Determining the Optimal Machine Learning Algorithm to Predict Pre-Term birth Maternal Health using Electronic Health Records in Kenya

By

Waswa Paul Wekesa

167860

Submitted in Partial Fulfillment of the Requirements for the Master's Degree of Data Science and Analytics to the Institute of Mathematical Sciences (SIMS), Strathmore University

Strathmore University Institute of Mathematical Sciences (SIMS)

Strathmore University

Nairobi, Kenya

June 2025

This dissertation is available for Library use on the understanding that it is copyright material and that no quotation from the dissertation may be published without proper acknowledgement.

Declaration and Approval

Declaration

I declare that this work has not been previously submitted and approved for the award of a degree by this or any other University. To the best of my knowledge and belief, the dissertation contains no material previously published or written by another person except where due reference is made in the Dissertation itself.

© No part of this Dissertation may be reproduced without the permission of the author and Strathmore University

Student's Name: Paul Wekesa Waswa

Sign: ______________________ **Date:** 04/06/2025

Approval

The Dissertation of **Paul Wekesa Waswa** was reviewed and approved for examination by the following:

Dr. Betsy Muriithi,
Lecturer, Institute of Mathematical Sciences (SIMS),
Strathmore University

Dr. Godfrey Achono Madigu,
Dean, Institute of Mathematical Sciences,
Strathmore University

Prof. Bernard Shibwabo,
Director of Graduate Studies,
Strathmore University

Abstract

Preterm birth (PTB) remains a leading cause of neonatal morbidity and mortality globally, with the burden disproportionately higher in low- and middle-income countries (LMICs). Despite advances in maternal healthcare, the early prediction of PTB remains challenging due to the multifactorial nature of its causes and limitations in traditional risk assessment tools. This study sought to develop and evaluate machine learning models for PTB prediction using routine maternal clinical data in a low-resource setting.

This study employed a quantitative, retrospective cohort study design using routinely collected maternal health data from a selected health facility in Kenya. Clinical data were retrospectively extracted from a Level 4 healthcare facility in Bungoma County, Kenya. The dataset included demographic, obstetric, and vital sign parameters collected during delivery between 2023 - 2025. After data cleaning, handling missing values, and addressing class imbalance using SMOTE, several machine learning models were trained and tested. These included Logistic Regression, Support Vector Machines (SVM), Random Forests, and gradient boosting models such as XGBoost, LightGBM, CatBoost, and AdaBoost.

Among the models, CatBoost demonstrated the most balanced performance with an accuracy of 0.582, recall of 0.693, F1-score of 0.591, and the highest AUC of 0.608. AdaBoost achieved the highest sensitivity (recall of 0.789) but had a lower overall accuracy (0.547). XGBoost and LightGBM also performed moderately well with AUCs of 0.607 and 0.606 respectively. Feature importance analysis revealed fundal height, temperature, and respiratory rate as the most influential predictors. SHAP analysis confirmed the non-linear and interactive contributions of these features to model predictions.

While the results show potential for ML-based risk stratification tools, the predictive performance of the models remains modest when compared to thresholds reported in literature (AUC greater than 0.70) for high income countries. However in LMICs the threshholds have been reported lower from 0.6161 which is consistent with our study findings. These limitations are likely due to the absence of longitudinal data such as ANC, nutrition data and demographic data that has been reported to be key in pre-term birth prediction. Nevertheless, the study underscores the feasibility of deploying explainable ML models in low-resource settings and highlights the need for data quality improvements, multi-site validation, and incorporation of additional clinical features to improve prediction accuracy.

Table of Contents

Declaration and Approval **ii**

Abstract **iii**

List of Figures **vii**

List of Tables **viii**

1 Introduction **1**

1.1 Major Risk Factors for Preterm Birth 2

1.2 Problem Statement 3

1.3 Research Objectives 5

1.4 Research Questions 5

1.5 Scope of The Work 5

1.6 Significance of the Study 6

2 Literature Review **7**

2.1 Maternal Health Service Provision in Kenya 7

2.2 Traditional Methods for Predicting Preterm Birth 9

2.3 Traditional Statistical Methods in Predicting Pre -term Birth 11

2.4 Machine learning applications in healthcare 12

2.5 Machine Learning in Predicting Preterm Birth 14

3 Methodology **20**

3.1 Study Design 20

3.2 Study Setting 21

3.3 Dataset description 21

3.4 Data Preparation 23

3.4.1 Gestation age calculation 23

3.4.2 Data Cleaning 24

3.5 Feature Engineering 24

3.6 Classification Models 25

3.6.1 Logistic Regression 25

3.6.2 Random Forest 25
3.6.3 Extreme Gradient Boosting (XGBoost) 25
3.6.4 Support Vector Machine (SVM) 25
3.6.5 Artificial Neural Networks (ANN) 26
3.6.6 LightGBM (Light Gradient Boosting Machine) 26
3.6.7 CatBoost (Categorical Boosting) 26
3.6.8 AdaBoost (Adaptive Boosting) 26
3.7 Evaluation Metrics 26
3.7.1 Accuracy 27
3.7.2 Area Under the Curve (AUC) 27
3.7.3 Precision 27
3.7.4 Recall (Sensitivity) 27
3.7.5 F1 Score 28

4 Results and Discussion **29**
4.1 Exploratory Data Analysis 29
4.1.1 Summary statistics 29
4.2 Univariate Distribution 30
4.3 Bivariate Analysis 31
4.4 Modelling 33
4.4.1 Splitting The Dataset 33
4.4.2 Handling Class Imbalance 33
4.4.3 Model Performance 34
4.4.4 Hyperparameter Tuning 35
4.4.5 Feature Importance 37
4.5 Shap Values 39
4.6 Boosting Algorithms 40
4.6.1 Hyperparameter Tuning for Boosting Models 41
4.6.2 Feature Importance Analysis of the CatBoost Model 43
4.6.3 Model Explainability Using SHAP Values 45
4.7 Discussion 48
4.7.1 Data Availability and Quality 49
4.7.2 Model Performance 50
4.7.3 Deployment 51

5 Conclusion And Recommendations **52**
5.1 Conclusion 52
5.2 Recommendations 53
5.3 Limitations and Future Directions 53

A Similarity Report **63**

VT OMNES VNVM SINT

B Ethical Clearance Confirmation **66**

C Letter of Introduction.pdf **68**

D Python Link **71**

List of Figures

1.1 Primary care networks model in Kenya 6

2.1 Comprehensive maternal service availability 8
2.2 Comprehensive Maternity Services Readiness 9
2.3 Pillars of machine learning for healthcare services 13

3.1 The CRISP-DM Process 20

4.1 Variable distributions for key maternal and clinical features 31
4.2 Correlation matrix for all variables 32
4.3 Distribution of the outcome variable before and after SMOTE 33
4.4 Model performance comparison 35
4.5 Evaluation metrics after hyperparameter tuning 37
4.6 The feature importance for XGBoost Model 38
4.7 The SHAP Values for XGBoost Model 39
4.8 Confusion Matrix for Boosting Models 42
4.9 Feature Importance for Catboost 44
4.10 SHAP Values for Catboost 46
4.11 Deployment using Streamlit 47

UT OMNES VNVM SINT

List of Tables

4.1 Descriptive Statistics for Clinical and Demographic Variables 29
4.2 Performance Metrics of Machine Learning Models for Preterm Birth Prediction 34
4.3 Performance Metrics of Machine Learning Models 36
4.4 Performance Metrics of Gradient Boosting Models 41
4.5 Updated Performance Metrics of Gradient Boosting Models 42

VT OMNES VNVM SINT

Abbreviations

ANC	Antenatal Care
AUC-ROC	Area Under the Receiver Operating Characteristic Curve
BEmONC	Basic Emergency Obstetric and Neonatal Care
CEmONC	Comprehensive Emergency Obstetric and Newborn Care
EHRs	Electronic Health Records
GA	Gestational Age
KDHS	Kenya Demographic and Health Survey
KEPH	Kenya Essential Package for Health
ML	Machine Learning
MOH	Ministry of Health
PCNs	Primary Health Care Networks
PTB	Preterm Birth
pPROM	Preterm Premature Rupture of Membranes
WHO	World Health Organization

Acknowledgements

First and foremost, I would like to express my sincere gratitude to my supervisor, Dr. Betsy Muriithi, for her continuous support, valuable guidance, and encouragement throughout the course of this research. Her insights and constructive feedback greatly enriched the quality of this work. I would also like to thank the faculty and staff of the SIMS - iLab, Strathmore University, for providing a conducive environment for research and learning. I am deeply grateful to my family for their unconditional love, patience, and understanding. Their unwavering belief in me has been a constant source of motivation.

VT OMNES VNVM SINT

Dedication

Dedicated to my family,
whose love and support made this journey possible.

“I can do things you cannot, you can do things I cannot; together we can do great things.”
– — Mother Teresa.”

Chapter 1: Introduction

Preterm birth (PTB), defined as delivery before 37 weeks of gestation, remains a critical public health challenge globally and in Kenya (Menon, 2012)(Wagura et al., 2018). Each year, around 130 million infants are born globally, and approximately 15 million are born pre-term (Wagura et al., 2018). The World Health Organization estimates approximately 13.4 million preterm births occurred worldwide in 2020, representing 10% of all live births (Zeng et al., 2024).

In sub-Saharan Africa, preterm birth rates are disproportionately high, with Kenya facing a dual burden of elevated prevalence and significant neonatal mortality linked to prematurity (Ajayi et al., 2025). Prematurity is not only a leading cause of death and illness in newborns but also poses long-term health risks that can extend into childhood and adulthood. Due to its impact on child survival, preterm birth significantly hinders progress toward international health goals, such as the Millennium Development Goal 4 (MDG-4) focused on reducing child mortality (Wagura et al., 2018).

National data from the 2022 Kenya Demographic and Health Survey (KDHS) reports a 7.14% prevalence of PTB (Ajayi et al., 2025) , though hospital-based studies between 2017 and 2021 documented higher rates of 15.9%–20.2% (Ajayi et al., 2025)(Wagura et al., 2018), reflecting disparities in healthcare access and regional risk factors. Preterm complications account for 37% of neonatal deaths in Kenya, contributing to a neonatal mortality rate of 21 deaths per 1,000 live births and representing 51% of under-five mortality (Anyango, 2024)(Business, 2023). The survival outcomes for preterm infants differ widely depending on geographic location and healthcare access. In low-resource settings, particularly in Africa, the likelihood of dying from complications related to prematurity is estimated to be at least twelve times higher than in high-income countries (Wagura et al., 2018). This burden is attributed to these regions where healthcare infrastructure often struggles to provide adequate antenatal care, and access to advanced diagnostic tools is limited (Karugu et al., 2023).

Globally, the prevalence of preterm birth ranges from 5% to 18% across 184 nations. The burden is heaviest in Sub-Saharan Africa and parts of Asia, where the majority of global births occur. These regions also contribute to over 60% of the world's preterm births and account for more than 80% of related neonatal deaths (Eyinda et al., 2024)(Wagura et al., 2018). Preterm birth (PTB) is not a uniform condition but a complex syndrome with varied underlying causes and is typically classified into three primary phenotypes based on clinical presentation and underlying causes; Medically Indicated (Iatrogenic), Spontaneous (Idiopathic), and Preterm Premature Rupture of Membranes (pPROM) (Menon, 2012). Iatrogenic deliveries occur due to intentional interventions for maternal/foetal health risks. Common indications include preeclampsia, foetal growth restriction, or placental abruption (Gimenez et al., 2016)(Goldenberg et al., 2008). Neonates in this category exhibit higher rates of small-for-gestational-age (SGA) status and metabolic complications like coagulopathy (X. Chen et al., 2021). Iatrogenic preterm birth, including labor induction and caesarean delivery without labor, constitutes about 30–40% of all preterm births, and pre-eclampsia/eclampsia and severe intrauterine growth restriction are the common causes (X. Chen et al., 2021). Spontaneous (Idiopathic) Preterm Birth category represents the largest proportion (Menon, 2012). These are births that occur due to spontaneous labor without any clear medical indication. Contributing factors may include uterine overdistension, infections, inflammation, cervical insufficiency, and psychosocial stress, though in many cases the exact cause remains unknown (Costello et al., 2025)(Goldenberg et al., 2008). Other risk factors include prior preterm delivery, low socioeconomic status, and maternal infections (Gimenez et al., 2016). Finally, PPROM involves membrane rupture before labor onset, often linked to intrauterine infections or inflammation.

Machine learning (ML) is significantly advancing the prediction of preterm birth (PTB), surpassing traditional statistical approaches by effectively analyzing complex datasets and identifying non-linear relationships among risk factors. Recent studies have shown that algorithms like random forests, gradient boosting machines, and deep learning models outperform conventional methods such as logistic regression, particularly in terms of sensitivity and specificity (Teshome et al., 2024). For example, a cohort study in Wenzhou, China, using the CatBoost algorithm, achieved an AUC of 0.70 and an accuracy of 0.81, identifying key predictors such as antenatal visits, maternal weight, and blood pressure (Zhang et al., 2024).

Additional research supports the effectiveness of ML in diverse settings. A study by Kloska et al. (2025) found that a linear support vector machine with boosted parameters performed best, with accuracy and recall over 80%, despite a lower F1-score (Kloska et al., 2025). In Western Australia, Wong et al. (2022) reported that a multi-layer perceptron model identified nearly half of preterm births using routine maternal data (Wong et al., 2022). A study in Sub-Saharan Africa using DHS data and random forests reached an AUC of 0.95, with important predictors including poor antenatal care and home deliveries (Ngusie et al., 2024). In the UK, Narice et al. (2024) used logistic regression to predict spontaneous PTB with good performance, based on demographic and clinical data (Narice et al., 2024). Collectively, these findings highlight the growing potential of ML to enhance early identification and improve maternal and neonatal outcomes.

1.1 Major Risk Factors for Preterm Birth

Infections play a pivotal role, with urinary tract infections (UTIs), syphilis, malaria, and HIV increasing PTB risk. In Kenya, UTIs elevate PTB likelihood by disrupting uterine integrity, while syphilis contributes to inflammation-triggered preterm labor (Eyinda et al., 2024)(Wagura et al., 2018). Across SSA, malaria accounts for 12–15% of PTB cases due to placental inflammation, and HIV is linked to immune dysregulation and intrauterine growth restriction (Mabrouk et al., 2022). Hypertensive disorders, including pregnancy-induced hypertension (PIH) and preeclampsia, independently increase PTB risk by 2.5–3.5 times (Etil et al., 2023)(Wagura et al., 2018). These conditions often necessitate early delivery to prevent maternal complications like eclampsia (Ajayi et al., 2025).

A prior preterm birth is the strongest predictor, doubling recurrence risk due to cervical insufficiency or chronic inflammation (Mabrouk et al., 2022). Multiple pregnancies (twins/triplets) elevate PTB rates to 30–50% in SSA, compared to 8–10% for singletons, driven by uterine overdistension and placental insufficiency (Tekeba et al., 2024)(The Alliance for Maternal and Newborn Health Improvement (AMANHI) GA Study Group, 2022). Short inter-pregnancy intervals (less than 6 months) further exacerbate risks by limiting nutrient replenishment (Mabrouk et al., 2022). Maternal age extremes (below 20 or greater than 35 years) correlate with PTB through pathways like cervical immaturity in adolescents and age-related placental dysfunction (Mabrouk et al., 2022)(Wagura et al., 2018).

Low socioeconomic status amplifies exposure to malnutrition, limited ANC access, and occupational stressors employed Kenyan women face 34% higher PTB odds compared to unemployed counterparts (Etil et al., 2023)(Tekeba et al., 2024). Behavioural factors such as alcohol use increase PTB likelihood by 19-fold in SSA, likely due to oxidative stress and foetal alcohol spectrum disorders (Tekeba et al., 2024). Inadequate antenatal care (ANC); fewer than four visits reduce early detection of PTB risks. Kenyan studies show ANC attendance lowers PTB odds by 87% through timely management of infections and hypertension (Etil et al., 2023)(Eyinda et al., 2024). However, disrespectful maternity care and facility understaffing deter service utilization, particularly in rural areas (Mabrouk et al., 2022).

This dissertation is structured as follows. Chapter 2 presents a review of the existing literature, focusing on the application of machine learning in healthcare and, more specifically, its use in predicting preterm birth. Chapter 3 outlines the methodology, including data sources, variable selection, preprocessing techniques, model development, and evaluation strategies. Chapter 4 reports the results of the various machine learning models applied in this study, highlighting their comparative performance. Chapter 5 provides a discussion of the findings in the context of existing literature, emphasizing the implications for clinical practice and data systems. Finally, Chapter 6 concludes the study by summarizing key insights and offering recommendations for policy, practice, and future research.

1.2 Problem Statement

Preterm birth (PTB) remains a significant contributor to neonatal morbidity and mortality worldwide, with the burden disproportionately higher in low-resource settings. Despite advances in clinical risk assessment, current assessment models often rely on limited variables such as previous PTB, maternal age, and obstetric history. These traditional methods are frequently inadequate in capturing the multifactorial and dynamic nature of PTB, leading to suboptimal accuracy in determining high risk pregnancies (Teshome et al., 2024)(W. Khan et al., 2023).

In Kenya, comprehensive maternal healthcare service provision is severely constrained by systemic gaps in service availability and facility preparedness. According to the Ministry of Health (2023), only 40% of health facilities offer maternity services, and among these, just 5% meet the full requirements for comprehensive care including infrastructure, equipment, and skilled human resources. Alarmingly, only 18% of these facilities conduct caesarean sections, 42% offer blood transfusion services, and a mere 16% have CPAP machines for neonatal respiratory support (MOH, 2023).

Findings from the Kenya Harmonized Health Facility Assessment (KHFFA) further show that while the mean availability of Basic Emergency Obstetric and Neonatal Care (BEmONC) tracer items is 63%, only 3% of facilities have all required items available (MOH, 2019). For Comprehensive Emergency Obstetric and Neonatal Care (CEmONC), only 43% of Level 4 hospitals offer caesarean section services, and just 23% meet the minimum criteria for full CEmONC readiness. Additionally, only 1% of hospitals had all essential items in place, and just 21% of those providing comprehensive services had a gynaecologist, while only 63% had a functional maternity theatre (MOH, 2023). These service gaps contribute to delays in the provision of timely and adequate maternal and neonatal care.

In response to this, Kenya is undergoing service delivery reform through the Primary Care Network (PCN) strategy, through strengthening referral across levels of care with maternal health services being a priority. However, the absence of reliable tools at the primary care level to detect PTB risk undermines the success of such referral systems. There is currently no robust system for identifying and prioritizing high-risk pregnancies for early referral within the PCN framework.

Compounding this problem is the poor quality of routinely collected clinical data in these settings. Electronic health records (EHRs) are often incomplete, inconsistent, or unavailable. These data challenges undermine efforts evidence-based decision making in maternal healthcare including utilization of advanced techniques. Moreover, most existing ML models are developed using high-quality datasets from high-income countries, limiting their applicability in resource-constrained environments like Kenya.

There is a pressing need to develop ML models that are robust to this type of data and tailor it to the realities of primary healthcare in LMICs. A context-sensitive ML tool could support frontline health workers in early identification of high-risk pregnancies and facilitate appropriate and timely referrals within PCNs, thus addressing the ”three delays” that contribute to poor maternal and neonatal outcomes: (1) delay in seeking care, (2) delay in reaching appropriate care, and (3) delay in receiving adequate care.

This dissertation addresses these gaps by determining and validating a machine learning-based tool using real-world data from Kenyan health facility. The aim is to support risk stratification and improve referral decisions for high-risk pregnancies, contributing to better outcomes for mothers and newborns.

1.3 Research Objectives

The primary goal of this study is to determine and evaluate machine learning (ML) models for predicting preterm birth (PTB) in the Kenyan healthcare setting, addressing current limitations in data quality and availability. The specific objectives include:

I. To evaluate the quality and completeness of clinical data relevant to preterm birth prediction in selected healthcare facilities in Kenya. This includes assessing data accuracy, consistency, and the extent of missing values for key maternal and neonatal variables.

II. To develop and validate machine learning models for predicting preterm birth using historical clinical data. The process will involve feature selection, model training (e.g., logistic regression, SVM, random forest, XGBoost, neural networks), and comparative evaluation using performance metrics such as AUC, precision, recall, and F1-score.

III. To explore the feasibility and practical considerations for integrating predictive ML models into existing clinical workflows in the Kenyan healthcare context.

1.4 Research Questions

I. What is the quality and completeness of clinical data available for preterm birth prediction in selected healthcare facilities in Kenya? What are the patterns of missingness, inconsistency, and variability in maternal and neonatal health records?

II. How accurately can machine learning models predict preterm birth using historical clinical data from Kenyan healthcare facilities? Which predictive variables are most important, and how do different ML algorithms (e.g., logistic regression, random forests, neural networks) perform in terms of AUC, precision, recall, and F1-score?

