, and help us in prioritizing pregnant women who are at high risk of having a preterm birth. The models can assimilate as varies as socioeconomic, cultural, and biological aspects to determine why most newborn deaths take place in domestic environments.

4. Research Methodology

4.1. Study Design and Participants

This study uses a secondary data of another study where infusion of iron was monitored in a group of pregnant women in Malawi's Blantyre and Zomba districts. It was done to ascertain whether treating moderate to severe maternal anemia in pregnant women with IV ferric carboxymaltose (FCM) or Oral infusion during the second trimester is safe and beneficial for improving maternal, neonatal, and infant outcomes.

Table 1: Data Collection Criteria

	Inclusion	Exclusion
1	Pregnancy of a singleton confirmed at 13–26 weeks.	Hypersensitivity to any study medication.
2	Moderate to severe anemia (Hb 10g/dl) without the need for an immediate blood transfusion.	Signs and symptoms of a bacterial infection or malaria.
3	A negative parasitemia for malaria.	Any illness or ailment that necessitates hospitalization and is serious concurrently.
4	Be a resident of the Blantyre and Zomba district study catchment region.	Chronic conditions that may be harmful to the growth and viability of the fetus.
5	Plan to give birth in a medical facility and give written consent (and, if under 18, assent.)	Clinically low hemoglobin (normally Hb 5g/dl) necessitating blood transfusion.

4.2. Conceptual Framework

Engaging pregnant women in their second trimester and tracking them up until 12 months after giving birth, this data is from Phase III of a 30-month study. This is an effectiveness experiment that was conducted in medical facilities. Since this data set consists of multiple visits from controlled group of patients, there were around 845 unique pregnant mothers who were finalized to take part in this study. Each was assigned a unique identification number, pid, after prescreening. However, almost half of the unique pregnant women had missing values in almost all major features, so a final group of 489 women is included in this study.

4.3. Output and Input Variables

Output variable in this study is a binary variable: either the pregnancy was normal or perinatal. A pregnancy has a perinatal outcome if any of these three outcomes occurred: the baby was stillborn, was born before 37 weeks of gestation, or weighed less than 2500 grams at the time of delivery. Any other form of birth was regarded as being normal or healthy. The study assumed that if either of these conditions met, the baby has no chance to survive, i.e., it is a perinatal morality – a baby who is born dead or died in the first 28 days of life.

For the ML models, we used a total of 56 predictor variables. The choice of these factors, the majority of which are accessible through documentation, was influenced by prior studies. They range from demographics, social, biological, and medical to past obstetric history. A detailed description of all of these variables is given in the following table.

Table 2: Variable Selection and Explanation

	Variable Names	Type of Variable	Variable Explanation
Identification	**pid**	**Objective**	Unique ID Number
1.Pregnancy outcome	**Results**	**Dichotomous**	1 for perinatal, 0 normal delivery
2.Fetal Biometry	**fbi_significant_findings_yes_no**	**Dichotomous**	0, No | 1, Yes Fetal biometry is measurement taken during a standard ultrasound.
	fbi_amniotic	**Ordinal**	1, Normal | 2, Moderately increased | 3, Polyhydramnios | 4, Moderately reduced | 5, Oligohydramnios | 6, Anhydramnios Amniotic is fluid surrounding a fetus within the amnion.
3.Demographics	**dem_tribe**	**Nominal**	1, Chewa | 2, Yao | 3, Tumbuka | 4, Lomwe | 5, Sena | 6, Tonga | 7,Ngonde | 8, Other Specify
	dem_formal_sch	**Dichotomous**	0, No | 1, Yes Mothers have formal schooling or not.
	dem_educ	**Ordinal**	0, None | 1, Lower Primary(1-5) | 2, Upper Primary(6-8) | 3, Lower Secondary(1-2) | 4, Upper Secondary(3-4) | 5, Tertiary Level Of Schooling
	dem_marital_status	**Nominal**	1, Single | 2, Married | 3, Widowed | 4, Divorced/Separated | 5, Others Specify Marital Status
	dem_religion	**Nominal**	0, None | 1, Christian | 2, Muslim | 3, Other Specify
	dem_income_source	**Nominal**	0, None | 1, Subsistence farming | 2, Large scale farming | 3, Employed | 4, Casual work for wages | 5, Business | 6, Other Specify
4.Piper Fatigue Scale	**pfs_fatigue_sch**	**Ordinal**	0, None | 1, 1| 2, 2 | 3, 3| 4, 4 | 5, 5| 6, 6 | 7, 7 | 8, 8| 9, 9 | 10, A great deal My fatigue interferes with my ability to do work or school activities?
	pfs_normal_abnormal	**Ordinal**	0, Normal | 1, 1| 2, 2 | 3, 3| 4, 4 | 5, 5| 6, 6 | 7, 7 | 8, 8| 9, 9 | 10, Abnormal I describe my fatigue as being? normal /abnormal
	eco_decision_make	**Dichotomous**	1, Me | 2, Someone else Who makes important decisions in your home?
	eco_decision_relation	**Nominal**	1, Sister / Brother | 2, Mother / Father |3, Mother in law / Father in law |4, Grandmother / Grandfather |5, Aunt / Uncle |6, Cousin |7, Does not want to disclose | 8, Other | 9, Husband Who takes decision in the home?
	eco_income_source	**Nominal**	1, Subsistence farming | 2, Large scale farming | 3, Employed | 4, Casual work for wages | 5, Business | 6, None | 7, Other Specify What is the main source of income of the decision maker?
	eco_decision_edu	**Ordinal**	1, None | 2, Lower Primary (1-5) |3, Upper Primary (6-8) | 4, Lower secondary (1-2) |5, Upper secondary (3-4) |6, Tertiary How far did the decision marker go with education?
	eco_breadwinner	**Nominal**	1, Me | 2, Decision Maker|3, Someone else Who is the main breadwinner?

5.Home Economics	**eco_drk_water_safe**	**Nominal**	1, Yes | 2, No |3, Don Do you do anything to the water to make it safer to drink?
	eco_toilet	**Nominal**	1, Flush toilet | 2, Pit latrine |3, Dug-out pit with roof |4, Dug-out pit without roof | 5, None |6, Does not wish to disclose |7, No facility, bush, outdoors |8, Other specify
	eco_house_roof	**Nominal**	1, Grass | 2, Iron sheets | 3, Clay | 4, Tiles | 5, Concrete | 6, Plastic Sheeting | 7, Does not wish to disclose | 8, Does not know | 9, Other, specify
	eco_transport_own_other	**Dichotomous**	0, No | 1, Yes Does anyone in the household own any other forms of transportation?
	eco_land_owner	**Nominal**	1, Own the structure | 2, Pay rent/ Lease |3, No rent, with consent of owner |4, No rent, without consent, squatting |5, Other specify ; Land Ownership status
	eco_agri_land	**Dichotomous**	0, No | 1, Yes Does any member of this household own any agricultural land?
	eco_livestocks	**Dichotomous**	0, No | 1, Yes Does this household own any livestock, herds, other farm animals, or poultry?
	eco_bank_account	**Nominal**	0, No | 1, yes | 3, I don't know Does any member of this household have a bank account?
6.Participant Random Drug	**ran_study_arm**	**Dichotomous**	1, Oral Iron | 2, FCM Which arm has the participant been randomized?
7.Obstetric history	**obh_gravidity**	**Continuous**	How many pregnancies have you had included this one (gravidity)?
	obh_parity	**Continuous**	How many of those pregnancies did you carry for more than 28 weeks(parity)
	obh_age_first_delivery	**Continuous**	At what age were you at delivery of first pregnancy?
	obh_live_delivery_num	**Continuous**	How many deliveries resulted in a live baby?
	obh_abortion_or_miscarriage	**Continuous**	Number of spontaneous abortions/miscarriages
	obhi_vag_deliveries	**Continuous**	Number of vaginal deliveries
	obhi_c_sections	**Continuous**	Number of C-sections
	obh_problem_complication	**Dichotomous**	0, No | 1, Yes | 99 Did you have any problems/complications during any of your deliveries?
	obh_smoke	**Dichotomous**	0, No | 1, Previous smoker | 2, Current smoker Do you smoke?
	obh_congenital_abnorm	**Dichotomous**	0, No | 1, Yes Is there family history of: Congenital abnormalities/genetic disease
	obh_perineal_lacerations	**Dichotomous**	0, No | 1, Yes Is there family history of: Consanguinity
	obh_hypertension	**Nominal**	1, No | 2, Previous pregnancies | 3, Current pregnancy

			Does the participant have any of these conditions: Hypertension
	obh_diabetes	**Nominal**	1, No | 2, Previous pregnancies | 3, Current pregnancy Does the participant have any of these conditions: Diabetes?
	obh_hiv_positive	**Dichotomous**	0, No | 1, Yes
	obh_anaemia	**Dichotomous**	0, No | 1, Yes | 99 Anemia in previous Pregnancies
	obh_syphillis	**Dichotomous**	0, No | 1, yes |99, Unknown/not done Syphilis (VDRL) Positive
	obh_obh_allergic_drug	**Dichotomous**	0, No | 1, Yes Have you ever experienced an allergic reaction to any drug?
8.Patient Cost Time Spent	**pct_cost_facility**	**Dichotomous**	Did you use any money at the health facility or because of the visit? 0, No | 1, Yes | 2, Don't know
	pct_total_spent	**Continuous**	How much in total did you spend at the health facility (disaggregated total)?
	pct_cost_undertable	**Continuous**	How much money did you spend on each of the following? Under-the-table payment (MK)
9.Health Quality Evaluation	**eq5_mobility**	**Ordinal**	1, I have no problems in walking about | 2, I have some problems in walking about | 3, I am confined to bed
	eq5_selfcare	**Ordinal**	1, I have no problems with self-care | 2, I have some problems washing or dressing myself | 3, I am unable to wash or dress myself
	eq5_activities	**Ordinal**	1, I have no problems with performing my usual activities | 2, I have some problems with performing my usual activities | 3, I am unable to perform my usual activities
	eq5_pain	**Ordinal**	1, I have no pain or discomfort | 2, I have moderate pain or discomfort | 3, I have extreme pain or discomfort
	eq5_anxiety	**Ordinal**	0, Worst imaginable health state | 1, I am not anxious or depressed | 2, I am moderately anxious or depressed | 3, I am extremely anxious or depressed
	eq5_health_state_today	**Ordinal**	indicate on this scale how good or bad your own health is today? best state you can imagine is marked 100 and the worst state you can imagine is marked 0
10.Maternal Physical Examination	**mpe_systolic**	**Continuous**	Systolic Sitting position Blood Pressure - mmHg
	mpe_diastolic	**Continuous**	Diastolic Sitting position Blood Pressure - mmHg
	mpe_pulse_rate	**Continuous**	Pulse Rate - beats/minute
	mpe_respiratory_rate	**Continuous**	Respiratory rate - cycles/minute
	mpe_axillary	**Continuous**	Axillary Temperature

4.4. Data Collection Strategy

The risk factors are divided into multiple visits from the start of the participation and confirmation of pregnancy to later delivery. The data used in this research consists of 7 visits. First visit leads to enrollment after screening of the women. Second visit happened in week 4 and it was hospital based. Third, Fourth and Sixth visits are home based which occurred in week 34, 36 and 40 of gestation, respectively. Fourth visit happens in week 36 of conception and lab and other fatigue examinations records are collected in this week. Finally, seventh visit is when delivery was done in hospital. Nevertheless, those babies who are born before week 28 are included in the results in the form of adverse outcome. The study evaluated mortality outcomes based on premature birth (before 37 weeks), low birth weight (<2500g) and stillbirth.

