# Improving Maternal and Neonatal Health Outcomes in Kenya by Leveraging Machine Learning for the Timely Detection of Preeclampsia

**By**

**Esther Ngaruiya**

**169526**

**Master of Science in Data Science and Analytics**

**2025**

# Improving Maternal and Neonatal Health Outcomes in Kenya byLeveraging Machine Learning for the Timely Detection of Preeclampsia

**By**

**Esther Ngaruiya**

**169526**

**Submitted in total fulfilment of the requirements for the degree of Master of Science inData Science and Analytics in Strathmore University**

**Institute of Mathematical SciencesStrathmore University**

**Nairobi, Kenya**

**2025**

## 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 thesis proposal contains no material previously published or written by another person except where due reference is made in the thesis itself.



Student's Name:

Esther Ngaruiya

Sign: ______________________________ Date: 26/03/2025

## Approval

The thesis of Esther Ngaruiya was reviewed and approved for examination by the following:

Dr. John Olukuru, PhD

Sign: ______________________ Date: 27th March 2025

Institute of Mathematical Sciences,Strathmore University

## Abstract

Preeclampsia remains a leading cause of maternal and neonatal morbidity and mortality globally, with its burden disproportionately high in low-resource settings such as Kenya. Timely diagnosis is critical to improving outcomes, yet healthcare systems often lack the tools for early and reliable risk assessment. This study aimed to develop an AI-powered machine learning (ML) model to predict preeclampsia among pregnant women in Kenya, based on routinely collected clinical and demographic data. The specific objectives were: (a) to identify key risk factors using statistical analysis and clinical insight; (b) to develop and evaluate various ML models for preeclampsia prediction; (c) to determine the most accurate model for risk classification; and (d) to deploy the predictive system across multiple platforms for practical use.

Using a dataset of 2,925 records collected from hospitals in Kenya's coastal region, the study applied extensive preprocessing to ensure data quality and clinical relevance. Variables included age, pre-pregnancy weight, systolic and diastolic blood pressure, proteinuria levels, parity, and history of hypertensive disorders. Five individual ML models—Logistic Regression, Random Forest, XGBoost, Support Vector Machine (SVM), and Decision Tree—were trained and evaluated. The XGBoost model emerged as the best individual performer, achieving an accuracy of 98.46%, precision of 95.59%, recall of 97.74%, and F1-score of 96.65%. However, a hybrid stacking classifier, which combined the predictions of multiple base models with Logistic Regression as a meta-learner, outperformed all individual models. It achieved an accuracy of 98.12%, a recall of 99.25%, and an F1-score of 96.00%, making it the most reliable for clinical deployment.

The final model was deployed as a functional web-based application, allowing healthcare providers to input patient data and receive immediate risk assessments. This implementation underscores the potential of AI in enhancing prenatal care by enabling early intervention. The study concludes that ensemble-based ML models, especially stacking classifiers, provide a robust and scalable solution for preeclampsia risk prediction. Future work should focus on expanding the dataset to include diverse populations, integrating additional biomarkers, and developing mobile and EHR-compatible interfaces for broader reach in underserved areas.

## Table of Content

## List of Tables

## List of Figures

## List of Abbreviations

**AI**: Artificial Intelligence

**EO-PE**: Early-Onset Preeclampsia

**GBM**: Gradient Boosting Machines

**HER**: Electronic Health Records

**ML**: Machine Learning

**NACOSTI**: National Commission for Science, Technology, and Innovation

**NMR**: Neonatal Mortality Rate

**ROC**: Receiver Operating Characteristic

**SVM**: Support Vector Machines

**TP**: True Positives

# 1 Introduction

## 1.1 Background to the Study

The health of mothers and newborns is an integral part of establishing the health status of a country in general and its progress in particular. Moller et al. (2019) clearly establishes that Morbidity and mortality have not escaped maternal and neonates despite the creation of partnership in efforts to reduce them. One of these challenges is preeclampsia, a severe hypertensive disorder of pregnancy that complicates pregnancy outcomes substantially. It is linked to stuffs like premature birth, placental abruption, organ damage, high likelihood of maternal and infant mortality (Overton et al., 2022). Worldwide, preeclampsia occurs in 2–8% of pregnancies; in developing countries, the disorder is responsible for 10–15% of maternal mortality (Nirupama et al., 2021).

In Kenya, the situation is particularly dire. The World Health Organization (WHO) estimates Kenya's maternal mortality rate at 530 deaths per 100,000 live births, with preeclampsia accounting for around 14% of these deaths (Karimi & Gitonga, 2020). Neonatal mortality is also significantly high, with 21-26 deaths per 1,000 live births, especially in rural areas where healthcare infrastructure is limited (UNICEF, 2024). Rural hospitals often suffer from inadequate healthcare resources including lack of trained personnel, basic equipment, and essential medications (KNBS and ICF, 2023). These limitations contribute to the late detection and treatment of preeclampsia, often after symptoms become severe. For example, a survey of rural health facilities in Kenya showed that only 35% had functional blood pressure monitors, and less than 30% had trained staff to recognize the early signs of preeclampsia (Kimani et al., 2020).

Many women in Kenya, particularly in rural regions, lack access to adequate prenatal care, with some receiving no antenatal services at all. Around 40% of pregnant women in Kenya do not receive the recommended minimum of four antenatal visits, and for those who do, routine screenings for high-risk conditions like preeclampsia are often missed due to resource constraints (UNICEF, 2019). This situation is further compounded by disparities in healthcare access between urban and rural areas, exacerbating the challenge of timely detection and intervention.