III. What are the key considerations in deploying and integrating the best-performing machine learning model for preterm birth prediction into clinical workflows in the Kenyan healthcare system?

1.5 Scope of The Work

The scope of this research encompasses; data acquisition and preparation, identification of significant risk factors and predictive variables for PTB using exploratory data analysis, development and training of multiple machine learning models for PTB prediction and model evaluation. Lastly development of a user-friendly, web-based application for deploying the best-performing machine learning model. The application will allow healthcare providers to input patient data and receive real-time PTB risk predictions.

1.6 Significance of the Study

This research holds significant relevance for several reasons, poised to transform maternal healthcare by enabling the early and accurate identification of pregnancies at high risk for preterm birth (PTB). Recognizing that Kenya is actively implementing Primary Health Care Networks (PCNs) this study will directly support these efforts by facilitating the timely referral of high-risk pregnancies identified within primary care facilities to higher levels of care for specialized management. This proactive approach facilitates the implementation of timely and targeted interventions, ultimately aiming to reduce PTB rates and improve neonatal outcomes.

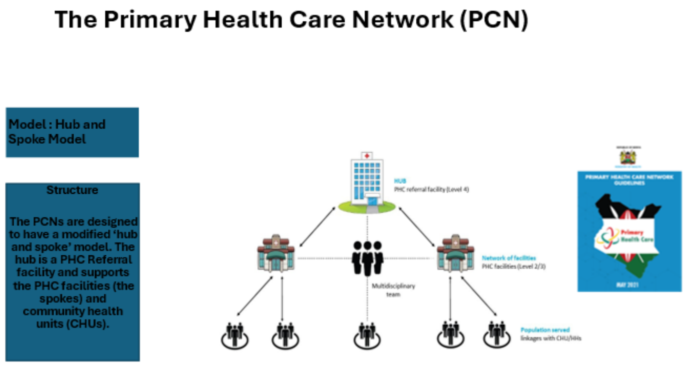

Figure 1.1: Primary care networks model in Kenya

Beyond its immediate clinical impact, the study contributes valuable insights to the burgeoning field of machine learning (ML) in global health. It addresses the critical issue of data limitations prevalent in resource-constrained settings like Kenya, demonstrating how ML solutions can be adapted and optimized within such contexts. Critically, the findings will directly inform the development of a minimum dataset for maternal health in Kenya, ensuring that essential data points are consistently and accurately collected to enhance the effectiveness data driven decisions. Moreover, the research will guide the formulation and enforcement of evidence-based policies designed to enforce collection of minimum datasets, reduce PTB rates and improve overall maternal and neonatal care. By identifying key risk predictors, the study empowers public health strategists and policymakers to create targeted interventions that address the specific challenges faced by pregnant women in Kenya.

Chapter 2: Literature Review

2.1 Maternal Health Service Provision in Kenya

This chapter explores the current state of maternal health service delivery in Kenya, assessing policy efforts, facility readiness, and service availability to support safe childbirth. It also highlights challenges in infrastructure, staffing, and emergency obstetric care that impact maternal and newborn outcomes.

Kenya has implemented policies aimed at enhancing reproductive health services, including improving service availability and readiness. The Global Strategy for Women, Children, and Adolescent Health (2016–2030) emphasizes reducing preventable maternal and newborn deaths, noting that over 15,000 children and 830 women die daily from preventable causes. Alarmingly, when a mother dies in childbirth, her newborn's chance of surviving the first month drops to just 19%. However, timely interventions during labor and immediate newborn care can prevent up to 51% of neonatal deaths (MOH, 2019). The National Reproductive Health Policy outlines operational guidelines and standards for health promotion and service delivery.

Health facility preparedness is a crucial component for delivering quality maternal and newborn care, particularly in minimizing preventable morbidity and mortality. In Kenya, a national health facility census revealed that only 40% of facilities across all levels of the Kenya Essential Package for Health (KEPH) were offering maternity services. However, the readiness to provide comprehensive maternity care remains alarmingly low. Among the facilities providing maternity services, only 5% had all the required equipment, infrastructure, and human resources necessary to offer full-scale comprehensive care. While 18% of these facilities reported conducting caesarean sections, just 42% had blood transfusion capabilities, and only 16% had Continuous Positive Airway Pressure (CPAP) machines for neonatal respiratory support (MOH, 2023).

Similar findings were reported from the Kenya Harmonized Health Facility Assessment (KHHFA), showing a critical shortage of both basic emergency obstetric and neonatal care (BEmONC), and comprehensive emergency obstetric and neonatal care (CEmONC). The mean BEmONC tracer item availability was 63%, but availability of all BEmONC tracers was only 3% nationally (MOH, 2019). These are critical interventions for managing complications during delivery and improving newborn survival. This gap compromises the lives of mothers and babies due to unavailability of life-saving services.

Comprehensive Emergency Obstetric and Newborn Care (CEmONC), which includes critical interventions such as caesarean sections and safe blood transfusion, is a key service expected at the hospital level to manage obstetric emergencies. However, the availability of these services in Kenya, particularly at Level 4 health facilities, remains limited and below expected standards. Specifically, only 43% of Level 4 facilities were found to offer caesarean section services, and 53% had blood transfusion capabilities. More concerning is that only 23% of these facilities met the criteria for providing full CEmONC services (MOH, 2019). This suggests that a significant number of Level 4 facilities may not meet the minimum service delivery standards required for their designation or are struggling to maintain them.

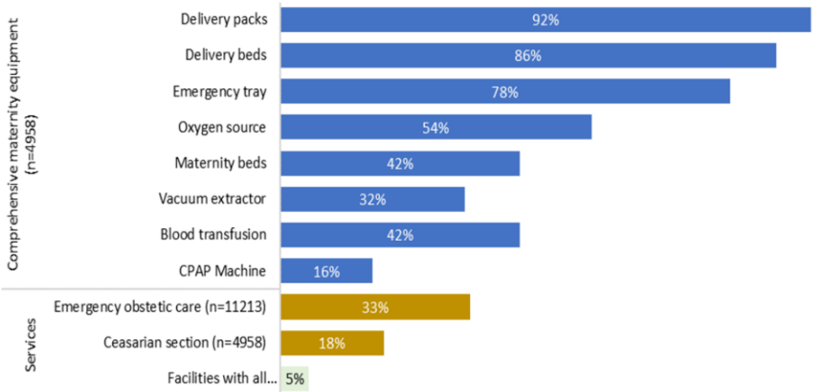

Figure 2.1: Comprehensive maternal service availability

Source: Adapted from Kenya Health Facility Census Report, 2023

According to KHHFA, nationally among hospitals offering caesarean section services, the average availability of essential tracer items required for full CEmONC readiness was just 70%. Alarmingly, only 1% of these hospitals had all the necessary items in place. The situation is further compromised by deficits in human resources and infrastructure. Only 21% of facilities offering comprehensive services had a gynaecologist, and just 63% had a functional maternity theatre. Additionally, essential infrastructure such as postnatal wards and handwashing stations, along with vital medical equipment like delivery beds, vacuum extractors, and oxygen supply systems, were not consistently available (MOH, 2023).

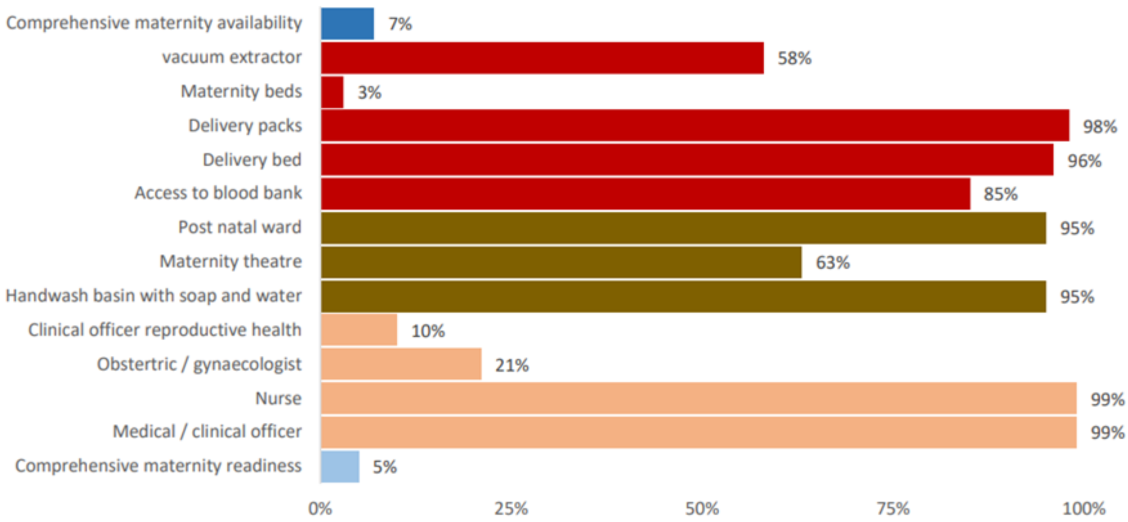

Figure 2.2: Comprehensive Maternity Services Readiness

Source: Adapted from Kenya Health Facility Census Report, 2023

These findings underscore a major gap between the availability of maternity services and their actual capacity to manage obstetric emergencies. The limited readiness of facilities driven by shortfalls in equipment, skilled personnel, and infrastructure compromises the quality and effectiveness of maternal care delivery, particularly in emergency situations where timely intervention can mean the difference between life and death.

2.2 Traditional Methods for Predicting Preterm Birth

Traditional methods for predicting preterm birth (PTB) involve assessing maternal characteristics, ultrasound markers, and biochemical markers (Oskovi Kaplan & Ozgu-Erdinc, 2018). These methods aim to identify women at risk of PTB and implement preventive strategies (Oskovi Kaplan & Ozgu-Erdinc, 2018)(Ramachandran et al., 2024). However, a single screening method with high sensitivity and specificity for PTB prediction remains elusive (Oskovi Kaplan & Ozgu-Erdinc, 2018). Maternal characteristics and obstetric history form a cornerstone of traditional methods for predicting preterm birth (PTB) (Tosto et al., 2023). This approach involves assessing various factors readily available during routine prenatal care to identify women at increased risk (Kassahun et al., 2024).

A history of previous preterm birth is one of the most significant risk factors for subsequent preterm deliveries. Women with a prior preterm birth have a substantially higher risk of delivering preterm again compared to those with a history of uncomplicated term deliveries (Tosto et al., 2023). Extremes of maternal age (younger than 18 or older than 35) are associated with increased PTB risk (He et al., 2018).

Additionally, adolescent pregnancies often carry higher risks due to factors such as inadequate prenatal care and socioeconomic disadvantages. Ultrasound markers play a crucial role in predicting preterm birth (PTB), with transvaginal cervical length (CL) measurement being the most important and widely used technique (Burgos-Artizzu et al., 2021)(Tosto et al., 2023) . CL screening via transvaginal ultrasound is a good predictor of PTB risk in singleton pregnancies. A cervical length of ≤ 25 mm at 24 weeks of gestation is often used as a threshold for increased PTB risk, demonstrating a sensitivity of 37.3% and a specificity of 92.2% (Tosto et al., 2023).

Other ultrasound markers, like the uterocervical angle (UCA) and cervical sliding sign (CSS), have also been investigated to improve PTB prediction. The UCA, which measures the angle between the lower uterine segment and cervical canal, has shown that an obtuse angle ($\geq 95°$ and $\geq 105°$) may be associated with an increased risk of premature birth (CAMEN et al., 2022)(Tosto et al., 2023). While CL remains the standard parameter for predicting PTB, combining it with other parameters like UCA and cervical consistency index (CCI) can significantly increase the prediction rate. However, there is currently no single or combined test with high sensitivity that can definitively identify women who will give birth prematurely (CAMEN et al., 2022).

Biochemical markers have been extensively studied as potential predictors of preterm birth (PTB), aiming to improve early identification of women at risk. These markers are often assessed in cervicovaginal fluids or maternal serum, reflecting underlying pathophysiological processes associated with PTB. Foetal Fibronectin (fFN), present in cervicovaginal secretions between 22 and 37 weeks of gestation, is a widely studied marker. Its presence may indicate a disruption of the maternal-foetal interface and increased risk of preterm labor(Oskovi Kaplan & Ozgu-Erdinc, 2018) . The study further reports that Insulin-Like Growth Factor Binding Protein-1 (IGFBP-1) is another marker for predicting PTB, with bedside tests developed for its detection. Studies suggest that bedside test for IGFBP-1 was more reliable in prediction of PB than foetal fibronectin test. Studies have explored the use of first-trimester biochemical markers such as Pregnancy-Associated Plasma Protein-A (PAPP-A) and free beta-human chorionic gonadotropin (fb-HCG). Low levels of PAPP-A and fb-HCG in the first trimester are associated with an increased risk of preterm birth before 35 weeks (Swiercz et al., 2024).

While traditional methods for predicting preterm birth (PTB) offer valuable insights, they face several limitations regarding cost, specificity, real-time applicability, and scalability for large volumes of patients, especially in resource-limited settings. Many traditional methods involve specialized equipment and trained healthcare professionals, increasing costs, especially in low- and middle-income countries (LMICs). Ultrasound examinations and biomarker tests, while effective, are less practical in resource-limited settings due to the lack of specialized medical equipment and trained healthcare providers. Access to ultrasound imaging remains a significant challenge for pregnant women in Kenya, particularly in rural areas (UNFPA, 2021)(Wanyonyi, 2025). Many women have to travel long distances to access this diagnostic service, typically available only at tertiary-level health facilities in major towns and urban areas (UNFPA, 2021). Availability of biochemical markers services in Kenya is limited to high end private facilities and national hospitals making accessibility to such services (PATH, 2025). In similar settings like Ethiopia, most primary care facilities lack access to advanced laboratory and imaging procedures, making it challenging to estimate PTB risk using these methods (Kassahun et al., 2024).

2.3 Traditional Statistical Methods in Predicting Pre -term Birth

For decades, traditional statistical models have been the backbone of research and clinical decision-making in predicting preterm birth (PTB). These models typically employ logistic regression, Cox proportional hazards models, or risk scoring systems, all of which rely on linear associations between known risk factors and outcomes. While statistically sound and easy to interpret, these methods are increasingly viewed as insufficient for capturing the complex, multifactorial nature of preterm labor.

Traditional statistical methods have been widely used to predict preterm birth (PTB) by identifying significant risk factors and developing prediction models (Arabi Belaghi et al., 2021). These methods involve analysing various maternal and pregnancy-related variables to estimate the probability of PTB. This method is frequently employed to model the relationship between a binary outcome (PTB or not) and a set of predictor variables. It identifies factors significantly associated with PTB and estimates adjusted odds ratios (AOR) to quantify the strength of these associations. Studies have used logistic regression to identify predictors such as diabetes and abnormal pregnancy-associated plasma protein A concentration in the first trimester (Arabi Belaghi et al., 2021).

Risk Scoring Systems is another method used. Risk scores are built by assigning points to clinical features like cervical length, history of PTB, or presence of infections. These scores are easy to implement but tend to oversimplify underlying risk patterns and often ignore interactions between variables.

Lastly the Cox Proportional Hazards Models used mainly for time-to-event analyses (e.g., gestational age at delivery), Cox models allow estimation of how risk factors affect the timing of birth. They require time-structured data and may suffer from the same assumptions as logistic models. Traditional models assume linearity and additive effects between predictors, which often fails to capture the nonlinear and interactive nature of PTB risk factors. For example, the combined effect of maternal infection and stress may be more significant than either factor alone, a relationship that linear models typically overlook (Mitrogiannis et al., 2023). In addition, statistical models generally rely on a narrow set of predefined variables, excluding non-traditional but potentially powerful predictors like environmental exposures, psychosocial stress, or dietary patterns. This restricts model adaptability to different populations or healthcare settings (Rao et al., 2018).

These limitations highlight the need for alternative approaches, such as machine learning models, that can leverage complex data and improve prediction accuracy and resource allocation in diverse healthcare settings.

2.4 Machine learning applications in healthcare

Machine learning (ML) is transforming modern healthcare, offering novel solutions across a wide range of domains from clinical decision support and early disease detection to personalized treatment planning, fraud prevention, and operational optimization. As healthcare systems become increasingly data-driven, the integration of ML is expected to improve both clinical outcomes and system efficiency. Globally, the AI in healthcare market, valued at over $11 billion in 2021, is projected to grow to $187 billion by 2030, reflecting a CAGR of 37% from 2022 to 2030 (Javaid et al., 2022).

ML facilitates individualized treatment plans by analysing patient-specific variables such as lifestyle factors, medical history, symptoms, and genetic data. This personalized approach ensures that medical interventions are tailored to each patient's unique profile, increasing treatment efficacy while minimizing side effects (Rani et al., 2024). It also allows clinicians to anticipate adverse reactions and optimize drug prescriptions accordingly. In addition AI enables a shift from reactive to preventive care by helping healthcare providers predict health risks before they materialize (Javaid et al., 2022)(Zhu et al., 2021) . ML enhances this process by identifying patterns that guide individualized treatment plans, reducing the risk of side effects, and improving treatment efficacy (A. Gupta & Katarya, 2020)(Roth et al., 2018). This application supports the broader goals of precision medicine, especially for complex or chronic illnesses.

ML is instrumental in identifying disease patterns earlier than traditional diagnostic tools. For instance, deep learning models analysing imaging data and EHRs have demonstrated superior sensitivity in detecting cancers, cardiovascular disease, and diabetic retinopathy (Javaid et al., 2020). Machine learning has shown remarkable potential in radiology, especially in processing medical images to detect tumours, classify lesions (e.g., benign vs malignant), and identify abnormalities. ML algorithms such as convolutional neural networks (CNNs) can outperform human radiologists in detecting early disease signs by learning from vast annotated image datasets. These technologies not only improve diagnostic accuracy but also reduce the time required for image analysis, aiding radiologists and robotic systems in making faster and more precise assessments (Islam et al., 2019)(Miotto et al., 2018) .

Healthcare providers are increasingly using ML for forecasting disease risks and identifying critical conditions like strokes, heart attacks, and sepsis. These systems analyse EHRs, real-time patient monitoring data, and even social media and environmental feeds to detect patterns. Predictive tools notify healthcare staff of possible patient deterioration in advance, enabling preventive care instead of reactive treatment (M. Khan et al., 2021)(Shah et al., 2019).

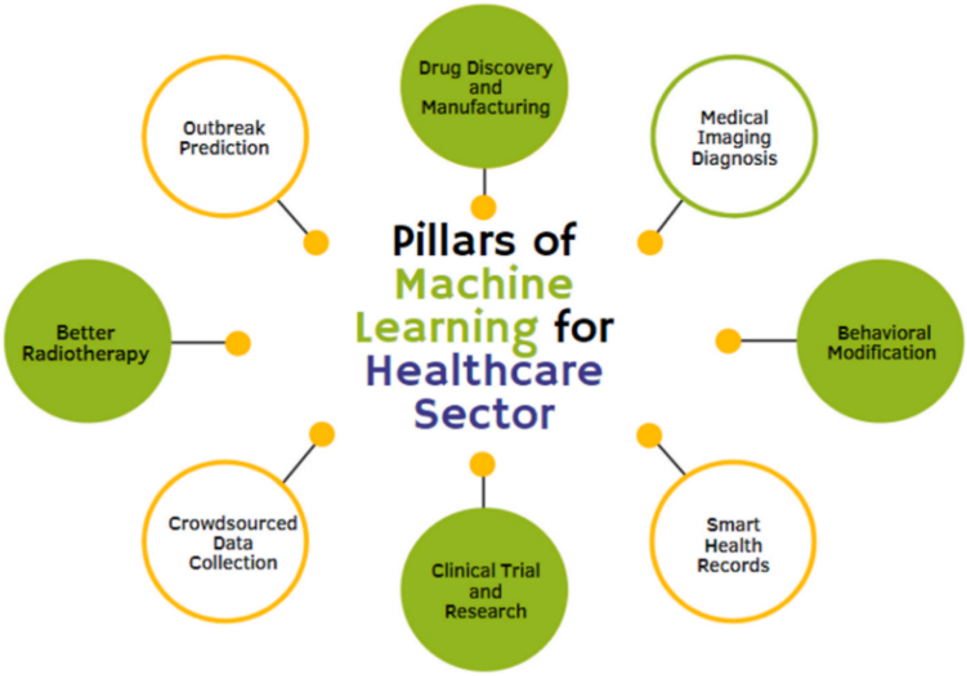

Figure 2.3: Pillars of machine learning for healthcare services

Adapted from: Javaid et al., 2022

ML is widely used in drug development and clinical trials, offering tools to identify ideal candidates for participation, predict therapeutic responses, and monitor safety outcomes. Unsupervised learning models explore raw biological data to identify new treatment pathways and predict success rates. These techniques not only accelerate drug discovery but also reduce errors in trial design and execution (Moyo et al., 2018)(Tong et al., 2020).