4.5. Data Preprocessing

Raw data is never complete; hence it cannot be passed through a model. To get the greatest results from what is done using ML algorithms on data, data preparation is necessary. Typical data preprocessing involved a careful labelling of variables, imputing missing values, dealing outliers, removing multicollinearity, balancing the unbalanced classes, scaling the data and handling low cardinality issue (using one hot encoder) for ordinal or nominal variables.

Input and output variables form above table shows that the final dataset has four different data types. They are dichotomous(binary), nominal (more than 2 options), ordinal(ranking) and continuous. Where there were null values, characteristics of variables were checked. For instance, mode was applied to fill the null values in categorical variables. Median and mean were applied for continuous variables. Moreover, standardization was applied column wise on continuous data to improve performance on models such as logistic regression. The ideal strategy for low cardinality variables is typically to convert the feature into one column per distinct value, with a 0 where the value is absent and a 1 where it is present. They are called dummy variables. Nominal and ordinal variables are typically the best candidates for this strategy. If we don't do this initially because they have no inherent order, the machine learning algorithm can mistakenly search for a relationship in the order of these values.

The data used for this study has unbalanced classes issue: there are more adverse cases than the normal cases in our output class. SMOTE is a method that evaluates data points in feature space by their closest neighbors as determined by Euclidean Distance (Chawla, et al., 2022). In order to construct synthetic data for both categorical and quantitative features in the data set, we utilized Synthetic Minority Over-sampling Technique for Nominal and Continuous Features (SMOTE-NC) from the imbalanced-learn toolkit. SMOTE-NC performs a specific action for the categorical features, which modifies the generation of a new sample in a small way. The most common category among the closest neighbors present throughout the generation is actually chosen to determine the categories of a newly created sample (Aguliar, 2019). Even though it is convenient way to harmonize classes, its efficacy is still debatable. A study by Eötvös Loránd University Hungry shows that all oversampling techniques produce minority samples that are most likely to be majority (AHMAD, et al., 2022). Despite having doubts, this study has used SMOTE-NC to augment the minority class by 60 percent in the training dataset.

When creating a predictive model, the process of feature selection involves lowering the number of variables. An important procedure called Principal Components Analysis (PCA) is also used to converts columns into a new set of features known as Principal Components. This effectively compresses a significant portion of the data from the entire dataset into fewer feature columns. It also assists the models process data more quickly by reducing the linear dimensionality of the data. To discover the ideal range of n estimators that best capture the data's variance, study has first depicted the PCA spectrum.

4.6. Feature Selection and Model Design

The objective of this study is to take predictive approach concerning perinatal mortality that incorporates ML algorithms. Conventional techniques like Ordinary Least Square or Logistic Regression are supplemented by these methodologies. Many machine learning techniques do not call for explicit model formulation before estimating or the distributional presumptions of more conventional techniques (Trent, et al., 2018). Machine learning techniques are often adaptable, nonparametric approaches to predicting or classifying data. These techniques can allow for a greater number of predictors, known as high dimensional data, and are often explained by the algorithm that explains how the predictions are created using the raw data.

Feature selection techniques were applied during the pre-processing of the data as well. For a layman, ranking features in terms of higher effect on delivery outcome makes more senses. This is where feature importance techniques play an important role in communicating technical factors in a ranked manner. First method used for this purpose is Filter based method. It's also called univariate variable importance technique where features can be ranked by measures of how they individually affect the output variable. Univariate feature ranking and selection (all measures) are all implemented in sklearn via the GenericUnivariateSelect object. Univariate importance seeks to quantify the degree to which each characteristic influences the target variable while ignoring all other factors. A prediction model or a closed formula, such as Pearson correlation, can be used to determine the score. Since humans typically think in terms of one variable, it is simple to understand. Additionally, it is unaffected by correlated features. However, it oversimplifies reality too much as it ignores the interactions between the features.

Model based feature ranking is another procedure of feature importance. One can rank features based on how frequently they are used in a trained model, as opposed to utilizing external metrics of how each feature affects the output feature. Random Forests is one example. Random Forests in Sklearn contain an element called feature importances_ that keeps the scores for each feature as a list indicating the importance of each feature. Random forest should be used carefully because in contrast to predictions on a test dataset, the feature importance from random forest is computed using only the training data provided to the model. Therefore, we require a substitute method that can determine the relative importance of each feature on a test dataset and do so based on many metrics, including accuracy, precision, and recall.

The permutation feature importance technique gets around the problems we discussed previously with a fairly straightforward idea: how does the model do if only the values of a feature are randomly shuffled? Therefore, we will also consider permutation importance. Permutation_importnace function from sklearn.inspetion will be used for implementing this technique.

It can be challenging to interpret machine learning models with above feature importance techniques. We need more algorithms if we are to comprehend what the key characteristics are that influence the model's output. The SHAP (Shapely Additive Explanation) method, which explains how each feature influences the model and enables local and global analysis for the dataset and issue at hand, is another technique. Shap values are based on the cooperative game theory, and it makes the machine learning models more transparent and understandable. Nevertheless, it is to be understood that SHAP values will reveal the feature which impact the most, but it does not tell the quality of prediction (Vinícius, 2022)

Six machine learning models -Logistic Regression, KNeighbors Classifier, Decision Trees, Random Forest, Gradient Boost, and Support Vector Classifier- were considered for the classification of pregnancy outcome. These six models are chosen while taking into account their use in similar studies. Each of these models have proven to be efficient in the medical studies. Besides, they all have the ability to improve after tweaking their parameters. When analyzing numerous explanatory variables at once and minimizing the effect of confounding variables, logistic regression is an effective model. Before fitting LR, we eliminate variables with high correlation to prevent multicollinearity issue. The ensemble techniques, known as Random Forest and Gradient Boost, enhances generalization by fusing various learning models. An ensemble approach is justified by the idea that, due to significant volatility, a pool of simple models may perform better than overfitted models (Do, et al., 2022). In order to create an ensemble output, RF constructs numerous decision trees that have been trained on training samples and combines the predictions. Gradient boosting builds weaker (simpler) prediction models in a sequential manner, with each model attempting to anticipate the error left by the previous model. A weak learner is a model that performs only marginally better than random predictions. SVM is a machine learning (ML) model with a linear delimiter that works well for binary classification. Using what are known as "kernels tricks," we expand SVM to handle non-linear issues. It uses a nonlinear map that is derived using a kernel function to implicitly transfer the input vectors into a high-dimensional feature space. K Nearest Neighbors uses a tree-like data structure instead of brute force to calculate the separations between points in the training set and the point of interest. The best approach in K-Neighbor is chosen based on the data's sparsity, the number of required neighbours, and the dimension and number of features.

The dataset is first split with simple train and test split ratio of 80:20. Then, continuous variables in training portion are standardized before they are augmented synthetically by SMOTE-NC. Test dataset is separated in the beginning form the original dataset and its left uncontaminated for later evaluations. Later, KFkold cross validation is done to crosscheck model performance on multiple random training sets. Lastly, hyperparameter tuning is employed on the three best models. Tuning the hyperparameters is essential for improving model performance. Effective hyperparameter tweaking may significantly boost performance when dealing with a complicated model with numerous hyperparameters. In terms of training time, infrastructure resource requirements (and consequent cost), model convergence, and model correctness, they can have a significant impact on model training. These models are evaluated based on area under the curve (ROC_AUC), accuracy, precision, recall and F1 scores. Since predicting perinatal outcome is more important than predicting normal delivery, roc_auc and precision scores are given more weight than the simple accuracy scores. Precision and F1 scores are also considered closely for better assessment. Python’s version of 3.8 is used to do all statistical modelling and visualization.

5.Results

Even though there were fewer less than a thousand pregnant women in controlled group study at the time of data collection, only 489 were selected for final modelling. Those who were dropped have either a lot of missing features or their outcome of birth was not traceable. As it can be seen in figure one, the selected women had following distribution. Almost 80 percent had normal deliveries and 20 percent had perinatal outcomes. As a result, we will use data augmentation techniques to increase the target occurrence using synthetic data (Chawla, et al., 2022).

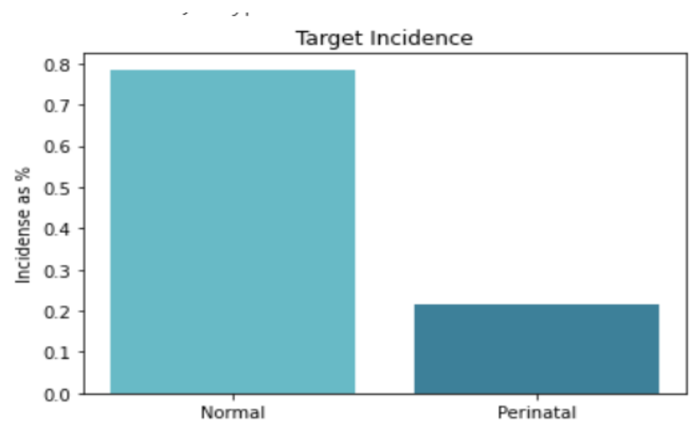


Figure 1: Distribution of Outcome Variable

Table 3: Summary of Continuous Variables in average form		**Pregnancy Outcome**	
Examination Type	**Featured Name and Explanation**	**Normal (n=384)**	**Perinatal (n=105)**
Obstetric history	**Live delivery Num:** How many deliveries resulted in a live baby?	1.975904	1.652174
	Abortion or Miscarriage: Number of spontaneous abortions/miscarriages)	0.132530	0.217391
	Vaginal deliveries: Number of vaginal deliveries)	1.975904	1.478261
	C sections: Number of C-sections	0.060241	0.217391
Patient Cost at Health Facilities	**Total spent:** How much in total did you spend at the health facility (disaggregated total)	395.061224	335.000000
Maternal physical Examination	**Systolic:** Systolic Sitting position Blood Pressure - mmHg	106.552083	105.228571
	Diastolic: Sitting position Blood Pressure – mmHg)	70.223958	68.885714
	Pulse rate: Pulse Rate - beats/minute	91.997396	92.923810
	Respiratory rate: Respiratory rate - cycles/minute	19.552083	20.180952
	Axillary: Axillary Temperature	36.206771	36.182857
	Weight: Avg weight Measurement (kg)	60.819427	59.008667

Table 3 shows average of continuous variables. The data shows significant difference between both groups. Obstetric history reveals that the mothers with perinatal outcome had on average experienced more abortion or miscarriage in their previous pregnancies. Those who had caesarian sections before are four times more likely to give birth to a vulnerable baby. Unfortunately, those patients who had spent more money have a greater chance of a normal birth. Lab examinations show that the ladies with slightly low blood pressure are more prone to risk. Those who weigh less than 60kgs have more risk than those above 60 kgs. Women with high respiratory rate (breathing issues) gave birth to ill babies.

5.1. Multicollinearity

Since our dataset has both numerical and categorical features, their distribution was observed carefully. The fowling graph shows the correlation between all variables.