In surgical settings, ML plays a pivotal role by enhancing robotic systems with contextual anatomical data, such as tumour location. This assists surgeons with greater precision during procedures. Additionally, AI-enabled platforms handle large datasets across healthcare institutions and deliver human-readable insights that facilitate real-time decision-making in minimally invasive surgeries (S. Gupta & Sedamkar, 2019)(Javaid et al., 2020). ML tools enhance hospital logistics by managing calendars, automating appointment bookings, and processing administrative workflows like billing and rescheduling. These applications streamline service delivery and improve access to care, particularly for facilities handling high patient volumes (McWilliams et al., 2019). ML-assisted systems aim to optimize the patient experience, particularly through intelligent hospital management platforms. These include chatbots, smart assistants, and mobile health apps that help patients with scheduling, treatment guidance, and medical inquiries. In addition, ML helps differentiate profit-driven systems from patient-centered ones by integrating empathy and service personalization (Hu et al., 2020)(Pollettini et al., 2012).

2.5 Machine Learning in Predicting Preterm Birth

Machine learning (ML) is revolutionizing the prediction of preterm birth (PTB), offering significant advancements over traditional statistical methods by effectively harnessing the power of complex datasets and capturing intricate, non-linear relationships between various risk factors. Studies consistently demonstrate the superiority of ML algorithms in identifying pregnancies at high risk of PTB, paving the way for earlier interventions and improved outcomes. A recent review highlighted the effectiveness of algorithms like random forests, gradient boosting machines (GBM), and deep learning models, such as neural networks, in outperforming conventional methods like logistic regression in terms of sensitivity and specificity. Deep learning models, in particular, have shown the potential to achieve the highest predictive accuracy, leveraging key predictors like maternal medical history, cervical length, and pregnancy-related complications (Teshome et al., 2024).

A cohort study conducted in Wenzhou, China, further underscored the potential of ML. Analysing data from 22,603 singleton pregnancies, the study found that the CatBoost algorithm, after 26 weeks' gestation, yielded impressive results with an AUC of 0.70, accuracy of 0.81, sensitivity of 0.47, and specificity of 0.83 (Zhang et al., 2024). This model highlighted the importance of antenatal care visits, maternal weight, blood pressure, and aspartate aminotransferase levels as key predictors. In a recent study evaluating machine learning models for preterm birth prediction, the linear support vector machine (SVM) with boosted parameters outperformed other models, achieving an accuracy of 82%, precision of 83%, and recall of 86%. Although its F1-score was relatively low (48%), it was the most reliable classifier overall. A linear regression model with boosted parameters followed closely with comparable performance (accuracy and recall at 80–82%, F1-score 82%) (Kloska et al., 2025). The study further shows that complex models like XGBoost and CatBoost underperformed due to the small dataset size and simple feature set, which made them prone to overfitting. Simpler models like decision trees and random forests also showed limited effectiveness for similar reasons. Feature importance analysis revealed that C-reactive protein (CRP), haematocrit (HCT), and white blood cell count (WBC) were the most influential predictors, suggesting a strong link between inflammatory and haematological markers and preterm birth risk.

A retrospective cohort study conducted in Western Australia analysed over 81,000 births from 1980 to 2015 to develop machine learning models for predicting preterm birth. The models included regularized logistic regression, decision trees, random forests, extreme gradient boosting, and multi-layer perceptron (MLP). Using routinely collected maternal data such as socio-demographics, medical history, and pregnancy complications the MLP model achieved the best performance. It correctly identified 49.1% of preterm births at a 5% false positive rate. Sensitivity improved to 52.7% in multiparous women when past obstetric history was included, demonstrating the potential of ML to support early and accurate antenatal risk prediction (Wong et al., 2022).

Elsewhere, a recent study conducted across 26 Sub-Saharan African countries aimed to predict adverse birth outcomes such as preterm birth, low birth weight, and stillbirth among 139,659 women using advanced machine learning techniques. Drawing data from the Demographic Health Survey (DHS), the researchers employed ten ML algorithms, with a focus on model interpretability using SHAP values and association rule mining.

VT OMNES VNVM SINT

Summary of ML Studies on Preterm Birth Prediction

Study	Methodology / Models Used	Model Performance	Research Gaps Identified
Abraham et al. (2022)	Logistic regression, random forest, gradient boosting on US EHR data	AUC = 0.75	- Lack of generalizability to LMICs - No interpretability - Limited to billing codes
Sun et al. (2022)	Decision trees, random forest, Naïve Bayes, SVM, KNN on EHR data	Random Forest: AUC = 0.885, Accuracy = 81.6%	- No external validation - No interpretability discussion - Imbalanced data not addressed
Belaghi et al. (2024)	Logistic regression, artificial neural networks (ANNs)	ANN: AUC = 0.688	- Moderate performance - No ensemble model comparison - Limited generalizability
Borboa-Olivares et al. (2023)	Ensemble models using cytokine measurements	Not specified numerically	- Specialized data impractical in low-resource settings - No structured clinical data
Lee et al. (2024)	Tree-based ensemble models on Korean insurance data	Not direct PTB prediction	- Black-box models - Limited to HIC setting - Focused on comorbidities
Zhang et al. (2023)	AdaBoost, logistic regression, random forest on	Accuracy: 100% (non-PTB), 72.7% (PTB)	- Likely overfitting - No SHAP/LIME

VT OMNES VNVM SINT

Among the tested models, the Random Forest classifier performed best, achieving an AUC of 0.95 and an accuracy of 88.0%. Key predictive features included home deliveries, lack of iron supplementation, inadequate antenatal care (fewer than four ANC visits), delivery interval extremes, unwanted pregnancy, first-time motherhood, and residence in West Africa (Ngusie et al., 2024). A UK-based retrospective study conducted at a tertiary obstetric unit between 2018 and 2020 developed a predictive model for spontaneous preterm birth using logistic regression. The model was built from socio-demographic and clinical data collected at booking, involving 917 women with a history of preterm birth and a matched cohort of 100 full-term births. Using seven key features maternal age, BMI, ethnicity, smoking status, gestational type, substance misuse, and parity the model was trained and tested with three-fold cross-validation (Narice et al., 2024). The model achieved an AUC of 0.76, with sensitivity of 0.71 and specificity of 0.78, indicating good predictive performance.

Another retrospective study involving 3,082 pregnant women was conducted to predict the risk of spontaneous preterm birth (sPTB) using clinical and laboratory data. Participants were categorized based on gestational age at delivery (Less than 37 weeks vs greater than or equal to 37 weeks), and five machine learning models were compared. Among the models, the XGBoost algorithm demonstrated the best performance with an AUC of 0.89 (95% CI: 0.88–0.90) during internal testing. It also maintained high predictive accuracy on training (AUC = 0.93), validation (AUC = 0.87), and external testing sets (AUC = 0.79) (Y. Chen et al., 2024). The study demonstrates that the top predictive features identified included alkaline phosphatase (ALP), alpha-fetoprotein (AFP), albumin (ALB), haematocrit (HCT), total cholesterol (TC), diastolic blood pressure (DBP), alanine aminotransferase (ALT), platelet count (PLT), maternal height, and systolic blood pressure (SBP).

The application of machine learning (ML) techniques in predicting preterm birth (PTB) has gained considerable traction in recent years, with various studies exploring different models and datasets to enhance predictive accuracy and clinical utility. Abraham et al. (2022) utilized logistic regression, random forest, and gradient boosting models on electronic health records (EHRs) from the United States and reported an AUC of 0.75 using billing codes. However, the study was limited by its lack of generalizability as there was limited diversity in features (mostly billing codes) which limit applicability to low- and middle-income countries (LMICs) and absence of interpretability mechanisms to support clinical decision-making (Abraham et al., 2022). Similarly, Sun et al. (2022) applied various models including decision trees, random forests, Naïve Bayes, SVM, and KNN on EHR data from 9,550 women. Their best-performing model, Random Forest, achieved an AUC of 0.885 and an accuracy of 81.6%. Nonetheless, the study lacked external validation and did not employ interpretability frameworks or explicitly reported addressing data imbalance (Sun et al., 2022).

Belaghi et al. (2024), working with the Ontario BORN registry, compared logistic regression and artificial neural networks (ANNs), with the ANN model achieving an AUC of 0.688 in first-trimester predictions. The study, however, was limited by moderate model performance and a narrow focus on multiparous women (Belaghi, 2024). In a different approach, Borboa-

Olivares et al. (2023) incorporated cytokine measurements from cervical-vaginal mucus into ensemble ML models, improving PTB prediction. Yet, the reliance on specialized biochemical data reduced the practicality of their approach in routine maternal health settings (Borboa-Olivares et al., 2023).

Lee et al. (2024) examined associations between PTB and comorbidities like temporomandibular and gastrointestinal disorders using ensemble tree-based models applied to Korea's national insurance claims data. While the models demonstrated strong associations, the study did not build direct PTB prediction models and contextual differences for application in LMIC settings (Lee et al., 2024). Zhang et al. (2023) employed AdaBoost, logistic regression, and random forest models on Chinese EHR data, achieving 100% accuracy for non-PTB cases and 72.7% for PTB. Despite strong results, the study likely suffered from overfitting due to imbalanced data (Zhang et al., 2024).

Włodarczyk et al. (2021) conducted a systematic review of ML methods for PTB prediction, emphasizing the potential of ML while pointing out key challenges such as lack of model transparency, inconsistent validation techniques, and limited application in LMICs (Włodarczyk et al., 2021) . Al Ghadban et al. (2023) used metabolomic and clinical data in supervised ML models like SVM and random forests to predict spontaneous PTB. Their study demonstrated strong performance but was constrained by the high cost and limited availability of metabolomic testing in standard care, as well as the absence of integration with facility-level HMIS platforms (Al Ghadban et al., 2024). Begum et al. (2021) developed a model using KNN, decision trees, Naïve Bayes, and SVM to predict preterm births, reporting an exceptionally high accuracy of 99% (Begum et al., 2021).

Machine learning (ML) has emerged as a promising tool for predicting preterm birth (PTB), offering improvements over traditional statistical methods by leveraging complex datasets and capturing non-linear relationships. This literature review synthesizes key studies highlighting the methodologies, evaluation metrics, and findings in ML-based PTB prediction. Studies consistently show that ML models, such as random forests, gradient boosting machines, and deep learning models, outperform traditional methods in predictive accuracy. Deep learning models, particularly transformer algorithms, have demonstrated superior performance in capturing long-term dependencies in patient data. Additionally, the common metrics used to assess model performance include accuracy, sensitivity, specificity, and the area under the receiver operating characteristic curve (AUC). Among the key predictors identified are the maternal medical history, cervical length, and pregnancy-related complications are identified as critical predictors. Other important factors include antenatal care visits, maternal weight, blood pressure, and aspartate aminotransferase levels.

While machine learning models such as random forests, gradient boosting, and neural networks have shown promise in predicting preterm birth, multiple limitations persist. These include poor model interpretability, class imbalance issues, lack of generalizability to LMIC contexts, and minimal consideration for integration into routine health information systems. This study seeks to address these gaps by determining the best ML models using real-world maternal health data from Kenya, incorporating SHAP values for interpretability, handling imbalanced data through appropriate techniques, and focusing on contextual relevance and deployment platform to support actionable clinical decision-making.

Chapter 3: Methodology

3.1 Study Design

This study employed a quantitative, retrospective cohort study design using routinely collected maternal health data from a selected health facility in Kenya. The study included a machine learning (ML) modelling component to develop and evaluate predictive models for preterm birth (PTB). The study design is grounded in the CRISP-DM (Cross-Industry Standard Process for Data Mining) framework. The framework is robust and widely adopted methodology for developing data-driven solutions as it provides a structured approach that ensures consistency, traceability, and iterative improvement in data mining and machine learning projects.

The framework consists of six well-defined phases: Business Understanding, Data Understanding, Data Preparation, Modelling, Evaluation, and Deployment (Chapman, 2000).

Figure 3.1: The CRISP-DM Process
Source: Chapman, 2000

The first phase, business understanding, focused on clarifying the objective of predicting preterm birth using clinical and maternal data to inform early interventions. In the data understanding data sources and description of variables, assessing for completeness, consistency, and patterns within the variables collected. During data preparation, raw data underwent cleaning, transformation, and feature selection to ensure compatibility with machine learning models. The modelling phase involved testing various algorithms to identify the most effective one for preterm birth prediction, with hyperparameter tuning for performance optimization. In the evaluation stage, models were assessed using metrics such as accuracy, AUC, sensitivity, and specificity to determine their reliability in achieving project goals. Finally, the deployment phase considered how the best-performing model could be integrated into clinical workflows, supporting healthcare providers with real-time risk predictions to enhance maternal and neonatal care.

3.2 Study Setting

This study was conducted at a Level 4 private healthcare facility located in Bungoma County, Kenya. The facility operates within a typical low-resource setting, reflecting many of the challenges commonly experienced in public and private healthcare institutions across rural Kenya. It was purposefully selected due to its role as a referral-level facility within the maternal care continuum, providing a full range of antenatal, delivery, and postnatal services to a broad catchment population.

The facility maintains an electronic health records (EHR) system, which was the primary source of data for this study. Despite being a private facility, it faces similar systemic challenges observed in public sector institutions, including intermittent documentation, resource limitations, and infrastructural constraints. These factors made it a suitable representation for evaluating the applicability and robustness of machine learning models in a real-world, low-resource maternal health context.

UT OMNES VNVM SINT

3.3 Dataset description

The dataset used in this study was sourced from a Level 4 public health facility located in Bungoma County, Kenya. The data were obtained from routinely collected hospital records and electronic health systems, focusing on pregnant women who delivered at the facility for deliveries between 2023 and 2025. The dataset comprises 3020 individual data points, each representing a unique maternal health case. For each case, a total of 21 variables were recorded. The data were retrieved from clinical records maintained in both electronic and manual formats within the facility. These variables included a mix of socio-demographic, clinical, and obstetric features.

Specifically, the dataset included date of admission, maternal age, religion, gravida, Abortions history, LMP (Last Menstrual Period), EDD (Estimated Due Date), blood pressure (systolic and diastolic), haemoglobin levels, history of preterm birth, fundal height, pulse, engagrment, foetal heart, dilation, temperature, respiratory rate, date of delivery and delivery type. Below is a description of each variable:

I. **Date of Admission:** The specific date when the patient was admitted to the health facility for antenatal care or delivery. Used for timeline referencing and gestational tracking.

II. **Maternal Age:** The age of the pregnant woman in years. A critical risk factor as both adolescent and advanced maternal age are associated with adverse birth outcomes.

III. **Religion:** Self-reported religious affiliation. Included as a socio-demographic characteristic potentially influencing health-seeking behaviors.

IV. **Gravida:** The total number of times a woman has been pregnant, including the current pregnancy. An indicator of reproductive history and maternal experience.

V. **Abortions History:** The number of previous spontaneous or induced abortions. Used to assess risks related to cervical insufficiency or uterine abnormalities.

VI. **LMP (Last Menstrual Period):** The first day of the woman's last menstrual cycle, used to estimate gestational age and calculate the EDD.

VII. **EDD (Estimated Due Date):** The expected date of delivery, derived from LMP or ultrasound scan. Crucial for determining whether a birth is preterm (¡37 weeks).

VIII. **Blood Pressure (Systolic and Diastolic):** Measured in mmHg. Hypertensive disorders such as preeclampsia are strong predictors of preterm birth.

IX. **Haemoglobin Levels:** Measured in g/dL, indicating maternal anaemia status, which may increase the risk of adverse pregnancy outcomes.

X. **History of Preterm Birth:** Binary variable (Yes/No) indicating whether the woman has previously delivered preterm. A strong predictor in risk models.

XI. **Fundal Height:** The measurement (in cm) from the top of the pubic bone to the top of the uterus. Helps estimate gestational age and foetal growth.

XII. **Pulse:** Maternal pulse rate in beats per minute (bpm), important for monitoring cardiovascular status.

XIII. **Engagement:** Refers to the descent of the foetal head into the maternal pelvis. Indicates readiness for labor and delivery stage.

XIV. **Foetal Heart Rate (FHR):** Measured in bpm, an indicator of foetal well-being and distress during labor.

XV. **Dilation:** Cervical dilation in centimetres during labor, ranging from 0 to 10 cm. Used to track labor progression.

XVI. **Temperature:** Maternal body temperature in degrees Celsius. Febrile states may be indicative of infection risk, which can contribute to preterm labor.

XVII. **Respiratory Rate:** Breaths per minute, used to assess maternal respiratory function and identify distress or infection.

XVIII. **Date of Delivery:** Actual date when delivery occurred, used with LMP/EDD to determine gestational duration.

XIX. **Delivery Type:** Categorized as spontaneous vaginal delivery, caesarean section, or assisted delivery (e.g., vacuum). This outcome variable helps define the clinical context of preterm birth.

However, after cleaning and calculation of the variable of interest, the data shape changed to 15x2291 as datapoints that did not allow for calculation of gestation age (due to both LMP and EDD missing) were dropped. The tools used were excel and python. Excel was used in the initial stages of data preparation while python was utilized to finalize the data preparation, pre-processing, exploratory data analytics, model building and evaluation.

3.4 Data Preparation

The dataset underwent a series of pre-processing steps to ensure data consistency, completeness, and readiness for modeling. These steps included gestational age calculation, missing value handling, and column renaming, checking and removing duplicates and outlier curation.

3.4.1 Gestation age calculation

Gestational age (GA) is a vital indicator of pregnancy progression and foetal maturity. In this study, GA was derived through a structured two-step approach, depending on the availability of either the Last Menstrual Period (LMP) or the Expected Due Date (EDD). Where both the LMP and the delivery date were reliably recorded, gestational age was calculated using the standard obstetric definition.

Gestation Age (days) = Date of delivery – LMP

Then the gestation age in weeks was calculated using the formula:

Gestation age (weeks) = [Gestation age (days) / 7]

This calculation aligns with international clinical guidelines, where the duration of pregnancy is traditionally counted from the first day of the last menstrual period, not from the date of conception. In the cases where the LMP was missing, perceived inaccurate from the gestational age outcome, or subject to recall bias the EDD was utilized. The Expected Due Date (EDD) typically derived from first-trimester ultrasound the LMP was back-estimated using the standard gestational length of 280 days (40 weeks):

Estimated LMP = EDD – 280 Days

This estimated LMP was then used in the same formula as above to calculate gestational age. This method is widely accepted and used in obstetric practice, especially when EDD is confirmed via early ultrasound, which is considered more accurate than LMP after the first trimester. Its a government policy to have at least two ultrasound checks for the entire period of pregnancy making this a reliable way of estimating the gestation age. A challenge remains access to ultrasound services. This data manipulation was done in excel.

3.4.2 Data Cleaning

To ensure the reliability and integrity of the dataset used for machine learning model development, extensive data preprocessing steps were undertaken. These included handling missing data, removal of duplicates, and outlier detection and curation.

Missing values were present across several clinical variables. These were addressed using univariate imputation methods, tailored to the nature of each variable. For variables such as age, fundal height, systolic pressure, diastolic pressure, pulse, dilation, temperature, haemoglobin concentration, respiratory rate and foetal heart rate missing values were imputed using the median of each respective column. Median imputation is robust to outliers and preserves central tendency. Variables such as gravida, abortions, stillbirths, and engagrment, missing values were imputed using the mode (most frequently occurring value). This method is appropriate for variables with a limited number of unique values.

Duplicate records were identified using the *.duplicated() method*. A total of 8 duplicate entries were found in the dataset. These were removed using the *.drop_duplicates(keep='last') function*. This ensures only the most recent entry for each duplicated observation retained.

Outliers in the continuous clinical variables were identified and addressed using the Interquartile Range (IQR) method. The variables assessed for outliers included fundal height, systolic blood pressure, diastolic blood pressure, body temperature, haemoglobin level, and respiratory rate.

For each of these variables, the first quartile (Q1) and the third quartile (Q3) were calculated to determine the spread of the central 50% of the data. The interquartile range was then computed as the difference between Q3 and Q1. Outlier thresholds were defined as values falling below Q1 minus 1.5 times the IQR, or above Q3 plus 1.5 times the IQR. Data points that lay outside these bounds were classified as outliers. To reduce their potential influence on the machine learning models, these extreme values were replaced with the median value of the respective variable. This approach ensured that the overall distribution of the data remained stable while minimizing the distortion that could arise from highly skewed or anomalous observations.

3.5 Feature Engineering

The outcome variable whether preterm or not was calculated from gestational age, where the gestational age was less than 37 weeks, it was considered preterm.

3.6 Classification Models

This study utilized five machine learning models to predict preterm birth: Random Forest, XGBoost (Gradient Boosting), Logistic Regression, Support Vector Machine (SVM) and Neural networks as base models. The evaluation metrics from these evaluation prompted exploration of the boosting models including LightGBM, Adaboost, and Catboost. Each model was selected for its unique approach to classification and predictive modelling, providing a comprehensive comparison of their effectiveness.