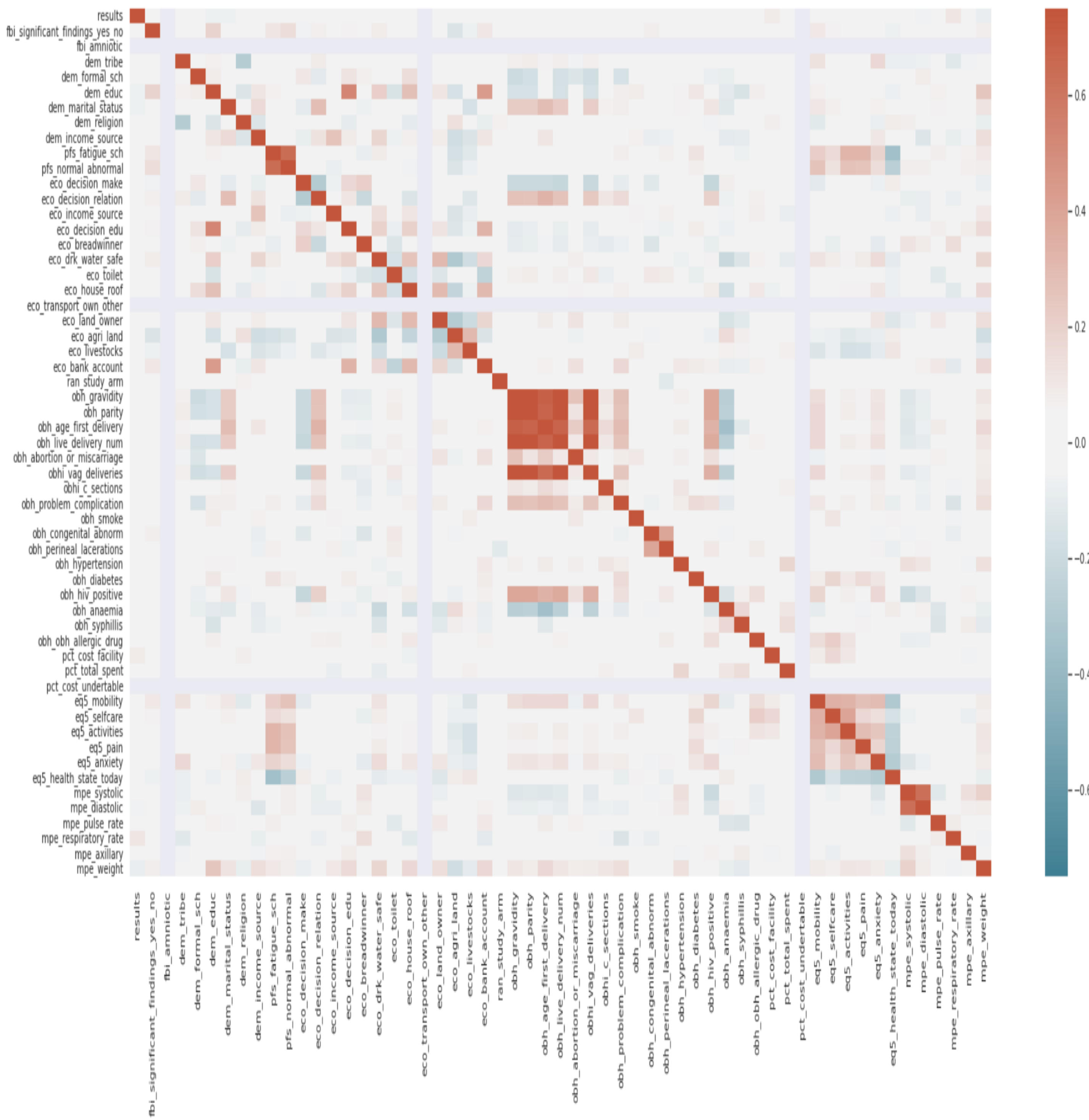


Figure 2: Multicollinearity in all variables

Figure 3: Multicollinearity in continuous variables

Figure 2 shows the correlation matrix of all variables. There is strong correlation between some of the continuous variables as extracted in figure 3. Gravidity (total number of pregnancies including current one) and parity (number of pregnancies with more than 28 weeks of gestation) are highly correlated as expected. Previous live delivery number is also correlated with normal (Vaginal) deliveries. These variables are dropped to solve multicollinearity issue. Following figure 4 shows the correlation matrix after dropping these variables.

Figure 4: Correlation Matric after removing highly corelated variables

SMOTE-NC was applied to the training dataset, and the target incidence was increased from 21 percent to 37.5 percent. The sampling strategy hyperparameter allows the generator of oversampling to be selected. This parameter was adjusted to 0.6, which resulted in a 60% increase in the minority sample in the training set. The new adverse to normal delivery class ratio in training set was 38 to 62. Test sample was not synthetically augmented to prevent contamination of data. K neighbors is another hyperparameter that was kept at default, i.e., k=5, for synthetic resampling.

5.2. Feature Selection

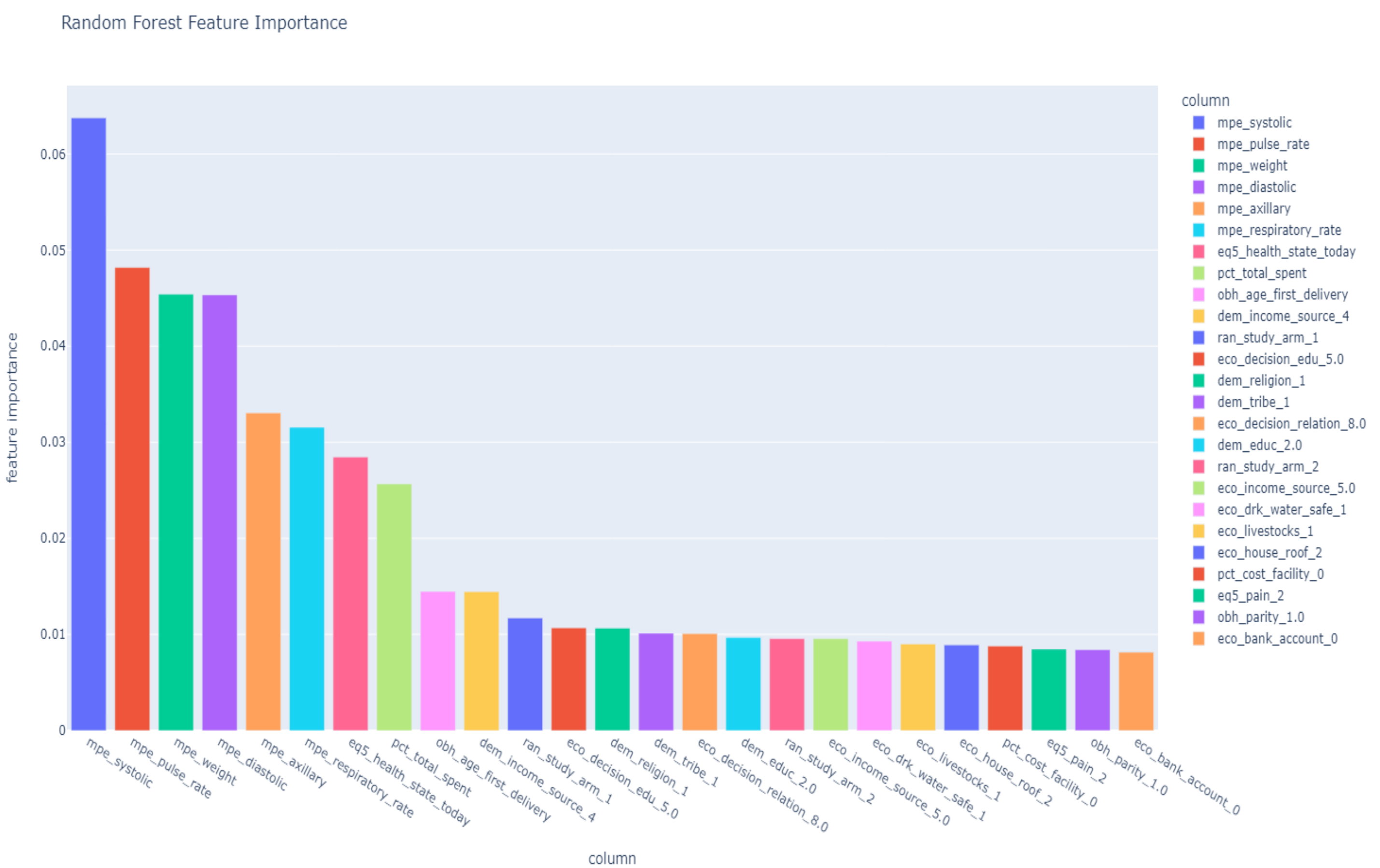


Figure 5: Random Forest Feature Selection

Random forest feature selection algorithm shows that the medical factors top the list in impacting the outcome of the pregnancies. Blood pressure levels (both systolic and diastolic) along with parity, heart rate and weight are in the top 25 features. However, there are other socioeconomic and cultural factors which are impacting the health of the conceived babies even before they are born. Economic factors include affordability of healthcare, profession, income source, possession of land and livestock and one's own house, access to safe drinking water and financial system. Social factors impacting the survival of new baby include the belonging of mother to a specific tribe, the education background of mother and the family head, and the religion they follow. Nevertheless, this study included women who were given iron supplements through oral or FCM means. This intervention has significant impact on the pregnancy outcome as well.