3.6.1 Logistic Regression

Logistic Regression serves as a baseline model for binary classification tasks and is widely used in medical research due to its simplicity and interpretability. It estimates the probability of an outcome based on a linear combination of input features. Although it assumes linearity between independent variables and the log-odds of the outcome, it provides a useful reference point for evaluating the performance of more complex models.

3.6.2 Random Forest

Random Forest is an ensemble learning method that constructs multiple decision trees and combines their outputs to improve generalization and reduce overfitting. It is particularly robust to noise, handles nonlinear relationships effectively, and can automatically rank feature importance, making it a popular choice in clinical applications. It has been shown to achieve high predictive accuracy in preterm birth prediction by leveraging ensemble learning techniques

3.6.3 Extreme Gradient Boosting (XGBoost)

XGBoost is a scalable and efficient implementation of gradient boosting machines. It sequentially builds an ensemble of decision trees by minimizing a regularized loss function, allowing it to handle complex interactions between features. XGBoost is well-suited for imbalanced datasets and offers strong predictive performance through optimized tree learning and regularization. Given its balance of accuracy, speed, and interpretability, it was a strong candidate for final model selection. XGBoost has demonstrated superior performance in predicting preterm birth by effectively capturing complex interactions between variables

3.6.4 Support Vector Machine (SVM)

SVM is a powerful classification algorithm that finds the optimal hyperplane to separate classes in a high-dimensional feature space. With the use of kernel functions (in this study, a radial basis function kernel), SVM can capture nonlinear relationships between features. However, SVM may be sensitive to imbalanced class distributions and requires careful parameter tuning.

3.6.5 Artificial Neural Networks (ANN)

A shallow feedforward neural network was implemented to explore the capabilities of deep learning techniques in modelling complex, non-linear relationships in the data. While ANNs are capable of capturing intricate patterns, they require larger datasets, longer training times, and are generally less interpretable compared to tree-based models.

3.6.6 LightGBM (Light Gradient Boosting Machine)

LightGBM is a gradient boosting framework developed by Microsoft that is designed to be highly efficient and scalable. It builds trees using a leaf-wise growth strategy, which often leads to better accuracy compared to the level-wise approach used by other algorithms. LightGBM is optimized for speed and memory usage and handles large volumes of data more efficiently by using histogram-based algorithms for binning continuous features. In this study, LightGBM was evaluated alongside XGBoost to test its suitability for handling the maternal health dataset.

3.6.7 CatBoost (Categorical Boosting)

CatBoost, developed by Yandex, is another gradient boosting algorithm specifically designed to handle categorical features efficiently without requiring extensive preprocessing (like one-hot encoding). It uses ordered boosting and symmetric trees to reduce prediction shift and overfitting, two common problems in traditional gradient boosting models. CatBoost was included in this analysis to evaluate its performance on clinical data where many predictors (e.g., gravidity, parity, history of stillbirths) may be naturally categorical or ordinal. Its ability to work with raw feature formats and produce stable results across folds made it a valuable addition to the ensemble.

VT OMNES VNVM SINT

3.6.8 AdaBoost (Adaptive Boosting)

AdaBoost is a foundational ensemble algorithm that combines multiple weak classifiers, typically shallow decision trees (decision stumps), in a sequential manner. Unlike gradient boosting, which minimizes loss through gradient descent, AdaBoost adjusts the weights of training samples based on their classification accuracy emphasizing incorrectly classified instances in subsequent rounds. Although AdaBoost is simpler than its gradient boosting counterparts, its inclusion in this study served as a comparative benchmark.

3.7 Evaluation Metrics

The performance of each machine learning model in predicting preterm birth was evaluated using a suite of metrics designed to provide a comprehensive understanding of their predictive capabilities. These metrics include accuracy, precision, recall, F1 score, and the area under the receiver operating characteristic curve (AUC).

3.7.1 Accuracy

Accuracy is the proportion of all correct predictions (true positives + true negatives) out of the total number of predictions. Accuracy is a general indicator of performance but may be misleading in imbalanced datasets, such as preterm birth prediction, where term deliveries significantly outnumber preterm cases. It indicates how well the model can correctly classify both preterm and non-preterm births.

3.7.2 Area Under the Curve (AUC)

AUC refers to the area under the Receiver Operating Characteristic (ROC) curve, which plots the True Positive Rate (Recall) against the False Positive Rate at various thresholds. AUC represents the ability of the model to distinguish between classes, providing a measure of the model's discriminative ability. It ranges from 0 to 1, where a higher AUC indicates better performance. An AUC of 0.5 suggests performance no better than random chance, while an AUC of 1 indicates perfect discrimination. Critical for evaluating the discrimination ability in preterm birth models, as it assesses the model's capability to rank high-risk pregnancies higher than low-risk pregnancies.

3.7.3 Precision

Precision quantifies the proportion of true positives among all instances predicted as positive.

Precision = True Positives / (True Positives + False Positives)

It assesses the model's ability to avoid false positives. In the context of preterm birth prediction, it indicates the proportion of pregnancies identified as preterm that are actually preterm. High precision is crucial to minimize unnecessary interventions and anxiety for expectant mothers.

3.7.4 Recall (Sensitivity)

Recall measures the proportion of true positives among all actual positive instances.

Recall = True Positives / (True Positives + False Negatives)

It evaluates the model's ability to capture all positive instances. In preterm birth prediction, it reflects the proportion of actual preterm births that are correctly identified by the model. High recall is essential to ensure that high-risk cases are not missed, allowing for timely medical interventions. Recall is of paramount importance in PTB prediction. Missing a high-risk case (false negative) may result in lack of antenatal corticosteroids, delayed referral, or inadequate neonatal preparation increasing the risk of neonatal morbidity and mortality. Thus, recall is a key metric when prioritizing patient safety.

3.7.5 F1 Score

The F1 Score is the harmonic mean of precision and recall. F1 balances the trade-off between false positives and false negatives. It is particularly valuable in situations where both types of errors are critical. Given the imbalanced nature of PTB datasets, the F1 Score serves as a robust summary statistic. A high F1 Score indicates that the model maintains both a low false positive and low false negative rate, making it an ideal metric for balanced model selection.

VT OMNES VNVM SINT

Chapter 4: Results and Discussion

4.1 Exploratory Data Analysis

4.1.1 Summary statistics

The dataset comprised 2,281 maternal cases, and descriptive statistics were calculated for all relevant clinical and demographic variables to characterize the study population and guide subsequent modelling. The mean maternal age was 25.9 years (SD = 6.12), with a range from 14 to 45 years. The median gravidity was 2, with an interquartile range (IQR) of 1 to 4, indicating that most women had experienced multiple pregnancies. The frequency of abortions and stillbirths was low, with respective means of 0.15 and 0.05, though individual cases experienced up to 6 abortions and 5 stillbirths, suggesting a subset of high-risk obstetric histories.

Table 4.1: Descriptive Statistics for Clinical and Demographic Variables

Variable	**Count**	**Mean**	**Std**	**Min**	**25%**	**50%**	**75%**	**Max**
Age	2281	25.896	6.124	14	21	25	30	45
Gravida	2281	2.688	1.802	0	1	2	4	12
Abortions	2281	0.150	0.493	0	0	0	0	6
Stillbirths	2281	0.051	0.289	0	0	0	0	5
Fundal height	2281	30.510	9.085	4	23	36	38	42
Systolic pressure	2281	113.009	10.390	85	106	115	119	140
Diastolic pressure	2281	70.106	6.685	53	67	70	72	88
Pulse	2281	86.371	12.097	55	77	86	94	145
Engagement	2281	3.174	0.783	0	3	3	3	8
Foetal heart	2281	140.132	4.413	102	140	140	140	184
Dilation	2281	4.829	1.859	1	4	4	6	10
Temperature	2281	36.107	0.118	35.8	36.0	36.1	36.1	36.5
Haemoglobin	2281	11.997	0.913	9.7	11.5	12.0	12.5	14.3
Respiratory rate	2281	20.494	1.167	17	20	20	21	25
Gestational age	2281	36.797	4.686	12.143	34.86	38.0	40.0	44.86

Measurements of fundal height, a proxy for foetal growth and gestational progression, had a mean of 30.5 cm (SD = 9.08), with a median of 36 cm. This aligns with late second to early third trimester gestational ages, though the wide range (4 to 42 cm) reflects variability in the timing of measurement or potential growth deviations. Maternal blood pressure readings were within expected antenatal ranges: systolic pressure averaged 113.0 mmHg (range: 85–140 mmHg) and diastolic pressure averaged 70.1 mmHg (range: 53–88 mmHg), suggesting a generally normotensive population. Similarly, pulse rate was normally distributed with a mean of 86.4 bpm and a range extending from 55 to 145 bpm. In terms of foetal assessment, foetal heart rate had a tightly centered mean of 140 bpm (SD = 4.41), consistent with normal foetal well-being. Engagement and cervical dilation values, while low in variance (mean dilation = 4.83 cm), suggest that the dataset included women in early to active labor stages. Maternal temperature was stable across the cohort, with a mean of 36.1°C (SD = 0.12), indicating an absence of systemic febrile illness in the majority of cases. The average haemoglobin (hb) level was 11.99 g/dL (range: 9.7–14.3), falling within normal antenatal ranges but still highlighting the need to monitor for anaemia, especially in low-resource settings. Respiratory rates also fell within expected ranges, with a mean of 20.5 breaths per minute. Finally, gestational age (ga) at the time of data collection averaged 36.8 weeks (SD = 4.69), with a widespread from approximately 12 to 45 weeks. This confirms the dataset includes both preterm and term cases, making it suitable for classification modelling aimed at predicting preterm birth.

4.2 Univariate Distribution

The distributional characteristics of key maternal and foetal variables were examined using histogram plots with overlaid density curves. These visualizations provide critical insight into the shape, central tendency, and spread of the data, as well as potential anomalies.

The age distribution of mothers was right-skewed, with most women concentrated between 20 and 30 years. Gravida, abortions, and stillbirths also exhibited right-skewed patterns, indicating that while a majority of mothers had limited reproductive history, a small proportion had high-parity or adverse outcomes such as abortions and stillbirths. Fundal height appeared approximately normally distributed but showed a slight tail, potentially reflecting clinical variability in foetal size or gestational age. Notably, systolic blood pressure had a highly skewed distribution with extreme outliers beyond 500 mmHg, suggesting potential measurement errors or data entry anomalies. These anomalies were later curated using interquartile range. In contrast, diastolic pressure was normally distributed, reflecting more physiological stability.

Pulse, foetal heart rate, and engagrment measurements were centered around expected clinical norms, with foetal heart rate sharply peaked at 140 bpm, indicating uniformity in measurement. Cervical dilation showed a moderate skew toward lower values, consistent with early labor presentations in a subset of the sample. Temperature, haemoglobin (hb), and respiratory rate were all normally distributed, with tightly clustered values indicating consistency in measurement. A notable outlier in temperature distribution below 30°C may warrant inspection, as it falls below physiological viability.

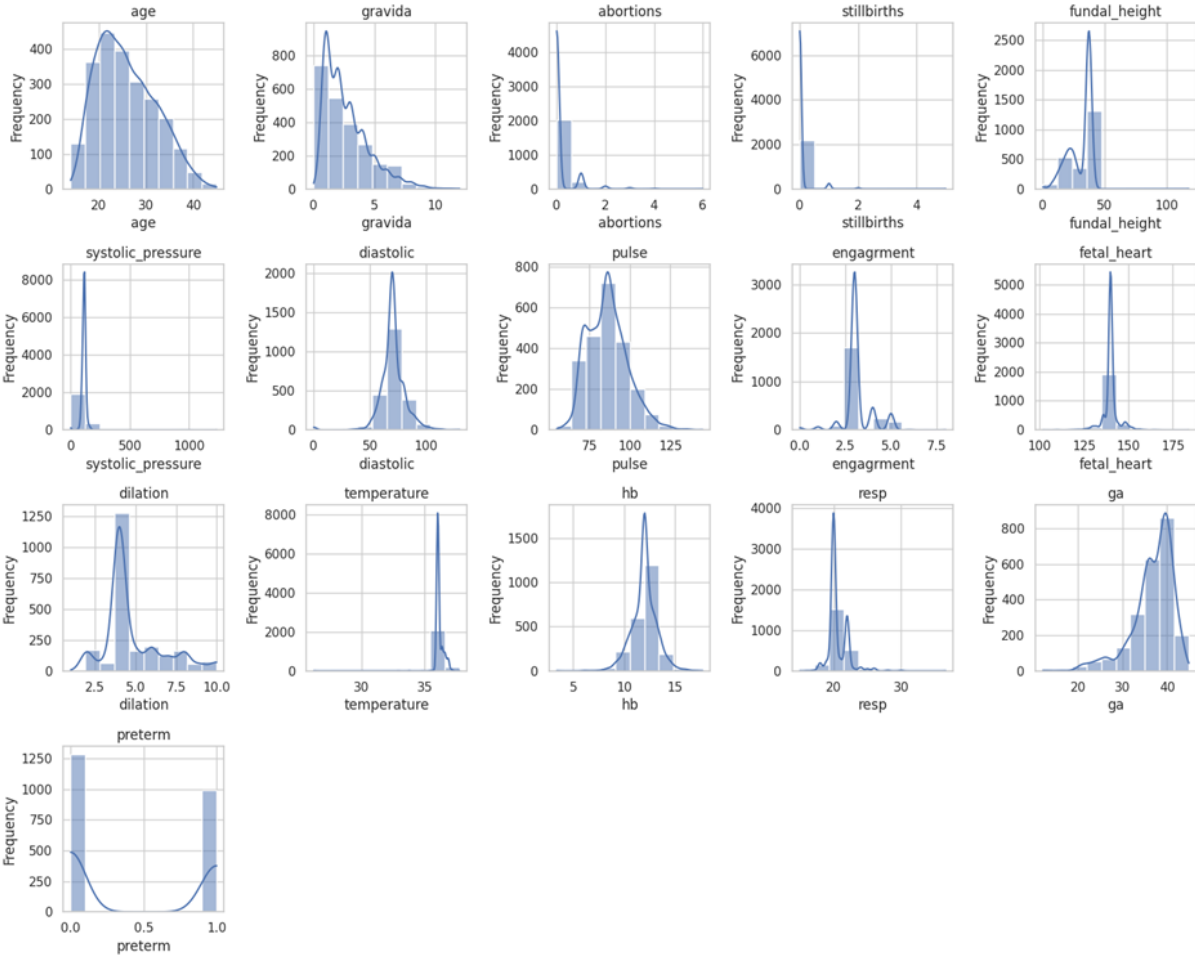

Figure 4.1: Variable distributions for key maternal and clinical features
Source: Author-generated plots from dataset analysis

Gestational age (ga) presented a left-skewed distribution, with a concentration around 38–40 weeks, suggesting that the majority of deliveries occurred at or near term. However, the presence of values below 30 weeks confirms that a subset of the sample represents preterm cases. Finally, the preterm variable, a binary outcome, demonstrated a bimodal distribution reflecting an approximately balanced representation of term and preterm deliveries. This balance is particularly important for supervised learning models aimed at classifying or predicting preterm birth outcomes.

4.3 Bivariate Analysis

The correlation matrix offers a comprehensive view of the pairwise linear relationships among maternal and foetal health variables, as well as their associations with preterm birth. Correlations are expressed as Pearson's r coefficients, where values closer to ±1 indicate stronger linear relationships.

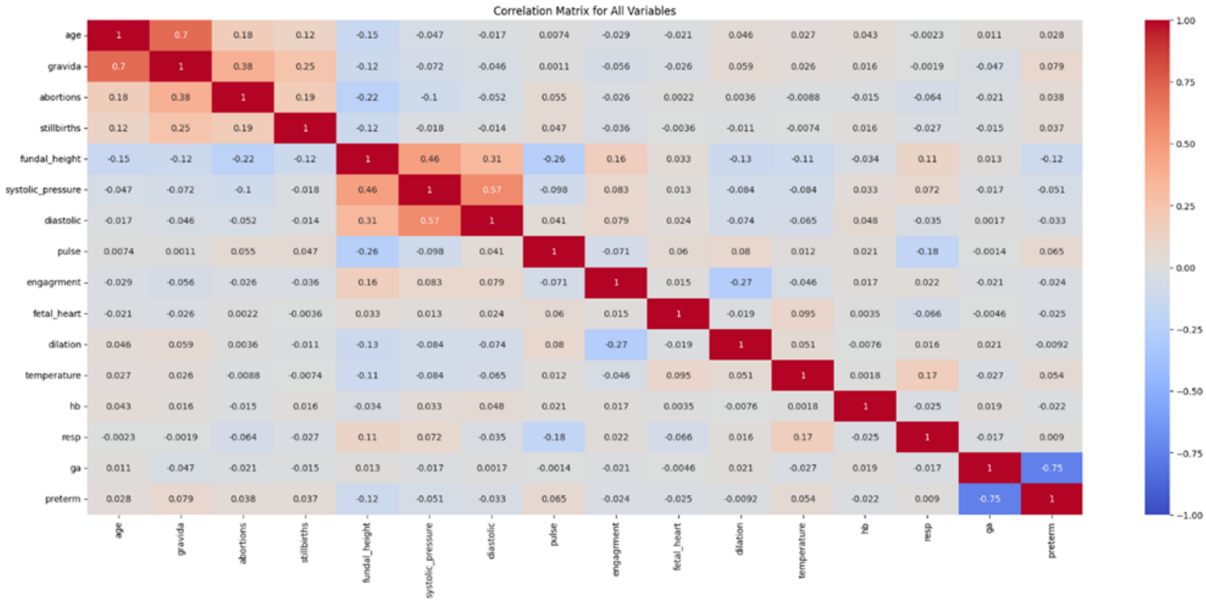

Figure 4.2: Correlation matrix for all variables

Source: Author-generated heatmap from dataset correlation analysis

A strong positive correlation was observed between age and gravida (r = 0.70), suggesting that older women are more likely to have had multiple pregnancies. Similarly, gravida was moderately correlated with abortions (r = 0.38) and stillbirths (r = 0.25), indicating a cumulative effect of reproductive events with increasing parity. These findings align with known reproductive health trajectories where prior obstetric complications can accumulate with successive pregnancies.

VT OMNES VNVM SINT

There were notable positive correlations among several physiological parameters. For instance, systolic and diastolic blood pressure showed a moderate correlation (r = 0.57), consistent with hemodynamic interdependence. Additionally, fundal height was moderately associated with systolic pressure (r = 0.46) and diastolic pressure (r = 0.31), likely reflecting gestational progression and uteroplacental demand. Most other features showed weak or negligible correlations with one another, suggesting that they contribute independently to maternal health profiles. For example, temperature, pulse, haemoglobin, and respiratory rate had low intercorrelations with both each other and the outcome variable.

Of particular interest is the strong negative correlation between gestational age (ga) and preterm birth (r = –0.75*), which is expected since lower gestational age directly defines preterm status. This strong inverse relationship validates the reliability of the dataset and confirms the predictive value of gestational age in classifying preterm outcomes. No other variable demonstrated a strong correlation with the preterm outcome. Variables such as gravida (r = 0.08), abortions (r = 0.04), stillbirths (r = 0.04), and fundal height (r = –0.12) had weak correlations with preterm birth.

4.4 Modelling

4.4.1 Splitting The Dataset

In the development of predictive models, particularly within the medical domain, rigorous dataset partitioning is essential to ensure that model evaluation is both reliable and generalizable. In this study, the complete dataset, consisting of 2,281 records, was randomly divided into training and testing subsets using an 80:20 split ratio, resulting in 1,824 observations in the training set and 457 in the test set. The training set was used to fit various machine learning algorithms and perform hyperparameter optimization. In contrast, the test set was held out as an independent evaluation subset, untouched during model training and resampling. This separation guarantees an unbiased estimate of model performance, simulating how the model would behave when deployed on unseen clinical data.

4.4.2 Handling Class Imbalance

In this study, the binary outcome variable preterm birth was initially imbalanced, as shown in the left panel of Figure below. Prior to resampling, the training dataset comprised 1,028 term (class 0) cases and 796 preterm (class 1) cases, resulting in a moderately imbalanced class distribution. Class imbalance can adversely affect the performance of machine learning models by biasing them toward the majority class, often leading to poor sensitivity for minority outcomes such as preterm birth.