5.3. Permutation Feature Selection

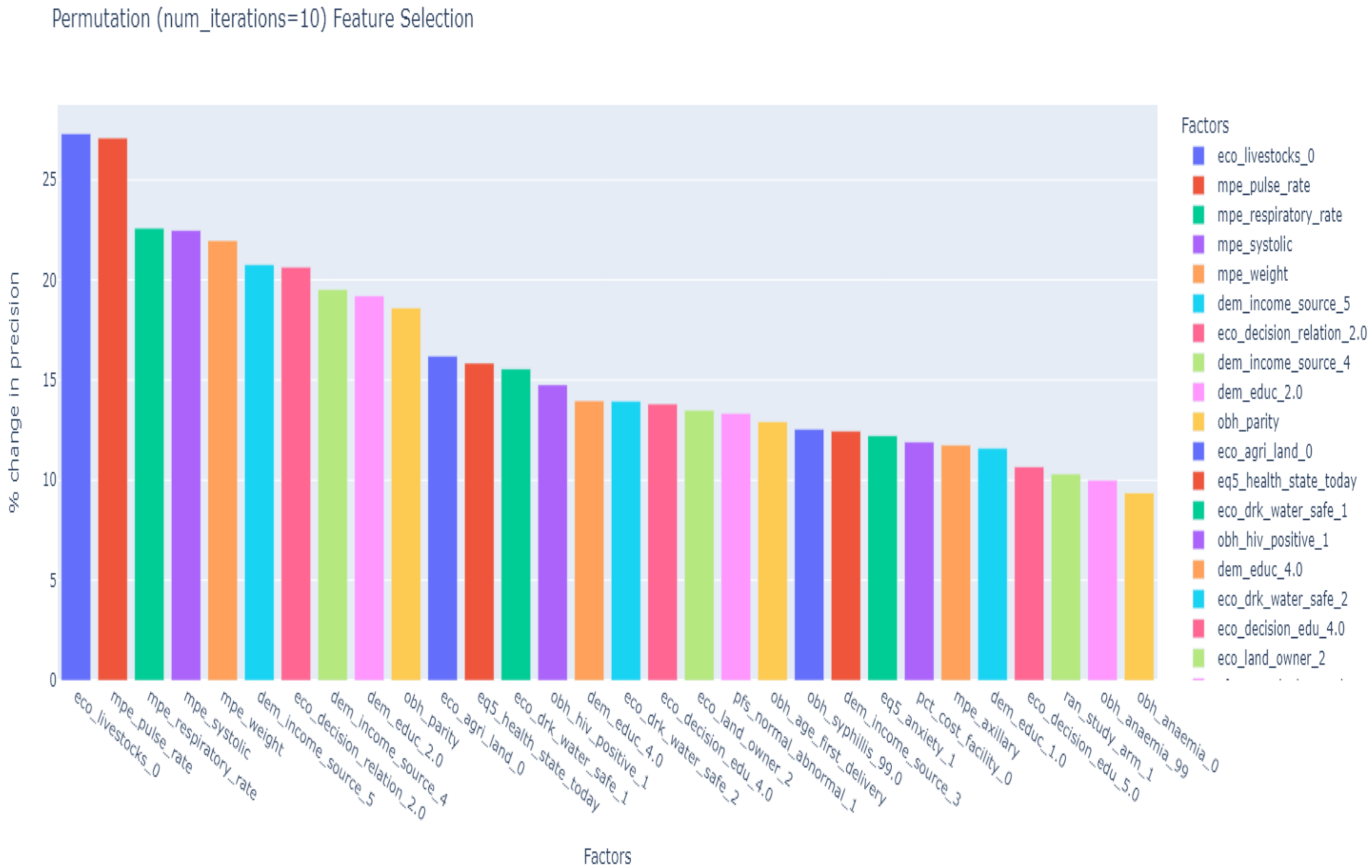


Figure 6: Permutation Feature Selection

Feature importance was also done through permutation method to get more accurate selection. Not surprisingly, most of the determinants are same but their ranking in top 25 has changed a little bit. Using the permutation importance technique, we may run the model while randomly rearranging the values of a single column to observe how the scores change. If the scores are significantly impacted, the feature is very crucial to the model; otherwise, it does not significantly improve the model. Therefore, the feature chosen by permutation are those who have higher impact on the precision of the predicting perinatal outcome. Anemia, presence of HIV, access to safe drinking water are three other factors which have made it to the top 25 factors affecting the perinatal mortality in Malawi.

5.4. SHAP (Shapely Additive Explanation) Values

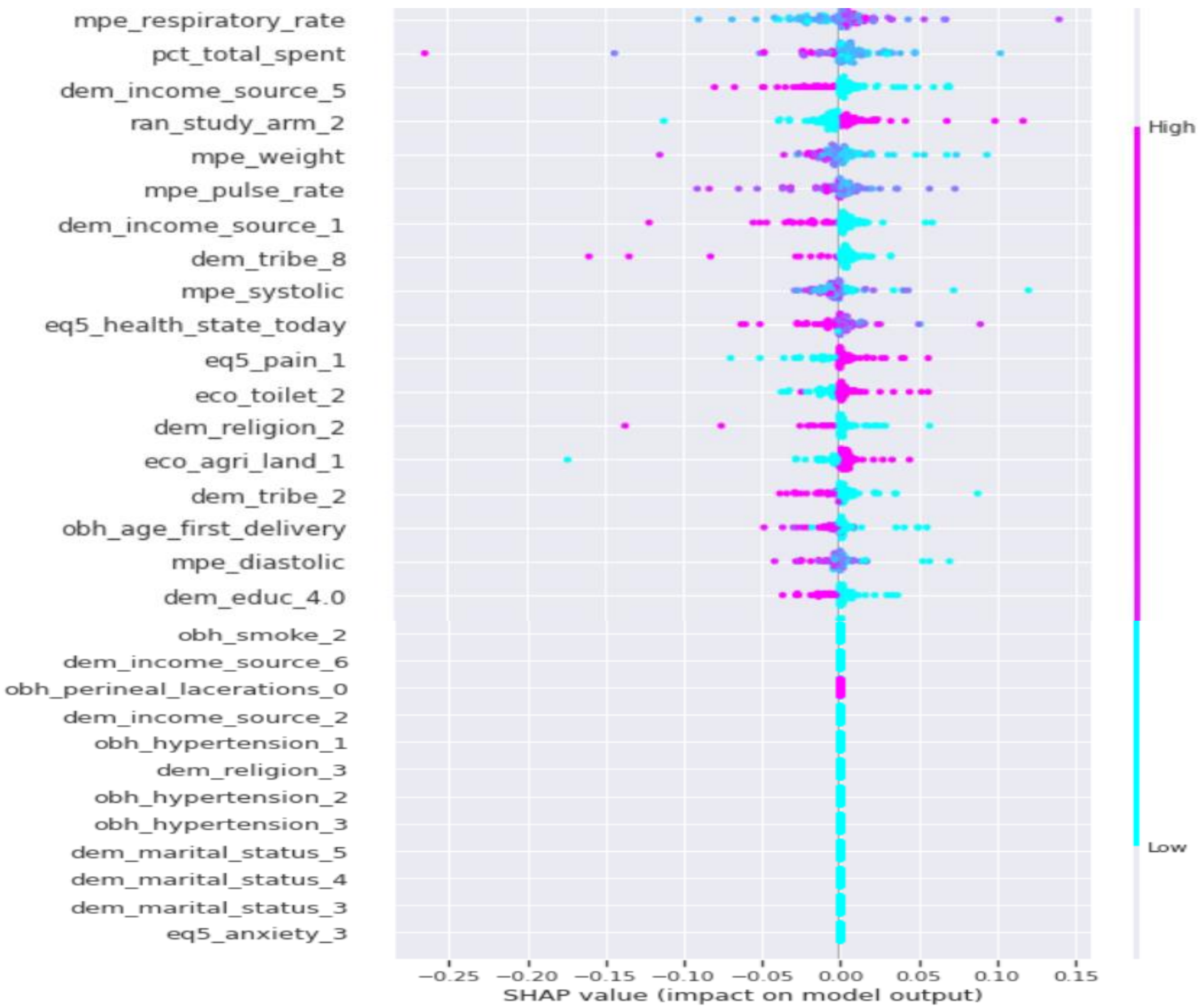


Figure 7: SHAP Values

Each dot in swarm plot in figure 7 stands for one observation. The color of the point tells us if that observation has a larger or lower value when compared to other observations, while the horizontal axis displays the SHAP value. A shap value is the average amount of contribution that a particular feature makes to the coalition value.

As it can be seen, high respiratory rate has high impact on the perinatal outcome. Similarly, lower respiratory rate has lower shap value and shows lower impact on mortality. The second prominent feature impacting the model outcome is total spending by mothers in the hospital at the time of pregnancy. Higher the spending, the greater the chance that the new baby will survive and vice versa. Income source also impacts the binary outcome of pregnany. Lower the weight in the form of blue dots, the higher the chances of perinatal mortality. Association with tribe 8 is also increase the risk of mortality. Noticeably, the feature systolic rate has similar impact with both low and high values. The lower the age at first delivery, the greater the impact it has on adverse outcome; higher the age at first delivery, lower the risk factor. The relevance of each variable is displayed with the most significant variable at the top and the least important variable at the bottom.

5.5. Performance of Machine Learning Models

Table 4: Performance of Machine Learning Models

Performance Comparison of Machine Learning Models					
Model Type	F1	ROC_AUC	accuracy	precision	recall
Logistic Regression	0.333333	0.575758	0.714286	0.333333	0.333333
Decision Tree	0.297872	0.543290	0.663265	0.269231	0.333333
K-Neighbors Classifier	0.258824	0.417749	0.357143	0.171875	0.523810
Gradient Boosting	0.476190	0.666667	0.775510	0.476190	0.476190
Support Vector Classifier	0.148148	0.521645	0.765306	0.333333	0.095238
Random Forest	0.275862	0.569264	0.785714	0.500000	0.190476

Table 4 and figure 8 show the performance of six machine learning models in predicting perinatal mortality. They are compared in terms of accuracy, ROC_AUC (receiving Operating Characteristic and Area Under the Curve), precision, recall and F1 scores. The accuracy of LR, DT, KN, GB, SVC AND RF are 0.71, 0.66, 0.35, 0.77, 0.76, and 0.78 respectively. Random forest demonstrated the highest accuracy and precision scores. K-Neighbor was best for scoring 0.52 in recall but its other evaluation measures, especially its accuracy score of just 0.35, were really poor against rival models. Compared to the LR, DT, RF, KN and GB models, the SVM model's recall was worse. Gradient bosting and random forest are two best performing models overall. ROC_AUC score of GB makes it superior to Random Forest in predicting adverse outcome of pregnancy in Malawi.

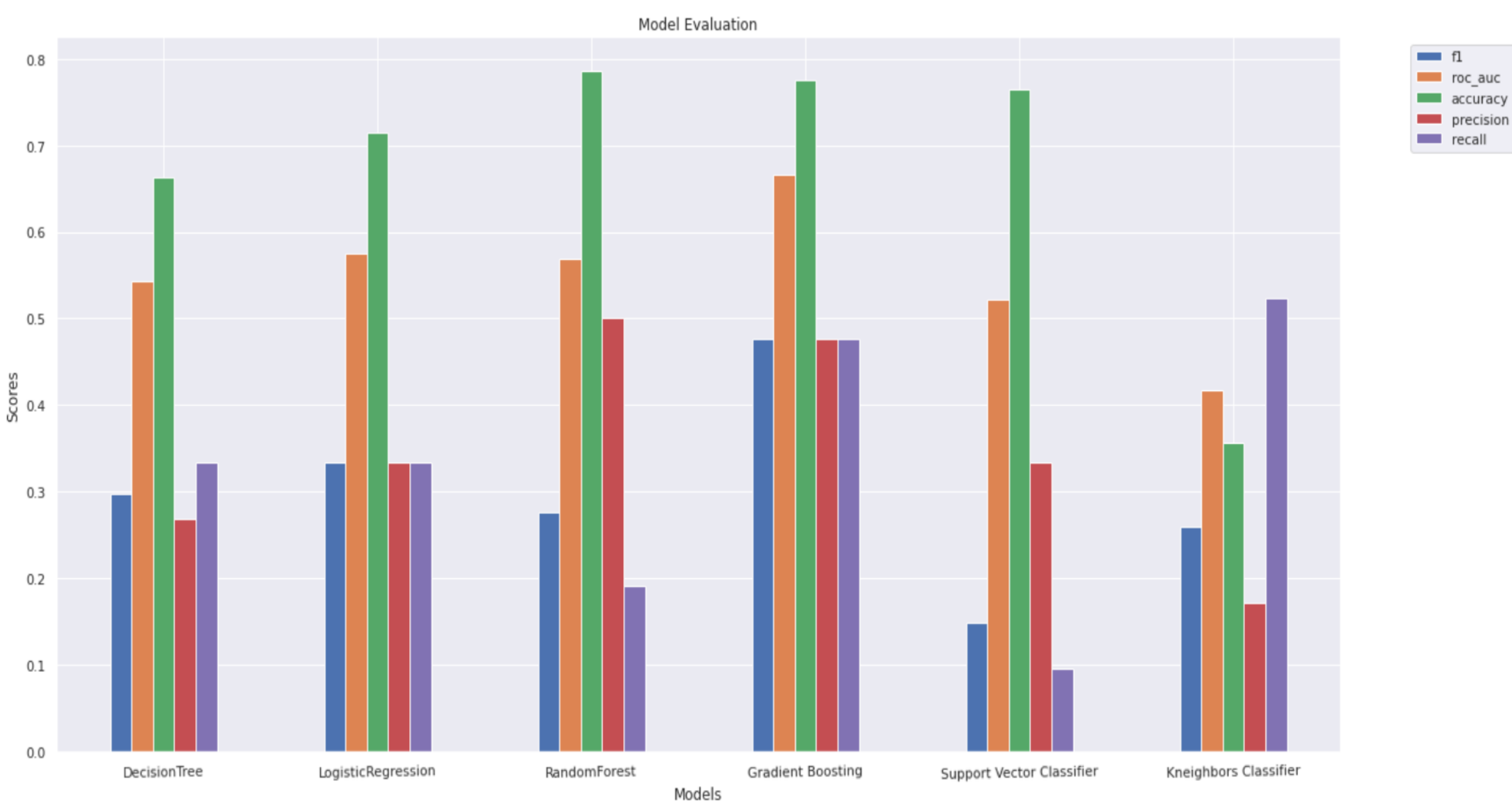


Figure 8: Pictorial comparison of Performance of Machine Learning Models

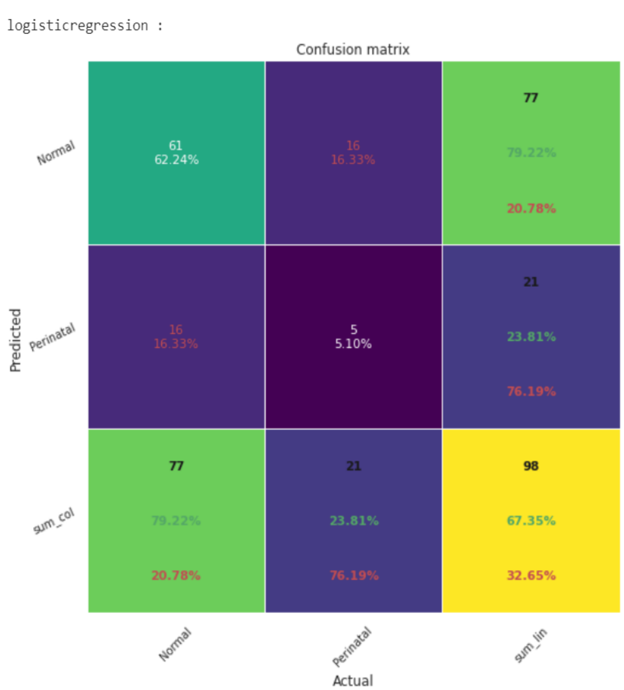
5.6. Hyperparameter Tuning and Enhanced Performance of Selective Models

Table 5: Hyperparameter Tuning

Best Parameters for LR, GB and RF.			
Classifier	**ROC_AUC_Score**	**time**	Hyper Parameters
Logistic Regression	0.731097	1.504910 s	{'clf__C': 0.1, 'pca__n_components': 40}
Gradient Boosting	0.805804	11.302566 s	{'clf_learning rate': 0.1, 'clf__max_depth': 5, 'clf__n_estimators': 80, 'pca__n_components': 40}
Random Forest	0.803132	52.884357 s	{'clf__criterion': 'entropy', 'clf__max_depth': 4, 'clf__min_samples_leaf': 0.05, 'clf__min_samples_split': 0.05, 'clf__n_estimators': 100, 'pca__n_components': 50}

Table 5 shows that the models have improved their ROC_AUC performance after hyperparameter tuning. For logistic regression, principal components analysis (PCA) best components are 40 and its penalty term, C, is 0.1. After parameter adjustment, the ROC AUC scores for Gradient Boosting and Random Forest were practically identical at 0.80. But in terms of timing, random forest is more expensive. For a gradient boosting classifier, the learning rate, maximum depth, n estimators, and pca n components are each 0.1, 5, 80, and 40, respectively. With a maximum depth of 4, the best criterion for random forests is entropy. For random forest, N Estimator and pca components are 100 and 50, respectively.

5.7. Confusion Matrix



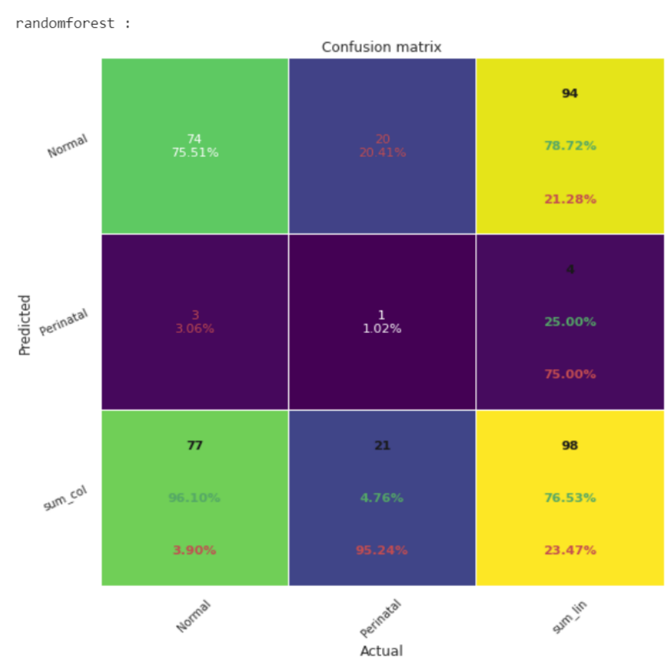


Figure 9: Logistic Regression Confusion Matrix
Figure 10: Random Forest Confusion Matrix

Figure 9-11 show the confusion matrixes of LR, RF and GB models respectively. Our test set has only 98 samples for prediction. Out of those samples, type I error, which is predicting perinatal outcome as normal, is 16 by logistic regression, 20 by random forest and 15 by gradient boosting. Type II error – predicting normal deliveries as perinatal – was 16, 1, and 11 for LG, RF and gradient boosting respectively.

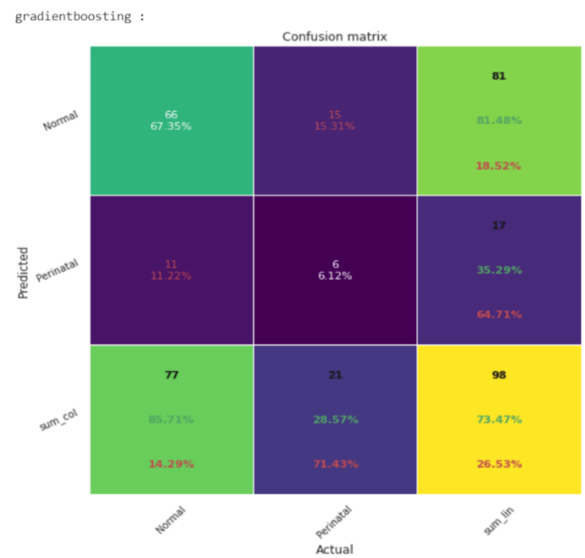


Figure 11: Confusion *Matrix Gradient Boost*

6. Discussion

6.1. Technical Discussion

This study's most important finding is that the machine learning method for predicting perinatal death in Malawi is quite convincing. Gradient boosting and Random Forest have shown 76 and 78 percent accuracy in identifying the outcome of a pregnancy beforehand. As shows in figure 10 and 11, these models have lower type I errors as well. As cost of predicting a potential adverse outcome as normal is way higher than the predicting an adverse outcome as a normal baby. Intervention is initiated after the model rightly picks mothers who need special care for them and their babies. ROC_AUC score for gradient boost, which shows a relationship of False positive (specificity) on X-axis and true positive (sensitivity) on y-axis, was 0.66 without hyperparameter tuning. It improved to 0.80 after tuning. It has comparable accuracy with Random Forest as well. However, Decision Tress, KNeighterst Neighbors, and Support Vector Machine were not really impressive in predicting the adverse outcome. Their ROC_AUC scores were lower than 0.60. Nevertheless, Logistic regression has proved to be efficient model as it takes the least time to fit and predict with a good accuracy. This is backed by Stoltzfus study as well where he emphasized that predictive ability of LR has been proven time and again and it has been frequently compared to models based on artificial intelligence (Stoltzfus, 2011).

The best performance was shown by ensemble algorithms, i.e., Random Forest and Logistic Regression, and they have been proved by other studies to be more effective (Do, et al., 2022). The foundation of an ensemble approach is the idea that, due to significant volatility, a collection of basic models may perform better than overfitted models. The RF implements decision tree ensembles. The decision-making process is described by each tree, and each node's value of one attribute is compared to a threshold to determine which branch to take. During the learning phase, the tree's structure and thresholds are decided. The predictions are combined to create an ensemble output using RF, which constructs numerous decision trees that have been trained on training data. Thus, ensemble models predict the perinatal mortality in Malawi with prominent accuracy.

Given the tradeoff of Logistic Regression time with accuracy of ensemble methods, the latter method is preferred because it is more precise in predicting the death of preterm children than that provided by the LR. Therefore, GB and RF may produce objective results and display great clinical applicability. Gradient Boosting ML technique has the benefit of being non-parametric as well, and it does not rely on distribution assumptions, and permits the combination of higher-order nonlinear correlation between the predictors. Cross-validation further reduce the chance of overfitting in ML models. Therefore, ML models in a clinical setting can aid in predicting prenatal mortality and aiding in the selection of appropriate treatments for this patient population.