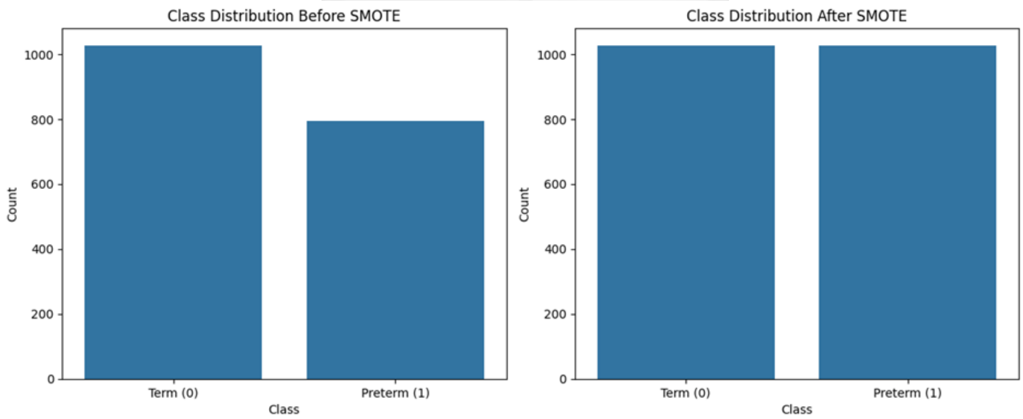

Figure 4.3: Distribution of the outcome variable before and after SMOTE

Source: Author-generated class balance visualization using SMOTE

In order to address this issue, Synthetic Minority Over-sampling Technique (SMOTE) was applied to the training set. SMOTE is an oversampling technique that generates synthetic samples for the minority class by interpolating between existing observations. As illustrated in the right panel of Figure 8 and confirmed by the printed output, SMOTE balanced the classes by increasing the number of preterm cases to match the term cases, resulting in 1,028 instances for each class. This ensures that the learning algorithm receives equal representation from both outcome categories during model training. Importantly, SMOTE was applied only to the training subset, preserving the integrity and representativeness of the test set for unbiased evaluation. This protocol aligns with best practices in machine learning to avoid information leakage and overestimation of model performance.

Balancing the training data through SMOTE is particularly important in clinical applications such as preterm birth prediction, where underrepresentation of adverse outcomes could lead to clinically unsafe decision thresholds. The post-SMOTE class distribution supports more robust training for predictive models by mitigating bias and improving recall on the minority class.

4.4.3 Model Performance

Five classification algorithms Logistic Regression, Random Forest, XGBoost, Artificial Neural Network, and Support Vector Machine (SVM) were trained on SMOTE-balanced data and evaluated on an unseen test set using standard performance metrics. These included accuracy, precision, recall, F1 score, and area under the ROC curve (AUC), each offering insight into the model's predictive performance, particularly in handling class imbalance relevant to preterm birth detection.

Table 4.2: Performance Metrics of Machine Learning Models for Preterm Birth Prediction

Model	**Accuracy**	**Precision**	**Recall**	**F1 Score**	**AUC**
Logistic Regression	0.571	0.508	0.487	0.497	0.555
Random Forest	0.554	0.486	0.452	0.469	0.549
XGBoost	0.575	0.512	0.518	0.515	0.585
ANN	0.554	0.488	0.497	0.493	0.567
SVM	0.582	0.525	0.422	0.468	0.598

Among all classifiers, the Support Vector Machine (SVM) achieved the highest accuracy (0.582) and AUC (0.598), suggesting slightly better overall discriminative ability and probability ranking compared to its counterparts. However, its recall (0.422) was the lowest among all models, indicating that despite better accuracy, it failed to identify a substantial proportion of actual preterm cases. This trade-off between precision and recall reflects the conservative nature of SVM in this context, favouring fewer false positives at the expense of more false negatives.

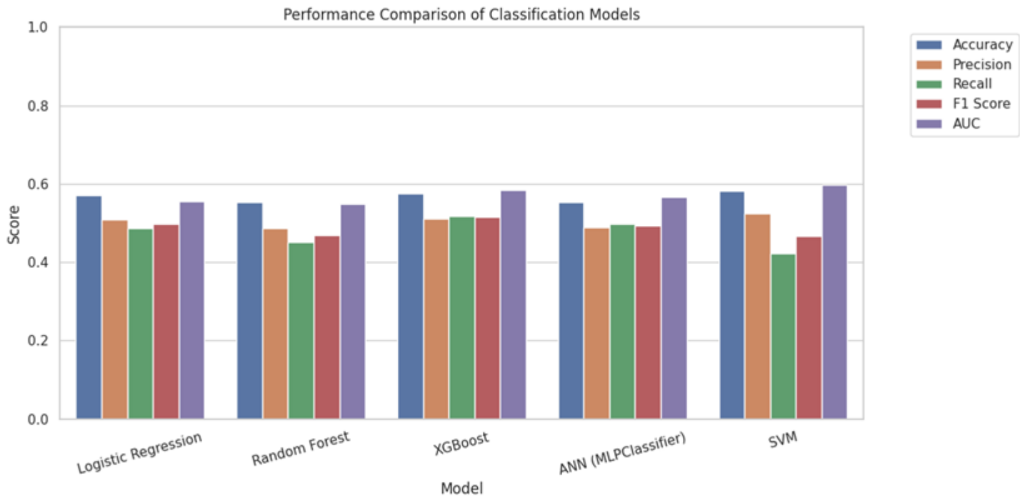

Figure 4.4: Model performance comparison
Source: Author-generated model evaluation comparison plot

The XGBoost classifier emerged as a strong performer with a balanced profile: an accuracy of 0.575, F1 score of 0.515, and AUC of 0.585. Notably, it exhibited the highest recall (0.518) and a competitive precision score (0.512), suggesting it is more effective at identifying true preterm cases while maintaining a tolerable false-positive rate. This makes XGBoost a strong candidate in clinical contexts where early identification of at-risk pregnancies is crucial. The Artificial Neural Network and Logistic Regression yielded similar results, with F1 scores of 0.493 and 0.497, respectively. These models achieved moderate recall and precision, showing acceptable performance but limited superiority over other models. The Random Forest classifier, despite being a powerful ensemble method, underperformed in this dataset, with the lowest F1 score (0.469) and AUC (0.549).

4.4.4 Hyperparameter Tuning

To optimize the predictive performance of the machine learning models, a grid search strategy was employed for hyperparameter tuning using GridSearchCV. For each algorithm, a predefined range of hyperparameters was specified, allowing the grid search to exhaustively explore combinations via cross-validation on the training set.

For the Logistic Regression model, two primary hyperparameters were subjected to tuning to optimize classification performance. The first parameter, C, controls the inverse of the regularization strength. It was explored across the values of 0.01, 0.1, 1, and 10. Lower values of C impose stronger regularization on the model, thereby reducing the magnitude of the coefficients and helping to prevent overfitting, particularly when multicollinearity or high-dimensional data is present. Conversely, higher C values reduce the regularization effect, allowing the model to fit the training data more closely. The second hyperparameter, solver, determines the optimization algorithm used to fit the logistic regression model. Two solvers were evaluated: 'liblinear' and 'lbfgs'. The 'liblinear' solver is particularly well-suited for smaller datasets and supports both L1 and L2 regularization, making it a robust choice when sparsity is desired. On the other hand, 'lbfgs' is a quasi-Newton method that is computationally efficient for larger datasets and supports multiclass classification in a one-vs-rest framework. Evaluating both solvers provided flexibility in balancing model accuracy, interpretability, and computational efficiency.

For the Random Forest classifier, a grid of hyperparameters was explored to optimize model performance. The number of trees in the ensemble, specified by the n_estimators parameter, was tested at values of 100 and 200. This parameter directly affects the model's complexity and ability to generalize. The max_depth parameter, which controls the maximum depth of each decision tree, was tuned using values of None, 10, and 20. A setting of None allows trees to grow without depth restriction, potentially capturing more complex patterns. Additionally, the min_samples_split parameter, determining the minimum number of samples required to split a node, was examined at values of 2 and 5 to prevent overfitting and encourage deeper splits only when necessary.

For the XGBoost classifier, a robust set of parameters was tuned to balance model learning and generalization. The number of boosting iterations, governed by n_estimators, was evaluated at 100 and 200 rounds. The max_depth of individual boosting trees was varied between 3 and 6 to control model complexity. Furthermore, the learning_rate, which dictates the contribution of each tree to the final model, was adjusted across values of 0.01, 0.1, and 0.2. These parameters work in tandem to fine-tune the model's learning behaviour, mitigating the risk of overfitting while maintaining predictive accuracy.

For the Support Vector Machine (SVM), tuning was focused on two key hyperparameters. The regularization parameter C was explored across values of 0.1, 1, and 10, reflecting varying trade-offs between maximizing the margin and minimizing classification error. The kernel parameter, which transforms the input space for optimal class separation, was tested using both 'linear' and 'rbf' (radial basis function) kernels. This allowed the model to learn both linear and non-linear decision boundaries, accommodating more complex patterns within the data.

Table 4.3: Performance Metrics of Machine Learning Models

Model	**Accuracy**	**Precision**	**Recall**	**F1 Score**	**AUC**
Logistic Regression	0.565	0.500	0.457	0.478	0.560
Random Forest	0.586	0.528	0.477	0.501	0.573
XGBoost	0.569	0.504	0.663	0.573	0.607
SVM	0.558	0.492	0.472	0.482	0.568

The Logistic Regression model achieved an accuracy of 0.565, with a precision of 0.500, recall of 0.457, and an F1 score of 0.478. The model AUC was 0.560, indicating a modest discriminative capacity. The best performance was achieved using an inverse regularization strength C = 0.01 and the 'lbfgs' solver. These settings suggest that stronger regularization yielded better generalization for the dataset. The Random Forest classifier produced slightly better overall performance with an accuracy of 0.586 and an F1 score of 0.501. Its AUC reached 0.573, and the model showed balanced precision (0.528) and recall (0.477). The best hyperparameters included n_estimators = 100, max_depth = 20, and min_samples_split = 2, reflecting a relatively deeper tree structure with minimal splitting constraints.

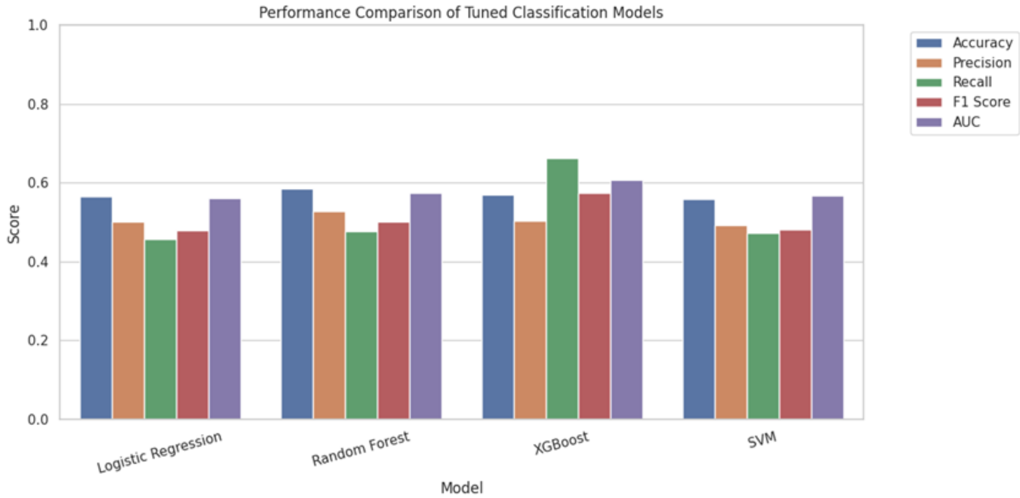

Figure 4.5: Evaluation metrics after hyperparameter tuning

Source: Author-generated evaluation metrics after model tuning

The XGBoost model outperformed other classifiers in most metrics, achieving an F1 score of 0.573 and the highest AUC at 0.607, suggesting superior capability in distinguishing between preterm and term classes. The model's recall (0.663) was notably the highest among all models, though its precision (0.504) was slightly lower than Random Forest. The optimal hyperparameters were n_estimators = 100, max_depth = 3, and learning_rate = 0.01, indicating that a shallow yet cautious boosting approach led to more generalizable learning.

The Support Vector Machine (SVM), with a linear kernel and regularization parameter C = 10, achieved an accuracy of 0.558, F1 score of 0.482, and AUC of 0.568. Although its recall (0.472) and precision (0.492) were moderate, the SVM performed consistently though not surpassing the tree-based models. The XGBoost emerged as the most effective classifier, particularly in recall and AUC, making it well-suited for clinical use cases where identifying preterm risk is critical. The Random Forest model also demonstrated competitive performance, while the ANN struggled under the current configuration, emphasizing the need for additional tuning or architecture revision.

4.4.5 Feature Importance

The XGBoost model being the best performing model was used to identify the most influential predictors contributing to preterm birth classification. Feature importance was calculated using the model's internal gain-based scoring mechanism, which quantifies the relative contribution of each feature to the model's decision-making process. The top ten features, as depicted in the bar plot, highlight a set of clinical and physiological variables that significantly impact predictive outcomes.

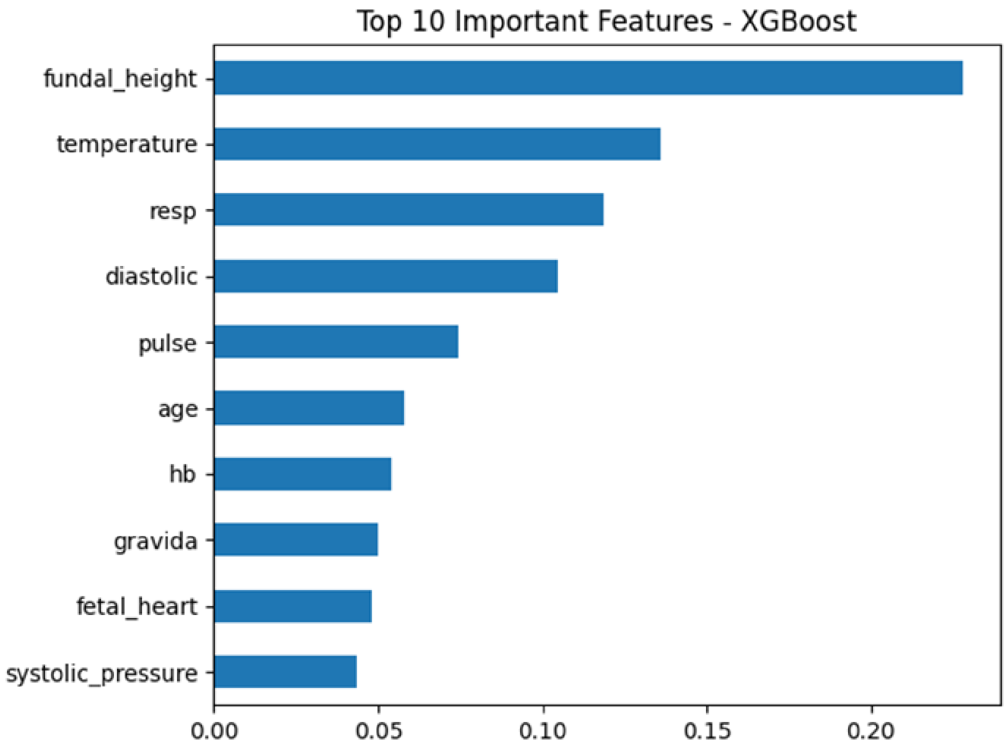

Figure 4.6: The feature importance for XGBoost Model

Source: Author-generated feature importance plot from tuned XGBoost model

The most prominent feature was fundal height, which exhibited the highest importance score among all variables. This finding is clinically intuitive, as abnormal fundal height measurements are often associated with intrauterine growth restriction or macrosomia, both of which may be linked to adverse perinatal outcomes, including preterm labor. The second and third most important features were temperature and respiratory rate (resp), respectively. These are vital signs that can reflect underlying maternal infections, metabolic stress, or systemic inflammation factors known to elevate preterm birth risk.

Diastolic blood pressure also ranked highly, suggesting a possible association between maternal cardiovascular function and preterm labor. Elevated or irregular blood pressure may indicate preeclampsia or gestational hypertension, both of which are established clinical precursors for medically indicated preterm deliveries. The mid-tier importance scores belonged to pulse, maternal age, and haemoglobin (hb). While these features were less influential than the aforementioned variables, they still contributed meaningfully to the model's decision boundaries. Maternal age has long been recognized as a risk factor for adverse birth outcomes, and haemoglobin levels reflect maternal oxygen-carrying capacity, which can impact foetal well-being.

Finally, gravida, foetal heart rate, and systolic blood pressure appeared in the lower ranks of the top ten. Though these variables contributed to the model, their relative influence was modest. Nevertheless, their inclusion supports the multifactorial nature of preterm birth risk, which often arises from subtle interactions among obstetric history, foetal health, and maternal physiology.

4.5 Shap Values

To gain a deeper understanding of how individual features influence model predictions, SHAP (SHapley Additive exPlanations) values were computed for the XGBoost classifier. This method provides both global feature importance (how features affect predictions overall) and local explanations (how features influence individual predictions). The SHAP summary plot below displays both the magnitude and direction of each feature's impact on the model output across all samples, along with the relative value of the feature indicated by colour (red for high values, blue for low).

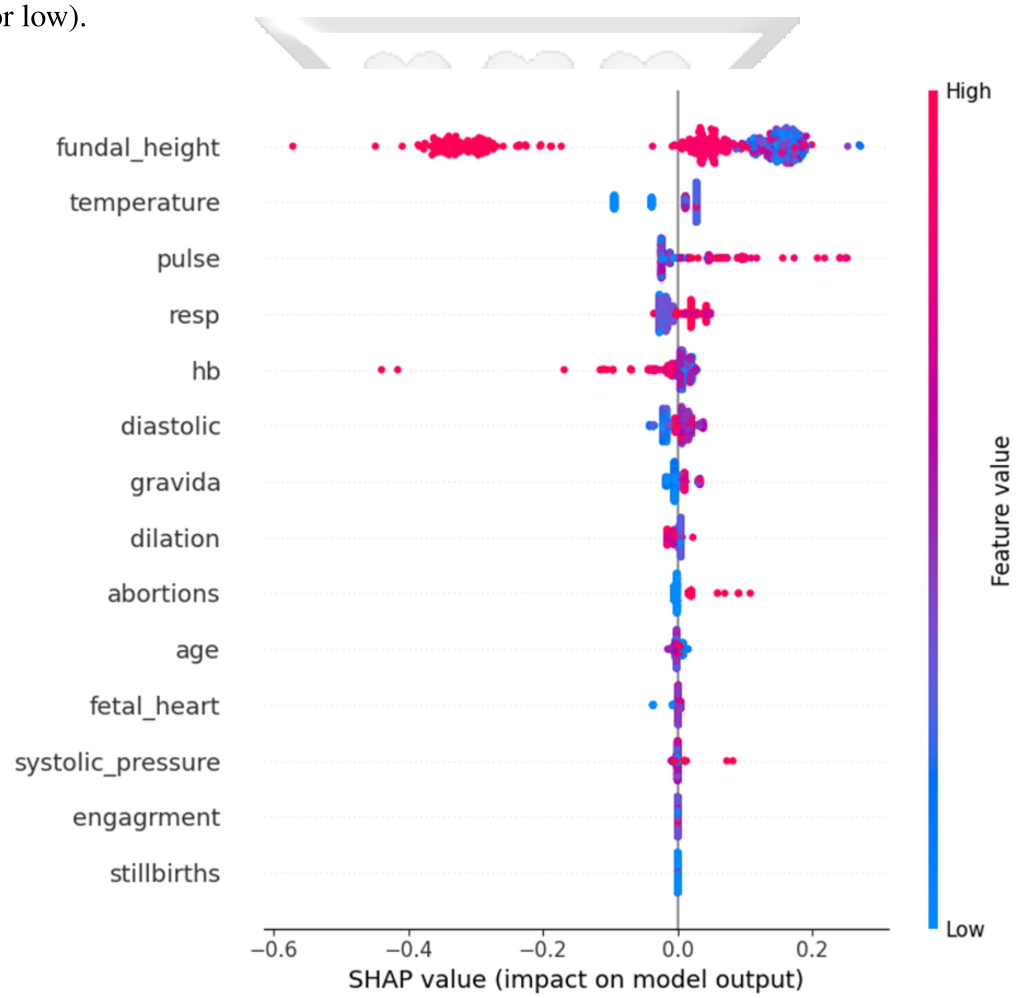

Figure 4.7: The SHAP Values for XGBoost Model
Source: Author-generated SHAP values plot from tuned XGBoost model

The feature fundal height emerged as the most influential predictor, with higher values (in red) generally associated with an increased likelihood of preterm birth. This aligns with clinical observations that abnormal uterine growth, whether excessive or insufficient, may reflect foetal or placental pathologies that trigger premature labor. The concentration and spread of SHAP values for fundal height suggest it consistently plays a major role in the model's decision-making process. Temperature, pulse, and respiratory rate (resp) also ranked highly in terms of impact. Elevated values of these vital signs (red points toward the right) were associated with positive SHAP values, indicating that higher physiological stress or potential maternal infection correlates with a higher predicted risk of preterm birth. These results underscore the importance of monitoring maternal vital signs as dynamic predictors of adverse obstetric outcomes.

Haemoglobin (hb) and diastolic blood pressure showed moderate contributions, with a mix of directions in SHAP values. For example, low hemoglobin levels (blue points with negative SHAP) appear to decrease predicted risk in some cases, which may reflect complex interactions with other clinical features or threshold effects in the data set.

Lower-ranked features such as gravida, dilation, abortions, and age had more modest but still observable effects. Their SHAP distributions were narrower and closer to zero, indicating less consistent influence on the model's output. However, their inclusion suggests subtle interactions and contributions to prediction refinement in certain subgroups. Of note, features like systolic pressure, engagement, and stillbirths contributed minimally, with SHAP values tightly centered around zero. This could imply limited predictive power in this cohort or possible redundancy with more influential covariates. The SHAP summary plot not only confirms the importance of clinically intuitive features like fundal height and vital signs but also provides a nuanced view of feature behaviour including the directionality and individual variability of their impact. Such interpretability is essential for integrating predictive models into clinical workflows, as it allows healthcare providers to understand and trust the rationale behind automated predictions.

4.6 Boosting Algorithms

Given the superior performance of XGBoost in the initial evaluation, the study advanced to a more focused investigation of boosting-based ensemble methods. Boosting algorithms are well-established for their capability to build highly robust, accurate, and interpretable models by sequentially combining multiple weak learners most commonly shallow decision trees into a strong predictive ensemble. Each successive model in the sequence is trained to correct the misclassifications made by its predecessor, making boosting especially effective in high-dimensional and noisy datasets such as clinical records used in this study.

The decision to pursue boosting algorithms was based on their adaptive learning structure, resilience to overfitting, and strong performance in binary classification tasks, particularly when dealing with unbalanced or complex data structures commonly seen in maternal health prediction. To establish a baseline for model performance, four popular boosting algorithms LightGBM, CatBoost, AdaBoost, and XGBoost were trained using their default hyperparameters. Their performance was assessed using five key classification metrics: Accuracy, Precision, Recall, F1 Score, and Area Under the Curve (AUC). These metrics offer a balanced evaluation of each model's predictive ability, particularly in the context of binary classification for preterm birth (PTB) prediction.

Table 4.4: Performance Metrics of Gradient Boosting Models

Model	Accuracy	Precision	Recall	F1 Score	AUC
LightGBM	0.573	0.511	0.477	0.494	0.580
CatBoost	0.584	0.526	0.462	0.492	0.598
AdaBoost	0.582	0.517	0.618	0.563	0.606
XGBoost	0.575	0.512	0.518	0.515	0.585

Among the evaluated models, AdaBoost demonstrated the highest recall (0.618) and F1 score (0.563), indicating a strong ability to correctly identify preterm birth cases. This makes it particularly suitable in clinical settings where sensitivity to true positives is critical i.e., minimizing missed cases of PTB. It also recorded a competitive AUC of 0.606, suggesting a reasonably good trade-off between sensitivity and specificity. Catboost achieved the highest AUC of 0.598 indicating strong overall discriminatory power although its recall was 0.462 which makes it less competitive where missed true positives are important.

XGBoost provided a robust overall performance across all metrics, with an AUC of 0.585, recall of 0.518, and F1 score of 0.515. It displayed more consistent performance than LightGBM and, as seen in prior stages of the study, continued to demonstrate its reliability in predicting complex clinical outcomes, reinforcing its selection as the leading candidate for deeper optimization. LightGBM, while efficient and scalable, showed the lowest scores in most metrics, including recall (0.477) and F1 score (0.494). Its AUC of 0.580 was comparable to XGBoost but lagged behind both CatBoost and AdaBoost, suggesting that further tuning would be required to enhance its predictive utility for this specific task.

4.6.1 Hyperparameter Tuning for Boosting Models

Following initial evaluations, four boosting algorithms LightGBM, CatBoost, AdaBoost, and XGBoost were fine-tuned using cross-validated grid search to optimize hyperparameters and enhance predictive accuracy in preterm birth (PTB) classification. Each model was evaluated using five core metrics: Accuracy, Precision, Recall, F1 Score, and Area Under the Receiver Operating Characteristic Curve (AUC). The results are as shown in the table below:

The models were also assessed using also using confusion matrices to provide a clearer picture of true positives, false positives, false negatives, and true negatives. This is shown in the figure below. The AdaBoost model emerged as the most sensitive, recording the highest recall (0.789) and F1 score (0.603) among all models. According to the confusion matrix, AdaBoost correctly identified 157 preterm birth cases (true positives) and correctly ruled out 93 non-preterm cases (true negatives). However, it also misclassified 165 non-preterm cases as preterm (false positives) and missed 42 true preterm cases (false negatives). Despite its relatively lower

Table 4.5: Updated Performance Metrics of Gradient Boosting Models

Model	Accuracy	Precision	Recall	F1 Score	AUC
LightGBM	0.580	0.513	0.678	0.584	0.606
CatBoost	0.582	0.515	0.693	0.591	0.608
AdaBoost	0.547	0.488	0.789	0.603	0.595
XGBoost	0.569	0.504	0.663	0.573	0.607

accuracy (0.547) and precision (0.488), AdaBoost's strength lies in its capacity to detect the majority of PTB cases, making it highly suitable for early warning systems and clinical triage where minimizing false negatives is critical.

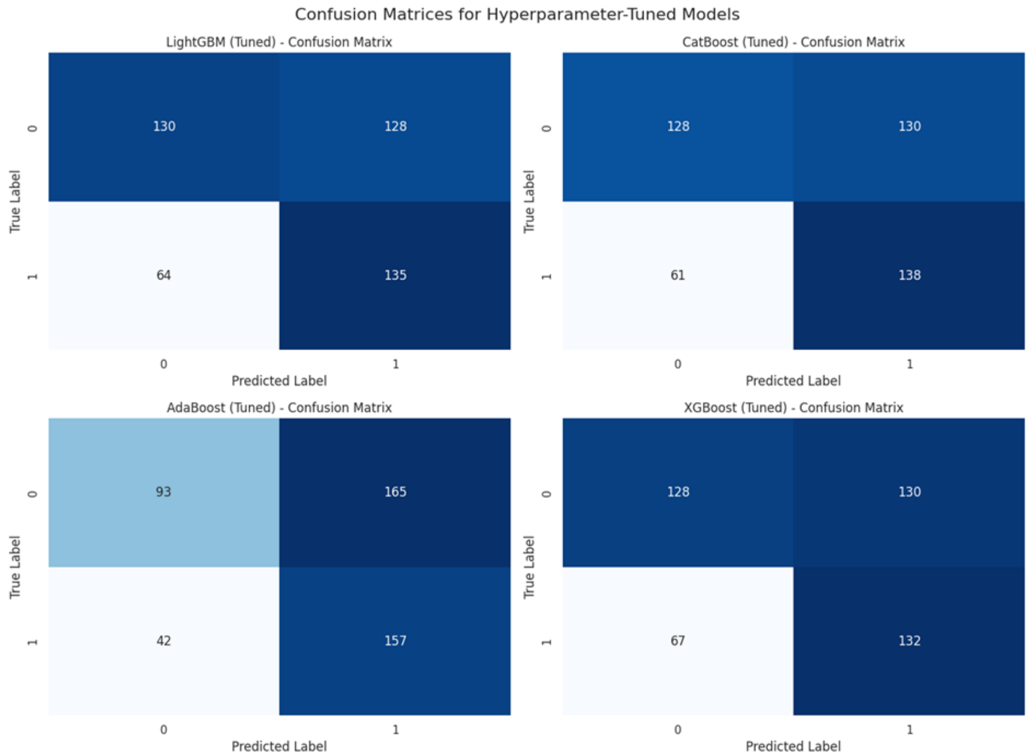

Figure 4.8: Confusion Matrix for Boosting Models

Source: Author-generated Confusion Matrix

CatBoost demonstrated the most balanced and consistent performance across all evaluated metrics. It achieved the highest AUC (0.608) and an F1 score of 0.591, with a precision of 0.515 and recall of 0.693. The model correctly identified 138 preterm births, while correctly classifying 128 non-preterm cases. It yielded 130 false positives and 61 false negatives. This balance of sensitivity and specificity suggests that CatBoost is well-calibrated for real-world clinical implementation, where both the cost of missed cases and the burden of false alarms must be minimized. Its built-in handling of categorical variables and regularization capabilities further reinforce its applicability in structured maternal health data.

XGBoost also performed reliably, with a recall of 0.663, F1 score of 0.573, and an AUC of 0.607. From the confusion matrix, XGBoost generated 132 true positives, 128 true negatives, 130 false positives, and 67 false negatives. These results indicate that XGBoost offers balanced classification capability and remains a dependable model for general-purpose maternal risk prediction, particularly in scenarios requiring stable and reproducible performance.

LightGBM showed comparable performance to XGBoost, achieving a recall of 0.678 and an F1 score of 0.584. The model produced 135 true positives, 130 true negatives, 128 false positives, and 64 false negatives. Although it had the lowest AUC (0.606) among the tuned models, its computational speed and memory efficiency make LightGBM well-suited for rapid prediction environments, especially in resource-constrained or real-time clinical applications.

These findings emphasize that each boosting model exhibits distinct strengths. AdaBoost is optimal for scenarios prioritizing high recall, such as early identification of high-risk pregnancies. CatBoost is best positioned for clinical integration due to its balance of precision and recall. XGBoost offers reliable, all-around performance suitable for varied deployment contexts, while LightGBM stands out for its scalability and efficiency. These models provide flexible and adaptable options for supporting preterm birth prediction in diverse maternal healthcare settings, particularly where timely referral and risk stratification are essential.

4.6.2 Feature Importance Analysis of the CatBoost Model

To interpret the inner workings of the tuned CatBoost model, a feature importance analysis was conducted using both CatBoost's internal method and SHAP values. This allowed for the identification of variables most influential in predicting preterm birth (PTB). The analysis helps establish clinical relevance and enhances transparency in model decision-making, which is essential for the potential integration of machine learning into maternal healthcare workflows.

VT OMNES VNVM SINT

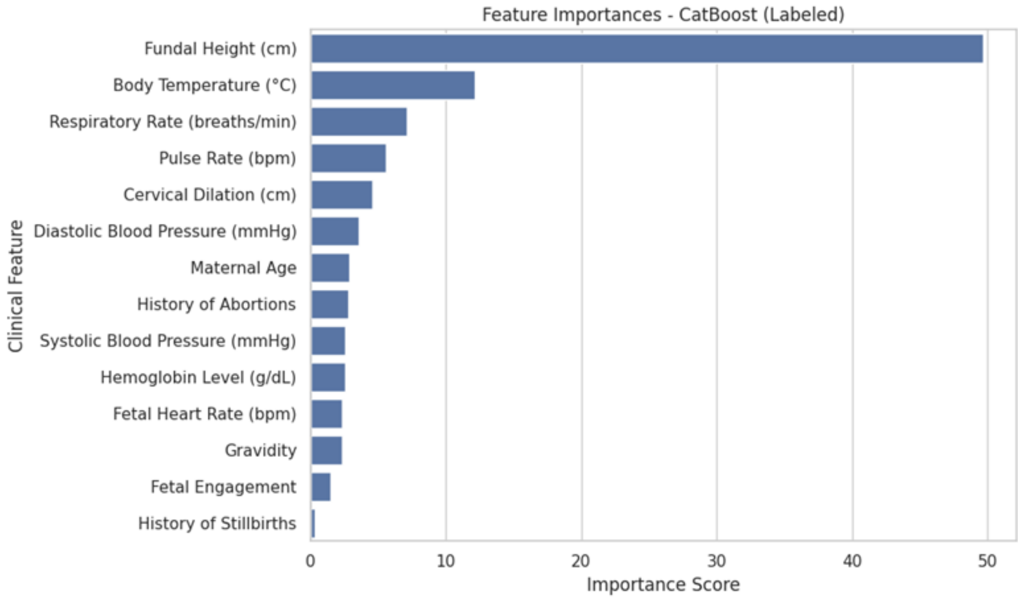

Figure 4.9: Feature Importance for Catboost

Source: Author-generated feature importances

The results, as visualized in the feature importance plot, indicate that Fundal Height (cm) was by far the most influential predictor in the model. It accounted for nearly half of the total importance score, underscoring its strong association with PTB risk. This finding aligns with clinical expectations, as abnormal fundal height measurements often signal foetal growth restrictions or pregnancy complications, both of which are associated with preterm delivery.

Following fundal height, Body Temperature (°C) and Respiratory Rate (breaths/min) emerged as the next most important features. Elevated maternal temperature may reflect systemic infection or inflammatory responses, which are known contributors to spontaneous preterm labor. Similarly, maternal respiratory rate can be a marker of physiological stress or infection, particularly in low-resource settings where conditions such as malaria or urinary tract infections may be prevalent.

Other physiologically relevant predictors included Pulse Rate (bpm), Cervical Dilation (cm), and Diastolic Blood Pressure (mmHg). These features likely reflect the maternal body's response to labor initiation or distress, further supporting their relevance in risk stratification.

Demographic and obstetric history variables such as Maternal Age, History of Abortions, and History of Stillbirths contributed modestly to the model's decision-making. While they are traditionally considered in risk assessments, their lower importance scores in this model may reflect the stronger signal provided by real-time clinical measurements.

Interestingly, Foetal Engagrment, Gravidity, and Foetal Heart Rate (bpm) had the lowest importance scores in this analysis. This does not imply irrelevance but suggests that in the presence of more dynamic and acute clinical indicators, these features contributed less to the model's predictive performance.

4.6.3 Model Explainability Using SHAP Values

To enhance the interpretability of the CatBoost model developed for preterm birth (PTB) prediction, SHAP (SHapley Additive exPlanations) values were employed. SHAP values quantify the contribution of each feature to the final prediction output for individual observations, allowing a more transparent understanding of model decision-making. The SHAP summary plot below presents both the magnitude and direction of each feature's influence on the model's output across the test dataset. Features are ordered by their importance, with those at the top exerting the most substantial impact on the prediction.

Fundal Height was confirmed as the most influential feature, with higher values strongly associated with increased predicted risk of PTB. As observed in the plot, red points (indicating higher values) are distributed on the right-hand side of the x-axis, meaning high fundal height increases the SHAP value and thus the model's prediction of preterm birth. This aligns with clinical evidence that deviations in fundal height may reflect abnormal foetal growth patterns or intrauterine complications.

Body Temperature followed closely in importance, with higher temperatures (red points) contributing positively to the PTB risk prediction. Elevated maternal temperature may signal underlying infection or inflammation, both recognized contributors to spontaneous preterm labor.

Respiratory Rate also demonstrated a positive relationship with PTB risk. Increased respiratory rates (shown in red) were generally associated with higher SHAP values, suggesting that physiological stress or systemic infection may play a role in the model's risk estimation

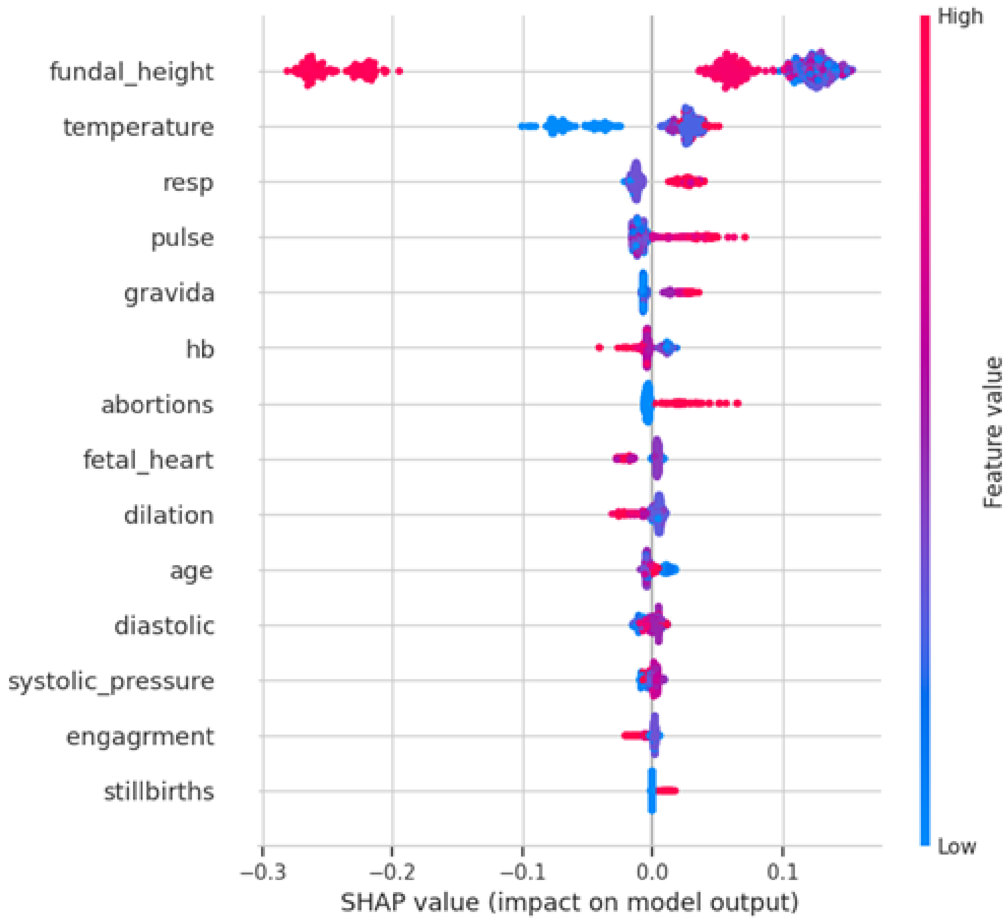

Figure 4.10: SHAP Values for Catboost

Source: Author-generated SHAP values

Features such as Pulse Rate, Gravidity, Haemoglobin Level, and Cervical Dilation exhibited more moderate influence. Notably, Cervical Dilation had both positive and negative SHAP contributions, indicating its nuanced role depending on individual cases low dilation may reduce risk, while abnormal early dilation could indicate labor onset.

Lower-ranked features such as Foetal Heart Rate, Maternal Age, Systolic/Diastolic Blood Pressure, Foetal Engagement, and History of Stillbirths showed minimal influence in the model's overall predictions. Although clinically relevant, these features may provide less unique signal when other acute indicators (like fundal height and temperature) are already captured.

Model Deployment

Following model development and performance evaluation, the best-performing machine learning model for predicting preterm birth (PTB) risk was selected and deployed as a user-friendly clinical decision-support tool. The deployment was done using Streamlit, an open-source Python library designed to simplify the creation of interactive web applications for machine learning models.

Figure 4.11: Deployment using Streamlit

Streamlit's ease of integration and minimal configuration made it ideal for rapidly deploying the model in a manner accessible to healthcare practitioners and field users. The deployed interface enables users such as clinicians, nurses, or community health workers to input a patient's clinical parameters, including age, systolic and diastolic blood pressure, pulse rate, foetal heart rate, haemoglobin levels, temperature, and other obstetric data relevant to PTB risk assessment. Once the inputs are submitted, the application processes them through the machine learning model and returns a probabilistic prediction indicating the likelihood of preterm birth.

For example, as illustrated in the deployed web interface, the model assessed a patient as having a Low Risk of Preterm Birth, with a predicted probability of 52.73%. This output provides not only a binary classification (e.g., "Low Risk" or "High Risk") but also a numerical confidence score, allowing practitioners to make more informed decisions about referral or escalation of care. Based on comprehensive evaluation, the CatBoost model was selected as the final model for deployment. This decision was informed by multiple factors including the fact that CatBoost achieved the highest area under the ROC curve (AUC = 0.608), indicating superior discrimination between PTB and non-PTB cases. It also maintained a strong balance between recall (0.693) and precision (0.515), making it both sensitive and reliable. The confusion matrix revealed that CatBoost had one of the best trade-offs between false positives and false negatives, correctly predicting 138 preterm births and 128 non-preterm births, while minimizing both missed cases and false alerts.

The model is also robust in Real-World as CatBoost handles missing values internally and is designed to work well with categorical and imbalanced datasets, which are common in maternal health records in low-resource settings. Its ability to manage raw categorical features without extensive preprocessing improves its suitability for deployment in live systems. While maintaining competitive accuracy and F1 scores, CatBoost also offers tools for interpretability such as feature importance scores and SHAP values, which are essential for model transparency in clinical environments. CatBoost consistently performed well across different test configurations and was less sensitive to the removal of dominant features like fundal height, suggesting better generalizability to other health facilities and data sources.

The CatBoost-based preterm birth risk prediction tool deployed combines technical rigor with clinical relevance, offering a practical solution to improve maternal care through timely risk stratification and intervention. Its integration into a web-based interface bridges the gap between complex machine learning algorithms and frontline clinical decision-making, especially in resource-constrained settings where early detection of PTB risk can significantly improve maternal and neonatal outcomes.