6.2. Practical Discussion

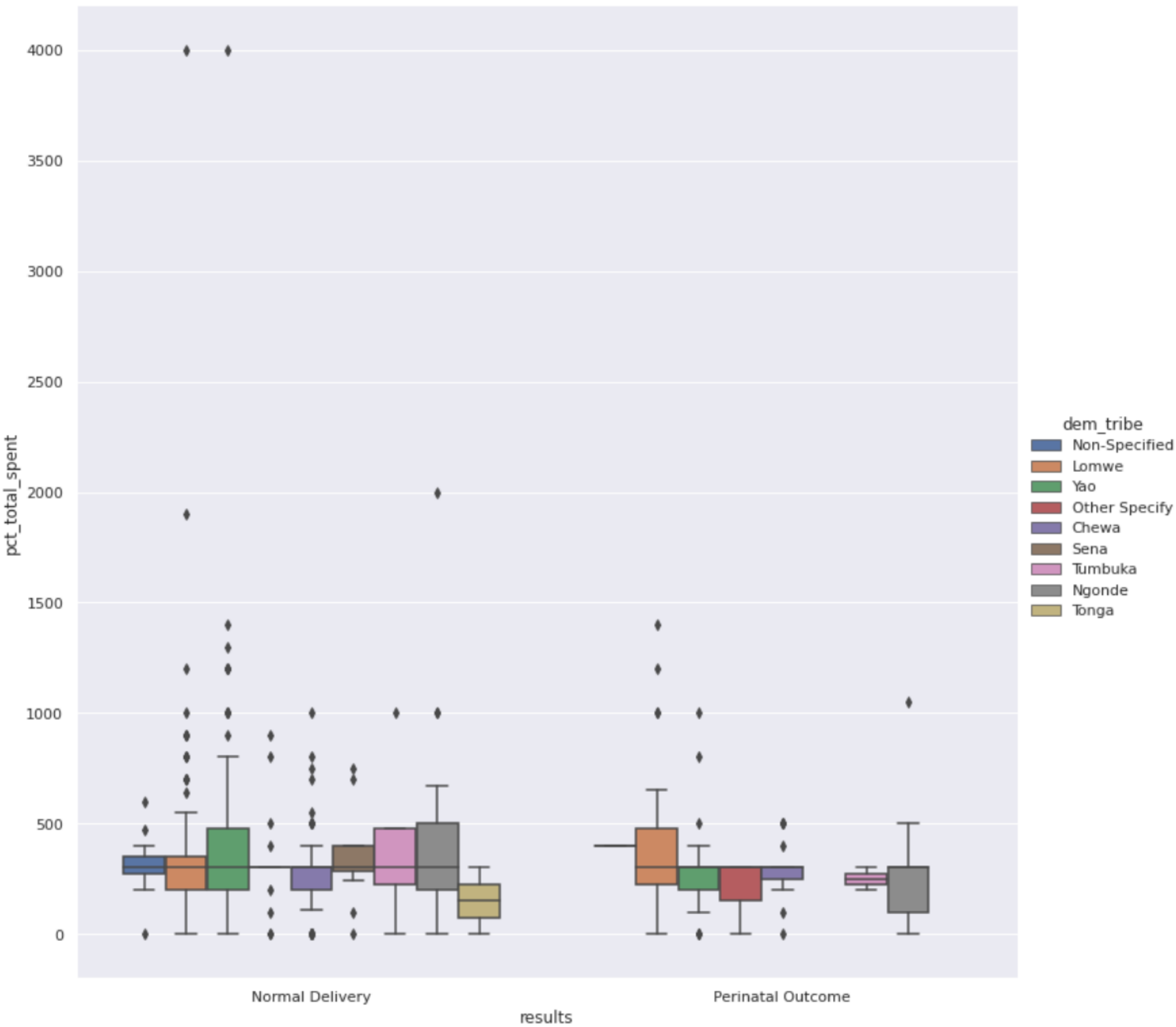


Figure 12: Tribe, spending and outcome relation

Feature selection and SHAP values have ranked risk factors in term of their impact on the outcome. Some of those identified elements are discussed here. Above figure 12 is really important. It shows how outcome is related with total spend. As we saw in table 3, average spend of those women who had perinatal outcome was 335 against 395 dollars for the other group. This picture illustrates the same relation: left side has higher total spend than the right side. Moreover, we also know form feature selection and SHAP values that certain tribes are more vulnerable than others. For instance, Lowme and Ngonde tribes have a greater number of casualties than they had normal deliveries. Interestingly, it shows both intratribal and intertribal differences in spending among tribes. For instance, we see women in Yao(green) and Chews tribe with higher spending have normal outcomes against women of the same tribes who spent less. Therefore, income inequality within tribes is also playing a role here. Tumbuka tribe has least number of babies dying in the study. Nevertheless, Lomwe tribe has an eccentric relation with very high number of perinatal outcomes despite higher spending from most of those deprived women. It should be investigated why the group has highest mortality rates among all the groups even with relatively higher spending.

The primary purpose of the data collection was to check whether the pregnant women with iron deficiency will have safer deliveries after giving them iron thorough oral or Ferric Carboxymaltose (FCM) infusion. Machine learning algorithm in the form of SHAP values indicate that the infusion is substantial in pregnancy outcome. As seen in figure 13, iron given through oral method produced better results than FCM. It is revealing for the government and international organizations, such as WHO and USAID, when it comes to launching next policy to treat iron deficient pregnant women. They should prioritize Oral iron over FCM for better results.

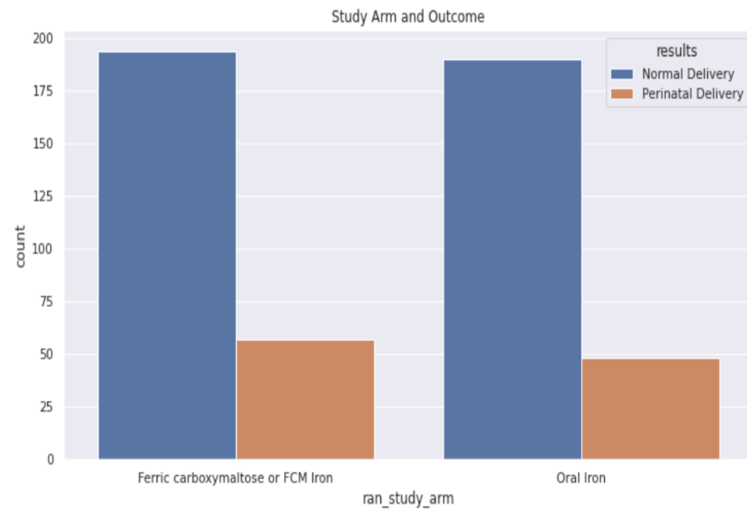


Figure 13: Study Arm and Outcome

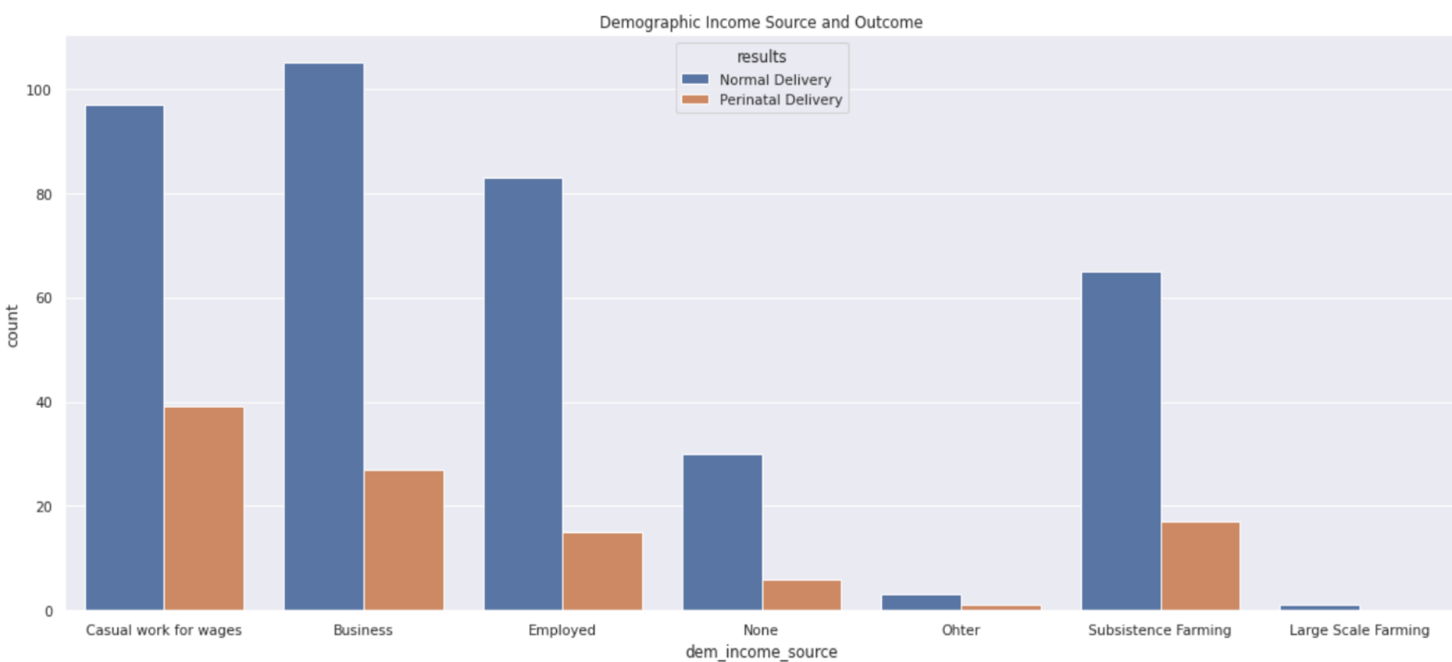


Figure 14: Demographic Income Source and Outcome count

Income source is also crucial for determining the pregnancy outcome. Above figure 14 shows that casual workers, i.e., daily wagers, are the most vulnerable people. Those who belong to this class have their babies dying in large numbers. It is understandable that the group might not get sufficient funds to look after health-related issues of the mother and the baby. It is noticeable that there was no mortality in the group with large scale farming. Therefore, income source also determines the outcome of the pregnancy in Malawi. Families belonging to businesses and farming have almost similar causalities.

Mother's education is another important factor which separated safe and perinatal deliveries. The same factor has been a recurrent theme in similar studies (WHO, 2006). Mothers whose education level is primary are more in numbers than the mothers with secondary education. Even though there were just 20 mothers with tertiary education, only one had an adverse outcome. However, out of 200 mothers with low primary background, 45 had negative outcome – the mortality rate reaches 22.5 percent form 5 percent of those mothers with tertiary-education background.

Below figure 15 shows the relationship of mothers age at the first delivery and outcome of the last studied pregnancy. First time mothers were higher in number when it comes to adverse outcome. First time mothers are known to have more risks associated with them for perinatal mortality (Mfateneza, et al., 2022). Almost 80 percent of the mothers whose babies dies belong to the group whose age at first delivery was less than 20 Years. It is startling social factor which is making huge impact on the pregnant women. Government should look into early marriages and pregnancies.