4.7 Discussion

This study evaluated multiple machine learning (ML) models for preterm birth (PTB) prediction, leveraging maternal demographic, clinical, and physiological data. This study employed various machine learning (ML) models Logistic Regression, Random Forest, XGBoost, Support Vector Machine (SVM), and Neural Networks to predict preterm birth (PTB) using maternal demographic and clinical data. The models were trained on a structured dataset and evaluated using standard metrics including accuracy, precision, recall, F1 score, and AUC. Class imbalance was mitigated using SMOTE, and model performance was further improved via hyperparameter tuning.

4.7.1 Data Availability and Quality

While the predictive models demonstrated moderate to high performance, the quality and completeness of the dataset posed critical limitations. Key socio-demographic characteristics such as education level, employment status, marital status, antenatal visits and blood sugar that have been poised to be key predictors of preterm birth was missing from the priority variables tracked at the facility. Studies across low-resource settings, including Kenya and Ethiopia, highlight systemic gaps in the quality and completeness of maternal health data, which hinder the prioritization of evidence-based predictors of preterm birth (PTB) in clinical practice. Despite literature identifying biomarkers (e.g., foetal fibronectin, PAMG-1) and ultrasound markers (e.g., cervical length, uterocervical angle) as strong predictors of PTB, these are rarely collected routinely in low-resource health facilities. Instead, facilities prioritize basic demographic and clinical variables (e.g., maternal age, gravidity) due to limited access to advanced diagnostics (Tosto et al., 2023)(Wanyoro et al., 2020). In Migori County, Kenya, gestational age documentation was inconsistent, with only 58% of stillbirths accurately reported in maternity registers. Early pregnancy ultrasound, critical for accurate gestational dating, remains unavailable in many rural settings (Wanyoro et al., 2020). Retrospective studies in Ethiopia and Uganda noted missing data on key predictors such as prior PTB history, antenatal haemorrhage, and hypertensive disorders due to reliance on paper-based systems. For example, 42% of stillbirths were underreported in Kenya's Ministry of Health (MOH) 711 forms compared to maternity registers (Wanyoro et al., 2020). In Ethiopia, models predicting PTB relied on variables like residence, haemoglobin levels, and pregnancy-induced hypertension, but excluded biomarkers and ultrasound data due to their absence in medical records (Feleke et al., 2022)(Getzzg, 2023). In Uganda, logistic regression models identified employment status and preeclampsia as PTB predictors but lacked data on cervical length or foetal fibronectin, which are standard predictors in high-income settings (Etil et al., 2023). These are often omitted from routine antenatal data collection due to logistical, infrastructural, or resource constraints in many low- and middle-income settings (Goldenberg et al., 2008). Moreover, inconsistencies in how routine variables are recorded, as well as substantial missingness in features such as temperature, dilation, and haemoglobin levels, impacted model training and generalization. This reflects broader challenges in leveraging real-world health data for predictive analytics, where data quality, standardization, and completeness often fall short of the requirements for robust machine learning models (Rajkomar et al., 2019).

4.7.2 Model Performance

The study employed several machine learning algorithms, including Logistic Regression, Random Forest, Support Vector Machines (SVM), and advanced ensemble methods such as XGBoost, LightGBM, CatBoost, and AdaBoost, to predict preterm birth (PTB) using maternal clinical data. The performance of these models was evaluated using standard metrics accuracy, precision, recall, F1-score, and the Area Under the Curve (AUC).

Among traditional machine learning models, Random Forest achieved the highest accuracy (0.586), followed by XGBoost (0.569) and Logistic Regression (0.565). In terms of recall a critical metric in identifying true PTB cases XGBoost outperformed the others with a recall of 0.663, indicating its capacity to detect the majority of positive (preterm) cases. However, the overall performance of traditional models remained modest, with AUC values ranging from 0.560 (Logistic Regression) to 0.607 (XGBoost). It achieved a balanced trade-off across all metrics, outperforming both linear and non-linear classifiers. XGBoost's gradient boosting framework and inherent ability to handle complex feature interactions likely contributed to its robustness (T. Chen & Guestrin, 2016). Although Random Forest also performed well, it was outpaced by XGBoost in terms of generalization and AUC. Similar studies have reported similar findings where XGBoost emerged as the most robust model, achieving high AUC scores (0.757–0.950) across diverse datasets (Ding et al., 2024)(Kloska et al., 2025). Support Vector Machines (SVM) lagged in precision, likely due to sensitivity to kernel selection and regularization parameters (Souza, 2025).

Logistic Regression and Neural Networks underperformed, possibly due to the former's linearity and the latter's sensitivity to architecture and training size. Studies have shown that Logistic Regression, despite its interpretability, often showed limited predictive power (AUC: 0.64–0.72), underscoring its inadequacy in modelling non-linear relationships inherent in PTB risk factors (W. Khan et al., 2023)(Teng et al., 2025). Neural Networks, though theoretically capable of capturing complex patterns, underperformed relative to tree-based models, possibly due to insufficient data size or architectural simplicity (Souza, 2025).

For the boosting algorithms, CatBoost emerged as the top performer across most metrics, attaining an accuracy of 0.582, precision of 0.515, recall of 0.693, F1-score of 0.591, and the highest AUC (0.608). LightGBM and XGBoost followed closely with AUCs of 0.606 and 0.607, respectively. AdaBoost, while posting a slightly lower AUC (0.595), achieved the highest recall (0.789), making it highly sensitive to true positives an important consideration in clinical settings where missing high-risk pregnancies could have severe consequences.

When compared to existing literature, our models' performance is within the range reported in similar studies, particularly those conducted in low- and middle-income countries (LMICs). For instance, a study utilizing data from the New Mothers-to-Be (nuMoM2b) dataset reported AUCs ranging from 0.6161 at the first prenatal visit to 0.7087 at the third visit, with improved performance upon incorporating ultrasound measurements such as cervical length and pulsatility index (Huang et al., 2024). Another study in a resource-limited setting developed a simplified risk prediction model with an AUC of 0.687, indicating moderate discriminative ability (Kassahun et al., 2024). However, studies in high-resource settings have reported higher AUCs. For example, an XGBoost model achieved an AUC of 0.893 on the validation set and 0.911 on an external dataset, utilizing comprehensive clinical and laboratory data (Teng et al., 2025). Similarly, a deep learning model using transformer architecture reported an AUC of 0.792, demonstrating the potential of advanced algorithms and richer datasets in improving predictive performance (Zhang et al., 2024).

The relatively lower performance of our models can be attributed to several factors. Firstly, the dataset was limited to routine maternal clinical parameters without the inclusion of longitudinal antenatal follow-up data, which have been shown to enhance predictive accuracy in other studies. Secondly, inconsistencies in electronic health records (EHRs), missing data, and variations in clinical documentation quality may have affected the model's training and performance. Furthermore, the imbalanced nature of PTB cases in the dataset, despite the use of SMOTE, may still contribute to suboptimal classification performance (Fergus et al., 2018).

Fundal height, and foetal heart monitoring emerged as meaningful predictors, underscoring the value of routine physiological monitoring in antenatal care. These findings resonate with existing clinical literature. For example, maternal anaemia and hypertensive disorders are both established risk factors for PTB (Teng et al., 2025)(Teshome et al., 2024).

4.7.3 Deployment

The deployment of a machine learning model for preterm birth risk prediction via Streamlit represents a significant step toward practical clinical integration. This framework enables healthcare providers to input maternal health indicators such as blood pressure, foetal heart rate, and obstetric history into an intuitive interface, generating real-time risk assessments. Streamlit's clinical utility by eliminating the need for coding expertise, allowing clinicians to interact with the model via dropdown menus, sliders, and input fields (Yu et al., 2024). Successful deployment requires harmonization with electronic health records (EHRs), as demonstrated in studies using perinatal datasets (Zhang et al., 2024). Streamlit's lightweight design allows deployment on mobile devices, potentially supporting rural or low-resource settings where preterm birth rates are highest.

Chapter 5: Conclusion And Recommendations

5.1 Conclusion

This study explored the application of machine learning techniques to predict preterm birth (PTB) using routinely collected antenatal care data. Five baseline models Logistic Regression, Random Forest, XGBoost, Support Vector Machines (SVM), and Neural Networks were developed and evaluated on a structured dataset.

Recognizing the inherent class imbalance, Synthetic Minority Oversampling Technique (SMOTE) was applied to enhance model sensitivity to the minority class (preterm births). In addition, hyperparameter tuning was conducted to further optimize model performance. Among all models, XGBoost emerged as the best-performing classifier, achieving the most balanced trade-off across key metrics such as accuracy, precision, recall, F1 score, and AUC. Random Forest also performed competitively, particularly after hyperparameter optimization, while SVM showed notable improvement in recall post-tuning. Logistic Regression and Neural Networks demonstrated moderate predictive capabilities but were limited by model simplicity and data constraints.

Based on this initial model performance the study explored deeply the boosting models. The investigation demonstrated the feasibility of utilizing models such as CatBoost, XGBoost, LightGBM, and AdaBoost to classify preterm birth risk, offering an alternative to traditional risk assessment tools that are often limited in sensitivity and specificity. Among the tested models, CatBoost exhibited the most balanced performance across evaluation metrics, including an AUC of 0.608 and recall of 0.693, while AdaBoost achieved the highest recall (0.789), underscoring its potential value in clinical triage where minimizing false negatives is critical. Feature importance analysis revealed that fundal height, body temperature, respiratory rate, pulse rate, and cervical dilation were the most predictive variables, with SHAP analysis further confirming their significant and interpretable contributions to model decisions.

The models achieved moderate predictive performance, however none exceeded the commonly accepted AUC threshold of 0.70. Although LMIC findings have reported similar findings to those in this study. This reflects limitations inherent in the available dataset, including the absence of advanced longitudinal data such as ANC coverage, nutritional data as well as challenges related to data quality and completeness typical in LMIC settings. As such, while the models demonstrate potential, their current performance may not yet support standalone clinical implementation. Nonetheless, the work highlights the promise of machine learning for maternal risk stratification in under-resourced environments. With enhanced data quality, incorporation of additional predictors, and external validation across diverse healthcare settings, such models can be optimized to support earlier detection and more effective management of high-risk pregnancies.

The findings from this study contribute to the growing body of evidence supporting the integration of artificial intelligence in maternal and neonatal health. They also serve as a foundation for future research aimed at developing context-specific, explainable, and scalable decision support tools that can augment clinical care and improve perinatal outcomes in low-resource settings.

5.2 Recommendations

1. Strengthen Data Collection and Quality in Maternal Health Records: Standardize the inclusion of evidence-based predictors such as maternal age, haemoglobin levels, blood pressure, prior PTB history, and gestational age (via last menstrual period (LMP) or simplified ultrasound protocols) in antenatal records. These variables have demonstrated strong predictive value even in resource-limited settings. Standardization of maternal health records should be backed by digital health systems. Transition from paper-based to electronic health records (EHRs) to reduce missing data and improve integration of maternal health indicators across facilities. Additionally, capacity building for clinicians and community health workers (CHWs) to recognize and document high-risk factors (e.g., hypertension, anaemia, prior PTB) during antenatal care. This aligns with WHO guidelines for PTB prevention in low-resource settings.

2. Policy and Resource Allocation: National maternal health strategies should formally integrate predictive analytics and machine learning tools into antenatal care (ANC) protocols. These tools can assist with early identification and referral of high-risk pregnancies, potentially reducing the incidence of preventable preterm births. Governments should invest in secure, interoperable digital health information systems (DHIS2, OpenMRS, etc.) that support automated data collection, model integration, and ethical AI deployment while safeguarding patient privacy. Finally, Regulatory frameworks should be established to monitor and mitigate algorithmic bias, ensuring fair performance across socio-demographic groups, especially in underserved populations disproportionately affected by PTB.

5.3 Limitations and Future Directions

This study is not without limitations. The dataset was derived from a single centre, limiting generalizability. The generalizability of the findings to other regions or populations may be limited due to differences in healthcare systems, patient demographics, and risk factor profiles.

Deployment in other regions outside the scope of this work may require revalidation owing to differences in contextual and data characteristics.

In addition, machine learning models can identify correlations between risk factors and PTB, but they cannot establish causation. The identified risk factors may not be directly causative but rather associated with other underlying factors. Additionally, gestational age estimation relied on proxy methods (LMP and EDD), potentially introducing misclassification. Lastly small sample size limited the performance of deeper models like neural networks. Future research should incorporate multi-centre data, expanded feature sets, and real-time clinical integration of models. A priority should be placed on improving electronic health record (EHR) infrastructure to support inclusion of underutilized yet clinically relevant variables.

VT OMNES VNVM SINT

References

Abraham, A., Le, B., Kosti, I., Straub, P., Velez-Edwards, D. R., Davis, L. K., Newton, J. M., Muglia, L. J., Rokas, A., Bejan, C. A., Sirota, M., & Capra, J. A. (2022). Dense phenotyping from electronic health records enables machine learning-based prediction of preterm birth. *BMC medicine*, *20*(1), 333. https://doi.org/10.1186/s12916-022-02522-x

Ajayi, K. V., Bolarinwa, O., Adekunle, T. E., Alawode, O. A., Siuluta, N., Shongwe, S., & McCallum, E. (2025). Prevalence and determinants of preterm birth among women of reproductive age in Kenya: A Multilevel Analysis of the 2022 Demographic Health Survey [Publisher: Sage]. *Therapeutic Advances in Reproductive Health*. Retrieved March 24, 2025, from https://ray.yorksj.ac.uk/id/eprint/11685/

Al Ghadban, Y., Du, Y., Charnock-Jones, D. S., Garmire, L. X., Smith, G. C. S., & Sovio, U. (2024). Prediction of spontaneous preterm birth using supervised machine learning on metabolomic data: A case-cohort study. *BJOG: an international journal of obstetrics and gynaecology*, *131*(7), 908–916. https://doi.org/10.1111/1471-0528.17723

Anyango, L. (2024, June). Kenya Races Against Time to Reduce Newborn Deaths. Retrieved March 24, 2025, from https://www.kisumu.go.ke/kenya-races-against-time-to-reduce-newborn-deaths/

Arabi Belaghi, R., Beyene, J., & McDonald, S. D. (2021). Prediction of preterm birth in nulliparous women using logistic regression and machine learning. *PLoS ONE*, *16*(6), e0252025. https://doi.org/10.1371/journal.pone.0252025

Begum, M., Redoy, R., & Anty, A. (2021, February). *Preterm Baby Birth Prediction using Machine Learning Techniques* [Pages: 54]. https://doi.org/10.1109/ICICT4SD50815.2021.9396933

Belaghi, R. A. (2024). Prediction of preterm birth in multiparous women using logistic regression and machine learning approaches [Publisher: Nature Publishing Group]. *Scientific Reports*, *14*(1), 21967. https://doi.org/10.1038/s41598-024-60097-4

Borboa-Olivares, H., Rodríguez-Sibaja, M. J., Espejel-Nuñez, A., Flores-Pliego, A., Mendoza-Ortega, J., Camacho-Arroyo, I., González-Camarena, R., Echeverría-Arjonilla, J. C., & Estrada-Gutierrez, G. (2023). A Novel Predictive Machine Learning Model Integrating Cytokines in Cervical-Vaginal Mucus Increases the Prediction Rate for Preterm Birth [Number: 18 Publisher: Multidisciplinary Digital Publishing Institute]. *International Journal of Molecular Sciences*, *24*(18), 13851. https://doi.org/10.3390/ijms241813851

Burgos-Artizzu, X. P., Baños, N., Coronado-Gutiérrez, D., Ponce, J., Valenzuela-Alcaraz, B., Moreno-Espinosa, A. L., Grau, L., Perez-Moreno, Á., Gratacós, E., & Palacio, M. (2021). Mid-trimester prediction of spontaneous preterm birth with automated cervical quantitative ultrasound texture analysis and cervical length: A prospective study [Publisher: Nature Publishing Group]. *Scientific Reports*, *11*(1), 7469. https://doi.org/10.1038/s41598-021-86906-8

Business, H. (2023, November). Kenya's Health Ministry to Boosts Preterm Survival with Key Interventions [Section: News]. Retrieved March 24, 2025, from https://healthbusiness.co.ke/7151/kenyas-health-ministry-to-boosts-preterm-survival-with-key-interventions/

CAMEN, I. V., MANOLEA, M. M., VRABIE, S. C., SANDULESCU, M. S., SERBANESCU, M. S., BOLDEANU, M. V., NOVAC, L., ISTRATE-OFITERU, A. M., NEAMTU, S. D., & DIJMARESCU, A. L. (2022). Prediction of Preterm Birth with Serial Measurements of Ultrasound Markers. *Current Health Sciences Journal*, *48*(2), 162–168. https://doi.org/10.12865/CHSJ.48.02.04

Chapman, P. (2000). CRISP-DM 1.0: Step-by-step data mining guide. Retrieved March 25, 2025, from https://www.semanticscholar.org/paper/CRISP-DM-1.0%3A-Step-by-step-data-mining-guide-Chapman/54bad20bbc7938991bf34f86dde0babfbd2d5a72

Chen, T., & Guestrin, C. (2016). XGBoost: A Scalable Tree Boosting System. *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 785–794. https://doi.org/10.1145/2939672.2939785

Chen, X., Zhang, X., Li, W., Li, W., Wang, Y., Zhang, S., & Zhu, C. (2021). Iatrogenic vs. Spontaneous Preterm Birth: A Retrospective Study of Neonatal Outcome Among Very Preterm Infants [Publisher: Frontiers]. *Frontiers in Neurology*, *12*. https://doi.org/10.3389/fneur.2021.649749

Chen, Y., Shi, X., Wang, Z., & Zhang, L. (2024). Development and validation of a spontaneous preterm birth risk prediction algorithm based on maternal bioinformatics: A single-center retrospective study. *BMC Pregnancy and Childbirth*, *24*(1), 763. https://doi.org/10.1186/s12884-024-06933-x

Costello, J. M., Takasuka, H., Roger, J., Yin, O., Tang, A., Oskotsky, T., Sirota, M., & Capra, J. A. (2025). Pre-conception clinical risk factors differ between spontaneous and indicated preterm birth in a densely phenotyped EHR cohort. *BMC Pregnancy and Childbirth*, *25*, 149. https://doi.org/10.1186/s12884-025-07166-2

Ding, L., Yin, X., Wen, G., Sun, D., Xian, D., Zhao, Y., Zhang, M., Yang, W., & Chen, W. (2024). Prediction of preterm birth using machine learning: A comprehensive analysis based on large-scale preschool children survey data in Shenzhen of China. *BMC pregnancy and childbirth*, *24*(1), 810. https://doi.org/10.1186/s12884-024-06980-4

Etil, T., Opio, B., Odur, B., Lwanga, C., & Atuhaire, L. (2023). Risk factors associated with preterm birth among mothers delivered at Lira Regional Referral Hospital. *BMC Pregnancy and Childbirth*, *23*, 814. https://doi.org/10.1186/s12884-023-06120-4

Eyinda, M., Nyaberi, J. M., & Orora, C. N. (2024). Predictors of Pre-Term Births Among Women Delivering at Migori County Referral Hospital, Migori, Kenya [Number: 2]. *East African Journal of Health and Science*, *7*(2), 1–14. https://doi.org/10.37284/eajhs.7.2.2333

Feleke, S. F., Anteneh, Z. A., Wassie, G. T., Yalew, A. K., & Dessie, A. M. (2022). Developing and validating a risk prediction model for preterm birth at Felege Hiwot Comprehensive Specialized Hospital, North-West Ethiopia: A retrospective follow-up study. *BMJ Open*, *12*(9), e061061. https://doi.org/10.1136/bmjopen-2022-061061

Fergus, P., Montanez, C. C., Abdulaimma, B., Lisboa, P., & Chalmers, C. (2018, January). Utilising Deep Learning and Genome Wide Association Studies for Epistatic-Driven Preterm Birth Classification in African-American Women [arXiv:1801.02977 [cs]]. https://doi.org/10.48550/arXiv.1801.02977

Getzzg. (2023, May). Development of risk prediction models for preterm delivery in a rural setting in Ethiopia. Retrieved March 26, 2025, from https://jogh.org/2023/jogh-13-04051/

Gimenez, L. G., Krupitzki, H. B., Momany, A. M., Gili, J. A., Poletta, F. A., Campaña, H., Cosentino, V. R., Saleme, C., Pawluk, M., Murray, J. C., Castilla, E. E., Gadow, E. C., & Lopez-Camelo, J. S. (2016). Maternal and neonatal epidemiological features in clinical subtypes of preterm-birth. *The journal of maternal-fetal & neonatal medicine : the official journal of the European Association of Perinatal Medicine, the Federation of Asia and Oceania Perinatal Societies, the International Society of Perinatal Obstetricians*, *29*(19), 3153–3161. https://doi.org/10.3109/14767058.2015.1118035

Goldenberg, R. L., Culhane, J. F., Iams, J. D., & Romero, R. (2008). Epidemiology and causes of preterm birth. *Lancet (London, England)*, *371*(9606), 75–84. https://doi.org/10.1016/S0140-6736(08)60074-4

Gupta, A., & Katarya, R. (2020). Social media based surveillance systems for healthcare using machine learning: A systematic review. *Journal of Biomedical Informatics*, *108*, 103500. https://doi.org/10.1016/j.jbi.2020.103500

Gupta, S., & Sedamkar, R. (2019, March). *Apply Machine Learning for Healthcare to enhance Performance and identify informative features*.

He, J.-R., Ramakrishnan, R., Lai, Y.-M., Li, W.-D., Zhao, X., Hu, Y., Chen, N.-N., Hu, F., Lu, J.-H., Wei, X.-L., Yuan, M.-Y., Shen, S.-Y., Qiu, L., Chen, Q.-Z., Hu, C.-Y., Cheng, K. K., Mol, B. W. J., Xia, H.-M., & Qiu, X. (2018). Predictions of Preterm Birth from Early Pregnancy Characteristics: Born in Guangzhou Cohort Study [Number: 8 Publisher: Multidisciplinary Digital Publishing Institute]. *Journal of Clinical Medicine*, *7*(8), 185. https://doi.org/10.3390/jcm7080185

Hu, L., Li, L., Ji, J., & Sanderson, M. (2020). Identifying and understanding determinants of high healthcare costs for breast cancer: A quantile regression machine learning approach. *BMC Health Services Research*, *20*(1), 1066. https://doi.org/10.1186/s12913-020-05936-6

Huang, C., Long, X., van der Ven, M., Kaptein, M., Oei, S. G., & van den Heuvel, E. (2024). Predicting preterm birth using electronic medical records from multiple prenatal visits. *BMC Pregnancy and Childbirth*, *24*(1), 843. https://doi.org/10.1186/s12884-024-07049-y

Islam, M. D. S., Liu, D., Wang, K., Zhou, P., Yu, L., & Wu, D. (2019). A Case Study of Health-Care Platform using Big Data Analytics and Machine Learning. *Proceedings of the 2019 3rd High Performance Computing and Cluster Technologies Conference*, 139–146. https://doi.org/10.1145/3341069.3342980

Javaid, M., Haleem, A., Pratap Singh, R., Suman, R., & Rab, S. (2022). Significance of machine learning in healthcare: Features, pillars and applications. *International Journal of Intelligent Networks*, *3*, 58–73. https://doi.org/10.1016/j.ijin.2022.05.002

Javaid, M., Haleem, A., Vaishya, R., Bahl, S., Suman, R., & Vaish, A. (2020). Industry 4.0 technologies and their applications in fighting COVID-19 pandemic. *Diabetes & Metabolic Syndrome*, *14*(4), 419–422. https://doi.org/10.1016/j.dsx.2020.04.032

Karugu, M., Oluchina, S., & Kikuvi, G. (2023). Maternal and obstetric factors associated with preterm births among mothers delivering at Kiambu Level 5 Hospital, Kenya [Number: 4]. *East African Medical Journal*, *100*(4), 5798–5808. Retrieved March 24, 2025, from https://www.ajol.info/index.php/eamj/article/view/248652

Kassahun, E. A., Gebreyesus, S. H., Tesfamariam, K., Endris, B. S., Roro, M. A., Getnet, Y., Hassen, H. Y., Brusselaers, N., & Coenen, S. (2024). Development and validation of a simplified risk prediction model for preterm birth: A prospective cohort study in rural Ethiopia [Publisher: Nature Publishing Group]. *Scientific Reports*, *14*(1), 4845. https://doi.org/10.1038/s41598-024-55627-z

Khan, M., Jan, M., Muhammad, Y., Do, D.-T., Rehman, A., Mavromoustakis, C., & Pallis, E. (2021). Tracking vital signs of a patient using channel state information and machine learning for a smart healthcare system. *Neural Computing and Applications*. https://doi.org/10.1007/s00521-020-05631-x

Khan, W., Zaki, N., Ghenimi, N., Ahmad, A., Bian, J., Masud, M. M., Ali, N., Govender, R., & Ahmed, L. A. (2023). Predicting preterm birth using explainable machine learning in a prospective cohort of nulliparous and multiparous pregnant women [Publisher: Public Library of Science]. *PLOS ONE*, *18*(12), e0293925. https://doi.org/10.1371/journal.pone.0293925

Kloska, A., Harmoza, A., Kloska, S. M., Marciniak, T., & Sadowska-Krawczenko, I. (2025). Predicting preterm birth using machine learning methods [Publisher: Nature Publishing Group]. *Scientific Reports*, *15*(1), 5683. https://doi.org/10.1038/s41598-025-89905-1

Lee, K.-S., Song, I.-S., Kim, E. S., Kim, J., Jung, S., Nam, S., & Ahn, K. H. (2024). Machine learning analysis with population data for the associations of preterm birth with temporomandibular disorder and gastrointestinal diseases. *PloS One*, *19*(1), e0296329. https://doi.org/10.1371/journal.pone.0296329

Mabrouk, A., Abubakar, A., Too, E. K., Chongwo, E., & Adetifa, I. M. (2022). A Scoping Review of Preterm Births in Sub-Saharan Africa: Burden, Risk Factors and Outcomes. *International Journal of Environmental Research and Public Health*, *19*(17), 10537. https://doi.org/10.3390/ijerph191710537

McWilliams, C. J., Lawson, D. J., Santos-Rodriguez, R., Gilchrist, I. D., Champneys, A., Gould, T. H., Thomas, M. J., & Bourdeaux, C. P. (2019). Towards a decision support tool for intensive care discharge: Machine learning algorithm development using electronic healthcare data from MIMIC-III and Bristol, UK. *BMJ open*, *9*(3), e025925. https://doi.org/10.1136/bmjopen-2018-025925

Menon, R. (2012). Preterm birth: A global burden on maternal and child health. *Pathogens and Global Health*, *106*(3), 139–140. https://doi.org/10.1179/204777312X13462106637729

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

Mitrogiannis, I., Evangelou, E., Efthymiou, A., Kanavos, T., Birbas, E., Makrydimas, G., & Papatheodorou, S. (2023). Risk factors for preterm birth: An umbrella review of meta-analyses of observational studies. *BMC Medicine*, *21*(1), 494. https://doi.org/10.1186/s12916-023-03171-4

MOH. (2019). KENYA HARMONIZED HEALTH FACILITY ASSESSMENT (KHFA) 2018/2019 MAIN REPORT.

MOH. (2023, September). KENYA HEALTH FACILITY CENSUS REPORT. https://www.health.go.ke/sites/default/files/2024-01/Kenya%20Health%20Facility%20Census%20Report%20September%202023.pdf

Moyo, S., Doan, T. N., Yun, J. A., & Tshuma, N. (2018). Application of machine learning models in predicting length of stay among healthcare workers in underserved communities in South Africa. *Human Resources for Health*, *16*(1), 68. https://doi.org/10.1186/s12960-018-0329-1

Narice, B. F., Labib, M., Wang, M., Byrne, V., Shepherd, J., Lang, Z. Q., & Anumba, D. O. (2024). Developing a logistic regression model to predict spontaneous preterm birth from maternal socio-demographic and obstetric history at initial pregnancy registration. *BMC Pregnancy and Childbirth*, *24*(1), 688. https://doi.org/10.1186/s12884-024-06892-3

Ngusie, H. S., Mengiste, S. A., Zemariam, A. B., Molla, B., Tesfa, G. A., Seboka, B. T., Alene, T. D., & Sun, J. (2024). Predicting adverse birth outcome among childbearing women in Sub-Saharan Africa: Employing innovative machine learning techniques. *BMC Public Health*, *24*(1), 2029. https://doi.org/10.1186/s12889-024-19566-8

Oskovi Kaplan, Z. A., & Ozgu-Erdinc, A. S. (2018). Prediction of Preterm Birth: Maternal Characteristics, Ultrasound Markers, and Biomarkers: An Updated Overview. *Journal of Pregnancy*, *2018*, 8367571. https://doi.org/10.1155/2018/8367571

PATH. (2025). Clinical Chemistry. Retrieved March 24, 2025, from https://www.pathcarekenya.com/our-services/clinical-chemistry/

Pollettini, J. T., Panico, S. R. G., Daneluzzi, J. C., Tinós, R., Baranauskas, J. A., & Macedo, A. A. (2012). Using machine learning classifiers to assist healthcare-related decisions: Classification of electronic patient records. *Journal of Medical Systems*, *36*(6), 3861–3874. https://doi.org/10.1007/s10916-012-9859-6

Rajkomar, A., Dean, J., & Kohane, I. (2019). Machine Learning in Medicine [Publisher: Massachusetts Medical Society _eprint: https://www.nejm.org/doi/pdf/10.1056/NEJMra1814259]. *New England Journal of Medicine*, *380*(14), 1347–1358. https://doi.org/10.1056/NEJMra1814259

Ramachandran, A., Clottey, K. D., Gordon, A., & Hyett, J. A. (2024). Prediction and prevention of preterm birth: Quality assessment and systematic review of clinical practice guidelines using the AGREE II framework [_eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1002/ijgo.155 *International Journal of Gynecology & Obstetrics*, *166*(3), 932–942. https://doi.org/10.1002/ijgo.15514

Rani, S., Kataria, A., Bhambri, P., Pareek, P. K., & Puri, V. (2024). Artificial Intelligence in Personalized Health Services for Better Patient Care. In S. K. Gupta, D. A. Karras, & R. Natarajan (Eds.), *Revolutionizing Healthcare: AI Integration with IoT for Enhanced Patient Outcomes* (pp. 89–108). Springer Nature Switzerland. https://doi.org/10.1007/978-3-031-65022-2_6

Rao, C. R., Bhat, P., Ke, V., Kamath, V., Kamath, A., Nayak, D., Shenoy, R. P., & Bhat, S. K. (2018). Assessment of risk factors and predictors for spontaneous pre-term birth in a South Indian antenatal cohort. *Clinical Epidemiology and Global Health*, *6*(1), 10–16. https://doi.org/10.1016/j.cegh.2017.07.001

Roth, J. A., Battegay, M., Juchler, F., Vogt, J. E., & Widmer, A. F. (2018). Introduction to Machine Learning in Digital Healthcare Epidemiology. *Infection Control and Hospital Epidemiology*, *39*(12), 1457–1462. https://doi.org/10.1017/ice.2018.265

Shah, P., Kendall, F., Khozin, S., Goosen, R., Hu, J., Laramie, J., Ringel, M., & Schork, N. (2019). Artificial intelligence and machine learning in clinical development: A translational perspective [Publisher: Nature Publishing Group]. *npj Digital Medicine*, *2*(1), 1–5. https://doi.org/10.1038/s41746-019-0148-3

Souza, H. F. d. (2025, February). AI predicts preterm birth risk with 82% accuracy [Section: Medical News]. Retrieved March 26, 2025, from https://www.news-medical.net/news/20250217/AI-predicts-preterm-birth-risk-with-8225-accuracy.aspx

Sun, Q., Zou, X., Yan, Y., Zhang, H., Wang, S., Gao, Y., Liu, H., Liu, S., Lu, J., Yang, Y., & Ma, X. (2022). Machine Learning-Based Prediction Model of Preterm Birth Using Electronic Health Record. *Journal of Healthcare Engineering*, *2022*, 9635526. https://doi.org/10.1155/2022/9635526

Swiercz, G., Zmelonek-Znamirowska, A., Szwabowicz, K., Armanska, J., Detka, K., Mlodawska, M., & Mlodawski, J. (2024). Evaluating the predictive efficacy of first trimester biochemical markers (PAPP-A, fβ-hCG) in forecasting preterm delivery incidences. *Scientific Reports*, *14*(1), 16206. https://doi.org/10.1038/s41598-024-67300-6

Tekeba, B., Techane, M. A., Workneh, B. S., Zegeye, A. F., Gonete, A. T., Ahmed, M. A., Wassie, Y. A., Wassie, M., Kassie, A. T., Ali, M. S., Mekonen, E. G., Tamir, T. T., & Tsega, S. S. (2024). Determinants of preterm birth among reproductive age women in sub-Saharan Africa: Evidence from the most recent Demographic and Health Survey data-2019-2022. *PloS One*, *19*(6), e0305810. https://doi.org/10.1371/journal.pone.0305810

Teng, X., Liu, M., Wang, Z., & Dong, X. (2025). Machine learning prediction of preterm birth in women under 35 using routine biomarkers in a retrospective cohort study [Publisher: Nature Publishing Group]. *Scientific Reports*, *15*(1), 10213. https://doi.org/10.1038/s41598-025-92814-y

Teshome, F., Gidi, N. W., Choe, S.-w., Kong, J. D., & Ayana, G. (2024, September). Machine Learning for Early Prediction of Preterm Birth. https://doi.org/10.2139/ssrn.4952965

The Alliance for Maternal and Newborn Health Improvement (AMANHI) GA Study Group. (2022). Population-based rates, risk factors and consequences of preterm births in South-Asia and sub-Saharan Africa: A multi-country prospective cohort study. *Journal of Global Health*, *12*, 04011. https://doi.org/10.7189/jogh.12.04011

Tong, Y., Messinger, A. I., & Luo, G. (2020). Testing the Generalizability of an Automated Method for Explaining Machine Learning Predictions on Asthma Patients' Asthma Hospital Visits to an Academic Healthcare System. *IEEE access: practical innovations, open solutions*, *8*, 195971–195979. https://doi.org/10.1109/access.2020.3032683

Tosto, V., Tsibizova, V., Clerici, G., & Di Renzo, G. C. (2023, November). Preterm Birth Prediction — Article — GLOWM. Retrieved March 24, 2025, from http://www.glowm.com/article/heading/vol-19--pregnancy-shortening-etiology-prediction-and-prevention--preterm-birth-prediction/id/419063

UNFPA. (2021). Midwives in rural Kenya embrace portable ultrasound technology for safer pregnancies. Retrieved March 24, 2025, from https://kenya.unfpa.org/en/news/midwives-rural-kenya-embrace-portable-ultrasound-technology-safer-pregnancies

Wagura, P., Wasunna, A., Laving, A., Wamalwa, D., & Ng'ang'a, P. (2018). Prevalence and factors associated with preterm birth at kenyatta national hospital. *BMC Pregnancy and Childbirth*, *18*(1), 107. https://doi.org/10.1186/s12884-018-1740-2

Wanyonyi, D. (2025, February). Ultrasound Imaging: A major Setback for Pregnant Women in Kenya, Health Experts Say. Retrieved March 24, 2025, from https://www.talkafrica.co.ke/access-to-ultrasound-imaging-a-major-setback-to-pregrant-women-in-kenya-health-experts-say/

Wanyoro, A., Ogolla, D., Merai, R., Otare, C., Simotwo, R., Mutindi, G., Mumo, E., Walker, D., & Otieno, P. (2020). Outcome of data quality assessments of maternity records across 17 health facilities in Migori County, Western Kenya [Publisher: International Society of Global Health]. *Journal of Global Health Reports*, *4*, e2020104. https://doi.org/10.29392/001c.18112

Włodarczyk, T., Płotka, S., Szczepański, T., Rokita, P., Sochacki-Wójcicka, N., Wójcicki, J., Lipa, M., & Trzciński, T. (2021). Machine Learning Methods for Preterm Birth Prediction: A Review [Number: 5 Publisher: Multidisciplinary Digital Publishing Institute]. *Electronics*, *10*(5), 586. https://doi.org/10.3390/electronics10050586

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

Yu, Q.-Y., Lin, Y., Zhou, Y.-R., Yang, X.-J., & Hemelaar, J. (2024). Predicting risk of preterm birth in singleton pregnancies using machine learning algorithms [Publisher: Frontiers]. *Frontiers in Big Data*, *7*. https://doi.org/10.3389/fdata.2024.1291196

Zeng, X., Jiang, W., He, X., & Zhang, J. (2024, May). Preterm Birth: Epidemiology, Risk Factors, Pathogenesis, and Prevention. https://doi.org/10.1093/acrefore/9780190632366.013.515

Zhang, F., Tong, L., Shi, C., Zuo, R., Wang, L., & Wang, Y. (2024). Deep Learning in Predicting Preterm Birth: A Comparative Study of Machine Learning Algorithms [Publisher: Chinese Medical Association Publishing House]. *Maternal-Fetal Medicine*, *06*(03), 141–146. https://doi.org/10.1097/FM9.0000000000000236

Zhu, J., Ren, Z., & Lee, C. (2021). Toward Healthcare Diagnoses by Machine-Learning-Enabled Volatile Organic Compound Identification. *ACS nano*, *15*(1), 894–903. https://doi.org/10.1021/acsnano.0c07464

VT OMNES VNVM SINT

Appendix A: Similarity Report

217038975 343773923

167860_Dissertation Final_2025 (1).pdf

Turnitin

Document Details

Submission ID

trn:oid:::27005:90133046

Submission Date

Apr 7, 2025, 5:10 PM UTC

Download Date

Apr 7, 2025, 5:12 PM UTC

File Name

167860_Dissertation_Final_2025_1.pdf

File Size

1.4 MB

48 Pages

15,803 Words

94,123 Characters

23% Overall Similarity

The combined total of all matches, including overlapping sources, for each database.

Filtered from the Report

- Bibliography

Match Groups

242 Not Cited or Quoted 20%
Matches with neither in-text citation nor quotation marks

44 Missing Quotations 3%
Matches that are still very similar to source material

1 Missing Citation 0%
Matches that have quotation marks, but no in-text citation

0 Cited and Quoted 0%
Matches with in-text citation present, but no quotation marks

Top Sources

16% Internet sources

15% Publications

17% Submitted works (Student Papers)

Integrity Flags

0 Integrity Flags for Review

No suspicious text manipulations found.

Our system's algorithms look deeply at a document for any inconsistencies that would set it apart from a normal submission. If we notice something strange, we flag it for you to review.

A Flag is not necessarily an indicator of a problem. However, we'd recommend you focus your attention there for further review.

VT OMNES VNVM SINT

Appendix B: Ethical Clearance Confirmation

Strathmore
UNIVERSITY

8th April 2025

Paul Wekesa Waswa

167860

paul.wekesa@strathmore.edu

Dear Paul,

RE: Utilising Machine Learning Techniques to Predict Maternal Health Adverse Events using EHR in Low resource Settings

This is to inform you that the Office of Graduate Studies on 7th April 2025 received your acknowledgement of breach in ethical processes given that you have already collected/analysed data and proceeded to write your Dissertation/Thesis prior to obtaining Ethical clearance. Consequently, it was noted that The Strathmore University Institutional Scientific and Ethical Review Committee (SU-ISERC) cannot review your study since you have already collected data and written the Thesis. The scientific & ethical review/approval process is ONLY done before the commencement of any experiments, implementation or any collection of data (primary or secondary-including desktop review).

This is a letter for you to proceed with the next steps of your academic requirements.

Please be advised, that in future, all research proposals should be submitted to the SU-ISERC through the RHInnO Ethics platform: https://strathmoreuniversity.rhinno.net/login

Disclaimer: *1) This is not in any way an ethical approval letter. 2) Should there be any legal implications/actions emanating from the research in terms of any ethical violations, you will be personally liable.*

Yours sincerely,

Prof. Bernard Shibwabo
Director of Graduate Studies

Appendix C: Letter of Introduction.pdf

Ref: iLab/DL/25/5

21st January 2025

Dear Sir/Madam,

RE: LETTER OF INTRODUCTION

I am writing to formally introduce Paul Wekesa, a student currently enrolled in the Master of Science in Data Science & Analytics Programme at Strathmore University's Strathmore Institute of Mathematical Sciences and @iLabAfrica.

As part of his academic requirements, Paul is undertaking a research project titled "Utilizing Machine Learning Techniques to Predict Maternal Health Adverse Events using EHR in Low Resource Settings." The successful completion and approval of his dissertation is a mandatory prerequisite for the fulfillment of the programme.

As part of his study, Paul requires access to relevant data to analyze patterns and develop predictive models. The insights drawn from this research have the potential to significantly contribute to improving maternal health outcomes in low-resource settings, and therefore, your assistance in providing the necessary data would be invaluable.

We kindly request your support in facilitating access to the relevant datasets, and we assure you that all data will be handled with the utmost confidentiality and in compliance with ethical research guidelines.
Should you require any additional information or have any questions regarding the research, please do not hesitate to contact the course administrator at Lmideva@strathmore.edu.

Your support in granting him permission to collect the required data will be invaluable for the success of his research. Thank you for considering this request.

Yours Sincerely,

Dr. Joseph Sevilla
Director of @iLabAfrica and @iBizAfrica,
Strathmore University
Email: joe@strathmore.edu

VT OMNES VNVM SINT

Appendix D: Python Link

For further details, refer to the supplementary materials available at:

https://colab.research.google.com/drive/1g5EKwR97QQsJizj5masJWdHTjHY4Ul4Xscroll
To=vyfMbB8E50rb

VT OMNES VNVM SINT