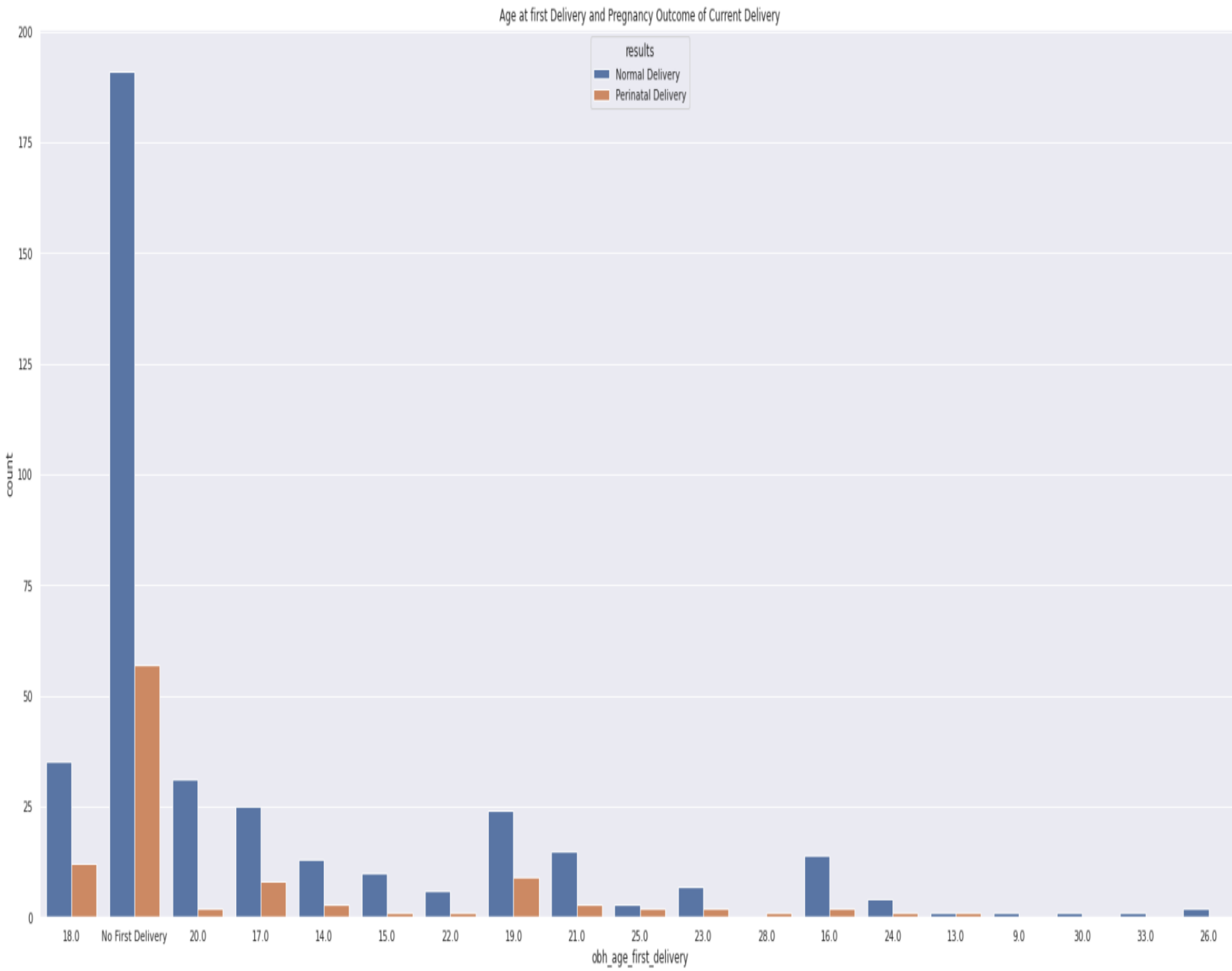


Figure 15: Age at first delivery and outcome of current delivery

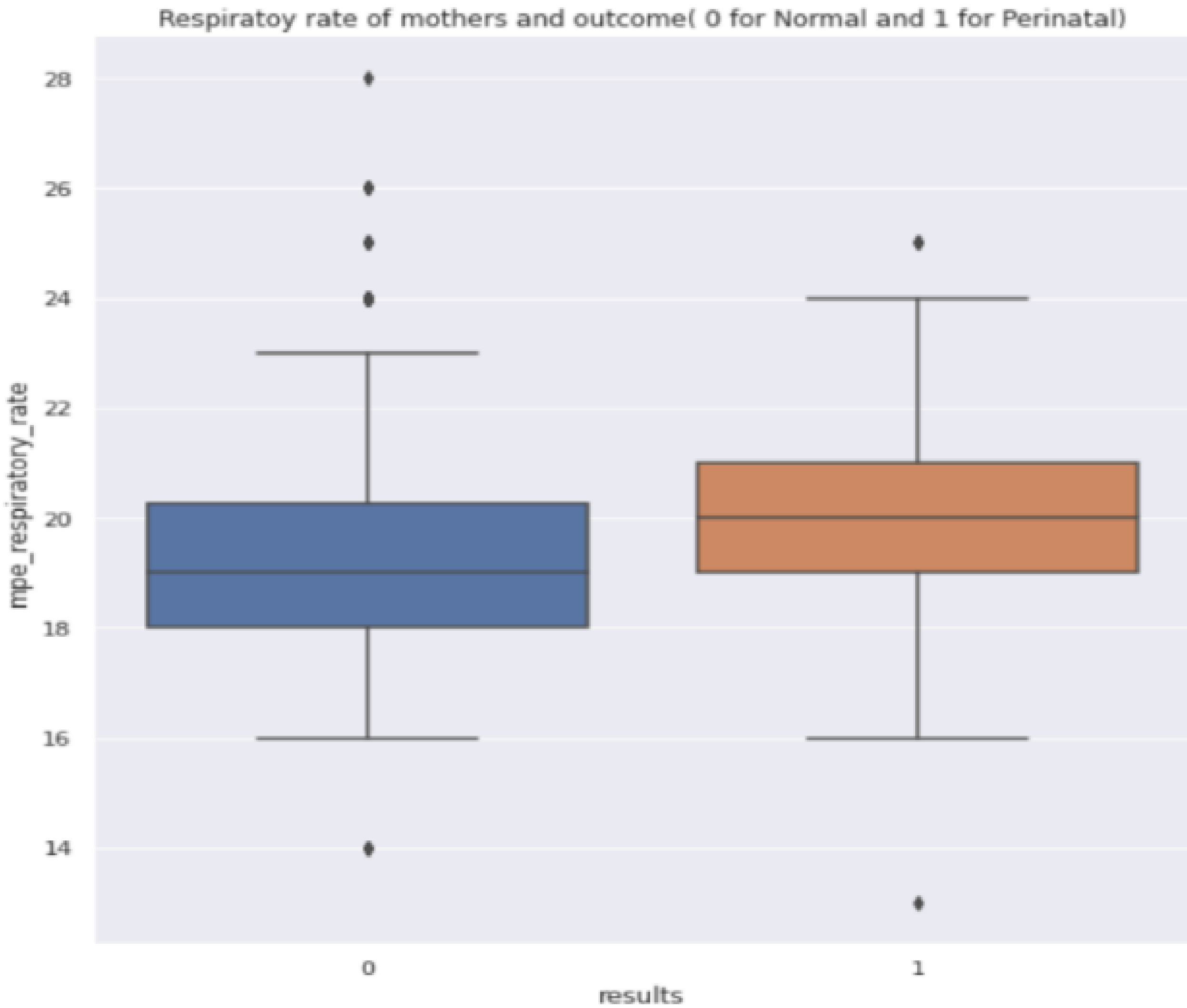


Figure 16: Respiratory Rates of Normal and Perinatal outcome

Respiratory rate has the highest SHAP value in predicting the binary outcome. This represents how many breaths you take in a minute. Respiratory rate can tell how well your body is working to get oxygen to all of your essential organs and tissues by observing the speed, pattern, and depth of your breaths (Cynthia, 2021). It can be seen in figure 16 that the average respiratory rate of those women who had adverse outcome is about 20. Normal range for an adult is considered to be 16-20 (Cynthia, 2021). However, about one fourth of women whose deliveries resulted in adverse outcome had respiratory rate above than the normal average range. The Healthline study claims that “If your respiratory rate is above average, it could indicate another underlying condition” (Cynthia, 2021).

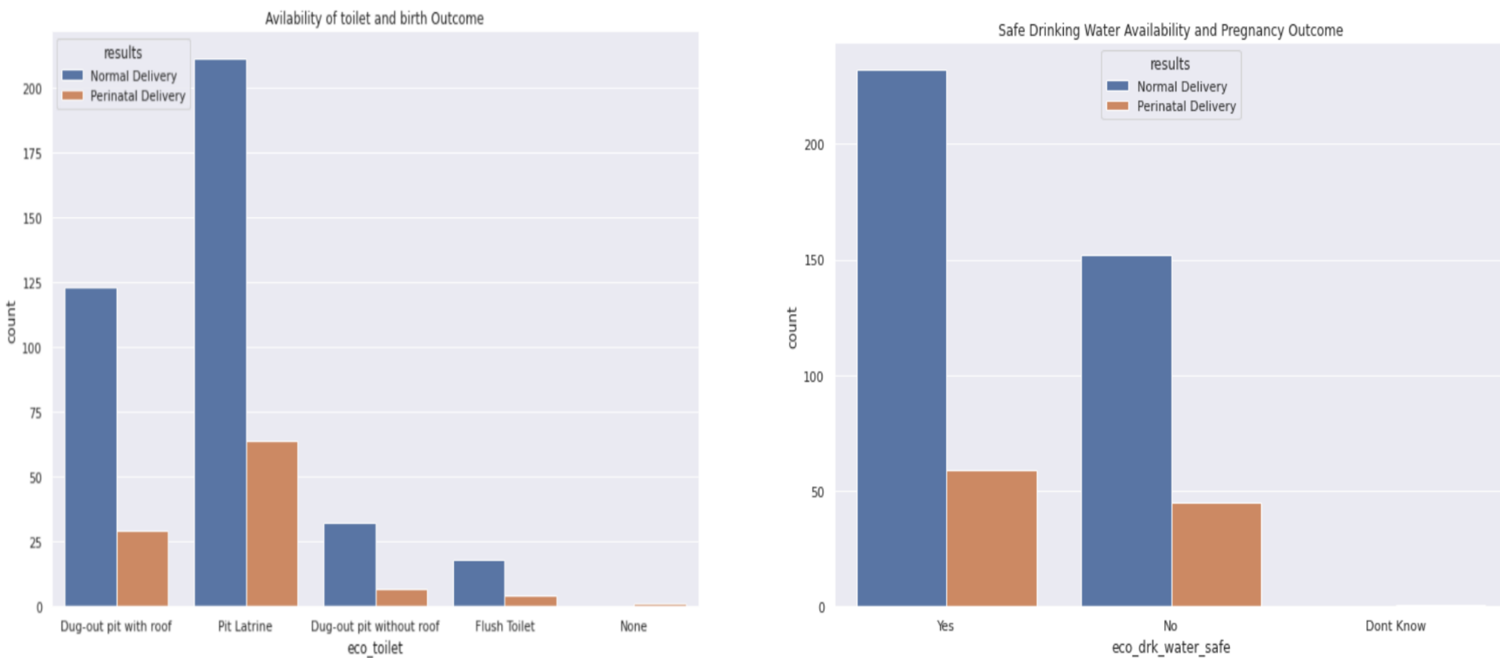


Figure 17: Access to Toilet and Safe Drinking water and outcome

Above figure 17(left) shows the relationship of availability of toilet and the binary outcome. The factor tells economic situation of the household as well. Household whose toilets are not made up of concrete, i.e., either dug-out or pit latrines, have shown high adverse outcomes. It clearly demonstrated how vulnerable people are in this part of the world. Unavailability of basic amenities such as water in figure 17(right) clearly has proportional impact on the adverse outcome. Lack of toilet and safe drinking facilities give rise to different diseases because of unhygienic way of living, and this then is reflected in the outcome of pregnancies. These two factors are identified by other studies in impacting perinatal mortality as well (Mfateneza, et al., 2022). Women who live with diseases are bound to suffer in the form of perinatal mortality.

Unfortunately, people who belonged to certain belief are identified as more vulnerable. Figure 18 shows that around 18 percent Muslim mothers had negative outcome compared to 22 percent of Christian women. This is reflected in both Random Forest feature selection and SHAP values as well. The underlying social factors should be investigated for true assessment of reasons. However, future policymaking should take religion into account for inclusive treatment. According to the 2018 census, 13.8 percent of the population is Muslim, and 77.3 percent is Christian. This study consists of just 489 women, so the results could not be generalized for all population. Nevertheless, it should be investigated.

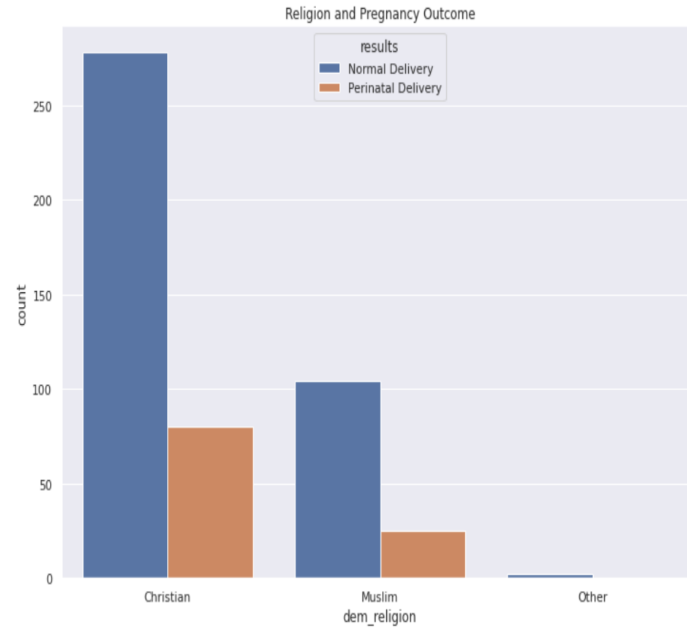


Figure 18: Religion and Pregnancy Outcome

7.Limitations

There are some Limitations of this study. First and foremost, it cannot be reflective of general population. This data was obtained from a controlled group of pregnant mothers who had iron deficiency. They were mainly observed for treatment of iron deficiency through FCM infusion or oral methods. Therefore, this study could not be reflective of all Malawi population. It can only help predict mortality in iron-deficient mother in Malawi.

Second limitation of this study is uneven distribution of its features. For instance, some tribes are overwhelming in majority in our dataset. It brings bias into prediction when one class within feature is dominating. Furthermore, there were mothers who were having their first pregnancy. Half of the mothers were those who had not previously given birth. Therefore, comparing first time pregnant women with those who have pregnancies would be challenging. For instance, simple imputation in missing values in the feature Age of mothers at the first delivery, was not easy. When we have unbalanced classes and they have missing values, it becomes harder to fill them without bias.

Third limitation of this study is that it misses some of the important features identified by previous studies in similar settings. For instance, it does not have gender of baby as a feature for pregnancy outcome. Similar studies have shown that the gender of the conceived baby impacts the survivability of the newborn (Nijkamp, et al., 2017). Past studies have also proven that the data collected at the time of delivery (respiratory rate of baby and a like), and in following two days after the birth, increase the predictive accuracy of perinatal mortality using machine learning models (Vivek V. Shukla, et al., 2020). Unfortunately, this study has not incorporated the first two days data of the newborn babies.

Fourthly, all the factors are narrowly focused on the health of mother. We don't know about the training of staff and other facilities. Stillbirths and neonatal mortality are influenced by the labor quality in hospital and availability of special care units (with working vents) for the newborn struggling babies. For instance, neonatal resuscitation units are instrumental in saving lives of struggling babies in the earlier hours.

Even though stillbirths are part of our adverse outcome in output variable, there is no distinction between intrapartum or antenatal stillbirths. It is important to know the nature of stillbirth as past studies have shown maternal illness as major cause of antenatal stillbirths and obstetric crises as main cause of intrapartum stillbirths (Lawn, et al., 2011).

Vivak Shukla warned in its paper in JAMA that it is yet to find out whether the machine learning technique is as effective on small datasets as it is on large data sets (Vivek V. Shukla, et al., 2020). The dataset for this dissertation has various features but small sample size. The study in being analyzes on 489 women and that makes it a small sample size. ML technique are proven to be elite when predicting outcome from large datasets.

8.Conclusion

There is no denying the fact that machine learning models can predict perinatal morality very well. Even in the case of Malawi data with relatively small sample, the highest accuracy score was 0.78. Ensemble methods especially Random Forest and Gradient Boost showed the ROC_AUC scores of 0.66 and 0.50 respectively for the complex data set. Moreover, precision and recall scores of these two models were also considerable. Elite performance of ensemble models is proven by Huan and Kyong as well for their study where they predicted mortality in low-weight infants (Do, et al., 2022). Nevertheless, Logistic regression also performed very well after hyperparameter tuning and showed an improved roc_auc score of 0.73 as well. As backed by Jill Stoltzfus in similar setting, LR also proved to be the fastest to fit and predict the target variable (Stoltzfus, 2011). Even though the study used various socioeconomic and medical features, top risk factors were identified through random forest feature selection and SHAP values. The pregnant women included in this study had underlying iron deficiencies. Medical risks were expected, and the models rightly identified high respiratory rate, blood pressure and anemia to be the major risk factors associated with the perinatal mortality in Malawi. Moreover, economic factors such as spending by mothers in hospital, possession of land and livestock, access to safe drinking water, and belonging to daily wage earner class were playing role in determining the adverse outcome of the delivery. Moreover, social factors such as pregnancy at very early age, affiliation with a particular religion or tribes, illiteracy or low education of mothers were also affecting the mortality rates. All of these factors have been identified in the similar studies on Rwanda (Mfateneza, et al., 2022) and Tanzania (Mitao & Philemon, 2016) in Africa.

This study has few limitations as well. It does not use a large dataset to be considered for general representation of the population in Malawi. Therefore, further studies should consider large dataset with inclusive representation of women from all over the country. Even though the study used 56 input variables, it misses data with few important dimensions. For instance, a study by World Health Organization (WHO, 2006) on neonatal mortality, a vital part of perinatal mortality, found that medical infrastructure and manpower in the form of midwives and trained gynecologists also play an important role in the survival of newborn babies. Lack of care for newborn baby in the first few hours can be deadly. Furthermore, precision of predicting perinatal mortality can be increases by incorporating the quality of health infrastructure as well. Both maternal illness and obstetric crises should be part of the future studies for better F1 and roc_auc scores.

9. Policy Recommendations

Based on the fining of this study, following recommendation are suggested. The government in Malawi and International organization such as WHO can find out risk factors associated with perinatal mortality form this study. While treatment of iron deficient pregnant women, the stakeholders should prioritize Oral infusion over FCM method. Secondly, a mass campaign should be launched to find out pregnant women who have high respiratory issues as they can be caused by other undiscovered diseases. Thirdly, medical professionals should visit pregnant women to monitor the conception over the gestational period of 9 months. This will help in finding the medically ill mothers and a correction course can be suggested to avoid any negative outcome at the time of delivery. Fourthly, those who cannot afford to pay the hospital bill should be brough under safety net to save lives of infants who cannot afford the luxury of simple deliveries in hospital. This exercise will encourage poor mothers to give birth in hospital settings. Fifthly, the government should enforce laws to protect daily wage earners as they are found to have higher mortality rate if all other professions.

Sixthly, the women should get at least access to clean drinking water and latrine. It is unfortunate that the unhygienic water and sanitation is major causes of illnesses in the pregnant women and their conceived babies. Besides, the government should promote education in women as low education is highly correlated with adverse pregnancy outcome. United Nation's Children Fund(UNICEF) should also promote higher education in Malawi for women to save newborn babies as well. Investigation should be initiate to find out high rates of mortality in Lowme tribe. If Malawi is to achieve its SDG goal 3.2 to reduce its perinatal mortality to 12 per 1000 livebirths, above suggestions will be indispensable.

.

10. References

Aguliar, F., 2019. SMOTE-NC in ML Categorization Models for Imbalanced Datasets. *Analytics Vidhya*, 16 October.

AHMAD, S. T., A. B. H., GHADA, A. A. & ABDULLAH, A., 2022. Stop Oversampling for Class Imbalance. *IEEE Open Access,* 10(1109).

Akalewold, T. G. et al., 2021. Gender differences in survival among low birthweight newborns and infants in sub-Saharan Africa: a systematic review. *International health,* pp. 122-131.

Allanson, E. & Tunçalp, Ö., 2016. Classifying the causes of perinatal death. *Bull World Health Organ,* pp. 79-79A.

Chawla, N. V., Bower, K. W., Hall, L. O. & Kegelmeyer, W., 2022. SMOTE: Synthetic Minority Over-sampling Technique. *Journal Of Artificial Intelligence Research ,* Volume 16, pp. 321-357.

Cynthia, T. C., 2021. *What Is a Normal Respiratory Rate for Adults and Children?.* [Online]
Available at: https://www.healthline.com/health/normal-respiratory-rate
[Accessed 27 08 2022].

Do, H., Moon, K. & Jin, 2022. Machine Learning Models for Predicting Mortality in 7472 Very Low Birth Weight Infants Using Data From a Nationwide Neonatal Network. *Diagonistics ,* Volume 12, p. 625.

Goto, T. et al., 2019. Machine Learning–Based Prediction of Clinical Outcomes for Children During Emergency Department Triage. *Emeergency Medicine .*

Lamping, F., Thomas, J., Rubsamen, N. & Sasse, M., 2018. Development and validation of a diagnostic model for early differentiation of sepsis and non-infectious SIRS in critically ill children - a data-driven approach using machine-learning algorithms. *BMC Pediatrics ,* pp. 18-112.

Lawn, J. E. et al., 2011. Stillbirths: Where? When? Why? How to make the data count?. *Lancet,* Apr 23.p. 1448–1463.

Mangold, C. et al., 2021. Machine Learning Models for Predicting Neonatal Mortality: A Systematic Review. *Neonatology,* pp. 394-405.

Mboya, I. B. et al., 2020. Prediction of perinatal death using machine learning models: a birth registry-based cohort study in northern Tanzania. *BMJ Open,* p. 10:e040132.

Mfateneza, E. et al., 2022. Application of machine learning methods for predicting infant mortality in Rawanda. *BMC Pregnancy and CHildhood,* p. 221.

Mitao, M. & Philemon, R., 2016. Risk factors and adverse perinatal outcome associated with low birth weight in Northern Tanzania: a registry-based retrospective cohort study. *Asian Pacific Journal of Reproduction,* pp. 75-79.

Nijkamp, W. et al., 2017. Perinatal death investigatons: What is current practice?. *Seminars in Fetal and Neonatal Medicine,* 22(3), pp. 167-175.

Pitt, C., Tawaih, T. & Seyi, S., 2015. Cost and cost-effectiveness of newborn home visits: findings from the Newhints cluster-randomised controlled trial in rural Ghana. *THE LANCET Global Health,* pp. E45-E56.

Raita, Y., T, G., Faridi, M. & Brown, D., 2019. Emergency department triage prediction of. *Critical Care,* pp. 23-64.

Stoltzfus, J. C., 2011. Logistic Regression: A Brief Primer. *Academic Emergency Medicine,* 18(10), pp. 1099-1104.

Trent, D. B., Antje, K., Adam, E. & Curtis, S., 2018. An Introduction to Machine Learning Methods for Survey Researchers. *Survey Practice,* Volume 11, p. 2718.

UNICEF, 2015. *Maternal and Newborn Health Disparities Malawi,* s.l.: UNICEF.

UNICEF, 2016. *UNICEF Annual Report 2015.* [Online]
Available at: https://www.unicef.org/reports/unicef-annual-report-2015

USAID, 2015. *Malawi : Profile of preterm and low birth weight prevention and care,* s.l.: Every Preemie Scale.

Velido, A. & Ribas, V., 2018. Machine Learning in Critical Care: state-of-the -art and Sepsis case study. *Biomedical Engineering online ,* p. 17:135.

Vinícius, T., 2022. *Using SHAP Values to Explain How Your Machine Learning Model Works.* [Online]
Available at: https://towardsdatascience.com/using-shap-values-to-explain-how-your-machine-learning-model-works-732b3f40e137#:~:text=SHAP%20values%20(SHapley%20Additive%20exPlanations,interpretability%20of%20machine%20learning%20models.
[Accessed 24 August 2022].

Vivek V. Shukla, M. et al., 2020. Predictive Modeling for Perinatal Mortality in Resource-Limited Settings. *JAMA Network Open,* pp. 3-11.

WHO, 2004. *Neonatal mortality rate,* s.l.: World Health Organization.

WHO, 2006. *Neonatal and Perinatal Mortality: country, regional and global estimates.,* s.l.: WHO Library Cataloguing-in-Publication Data.

WHO, 2016. The WHO application of ICD-10 to deaths during the perinatal period: ICD-PM. *World Health Organization.*

WHO, 2018. *Country Cooperation Strategy at Glance,* s.l.: WHO.

WHO, 2022. *Newborn Deths.* [Online]
Available at: https://www.who.int/data/gho/data/themes/topics/indicator-groups/indicator-group-details/GHO/newborn-deaths

WHO, 2022. *World health statistics 2022: monitoring health for the SDGs,* Geneva: WHO.

Zacharias, N., 2020. *Perinatal Mortality.* [Online]
Available at: https://www.uptodate.com/contents/perinatal-mortality/print