The inefficiencies in Kenya's healthcare system, particularly in detecting and managing preeclampsia, are glaring. A 2020 report by the Kenya Ministry of Health highlighted that delays in diagnosis due to poor infrastructure and limited access to diagnostic tools were key contributors to maternal mortality. Specifically, 43% of maternal deaths related to hypertensive disorders were preventable with timely diagnosis and treatment (Ndwiga et al., 2020).

New technologies of the fourth industrial revolution especially in the Artificial Intelligence (AI), Machine Learning (ML) have opened new possibilities in tackling these challenges. Significantly, machine learning models are promising for the realisation of comprehensive clinical data and increased accuracy of the preeclampsia risk compared to the regular observation-based approach, or blood pressure and proteinuria check-ups (Wulff et al., 2019). In resource- constrained regions such as Kenya, using ML models for prediction means that health care workers can diagnose diseases before they escalate to complicated stages.

Specifically, in the field of maternal and neonatal care, the utilization of AI and ML models has demonstrated potential all over the world. AI models in high-resource environments have presented percentages of over 90% in the assessment of the risk of preeclampsia from the clinical, demographic, as well as the biochemical data (Chalupka & Vizentini, 2020). The application of such solutions based on artificial intelligence can open significant gaps in the detection of the disease in Kenya and help save the lives of women and newborns against high rates of maternal and neonatal mortality due to preeclampsia.

These technological inventions are in support of the United Nations Sustainable Development Goal 3 that seeks to promote maternal health by ensuring the global maternal mortality ratio is reduced to fewer than 70 per 100,00 live birth by the year 2030. But for; AI & ML solutions to work in Kenya, they will need to be tailored with the Kenyan environment regarding infrastructure and health facilities, accessibility and the training of other health professionals.

### 1.1.1 Maternal and Neonatal Health Outcomes

Globally, maternal and neonatal outcomes are significantly influenced by a range of health conditions and socio-economic factors. Preeclampsia remains a key contributor to adverse outcomes for both mothers and newborns, leading to complications such as instrumental and caesarean deliveries, postpartum hemorrhage, and increased rates of neonatal mortality (Sylla *et* *al.*, 2020). Studies have also highlighted other health challenges, such as female genital mutilation (FGM), which can result in higher risks of adverse maternal and perinatal complications (Sylla *et al.*, 2020). In women with pre-existing conditions, such as congenital heart disease (CHD), maternal complications like heart failure and cardiac arrhythmias are more frequent, leading to an increased likelihood of low birth weight, neonatal resuscitation, and ventilation needs (Lammers *et al.*, 2021). Furthermore, the lack of effective and systematic data collection restricts interventions aimed at improving maternal and neonatal health outcomes, as seen in European settings where better data could enhance management strategies for premature births (Haumont *et al.*, 2020).

In Sub-Saharan Africa, maternal and neonatal health outcomes are deeply affected by resource limitations and healthcare access challenges. While performance-based financing (PBF) projects have been introduced across the region to improve maternal and neonatal care, their impact on early neonatal mortality and health outcomes remains limited (Gage and Bauoff, 2021). Despite improvements in facility deliveries and antenatal care in some countries, PBF has shown inconsistent results in directly improving neonatal health outcomes across the continent. Moreover, crisis settings and conflict zones present additional challenges, with barriers such as insecurity, lack of resources, and inadequate staffing hindering the effective delivery of maternal and neonatal healthcare interventions (Munyuzangabo *et al.*, 2021). Multi-stakeholder collaboration and innovative approaches are therefore necessary to address these gaps and ensure better health outcomes in conflict-affected populations.

Kenya reflects many of the broader regional challenges, with neonatal mortality contributing approximately 40% of under-five mortality rates, remaining a significant public health concern (Imbo *et al.*, 2021). Key determinants include maternal education levels, access to antenatal care (ANC), and neonatal birth weight, with low ANC attendance and low birth weight being associated with higher risks of neonatal mortality (Kenya National Bureau of Statistics, 2023). Additionally, the disruption of maternal and child health services due to the COVID-19 pandemic has had a profound impact on access to healthcare, necessitating community-based strategies to mitigate potential rises in peripartum deaths (Kimani *et al.*, 2020). Addressing these challenges requires targeted interventions to improve ANC access and strengthen community-based midwifery to provide safe maternal and neonatal care in Kenya's strained healthcare system.

### 1.1.2 Machine Learning in Maternal and Neonatal Health

Machine learning is a discipline of artificial intelligence that offers a host of techniques and procedures to make a computer learn from the data to make patterns that can be used to make predictions and decisions (Mennickent et al., 2023). The fundamental categories of ML are supervised learning because models are trained to predict outcomes from previously categorized data such as logistic regression, decision trees; unsupervised learning that uses unclassified data to look for similarities such as clustering, principal component analysis and reinforcement learning on training models to make decisions on the basis of reward and penalty (Sutton and Barto 2018). Supervised models are predominately used in healthcare systems since they remove the uncertainty of predictions. In addition, there are other methods comprising deep learning, which is a kind of ML embedding artificial neural networks to identify intricate patterns in medical data and has the prospect of application in maternal and neonatal health (Spurri Forroti, 2023).

In maternal and neonatal health, ML has demonstrated significant potential for improving care and outcomes. For instance, the prediction of neonatal conditions, such as infections and the evaluation of maturation in premature infants, has benefited from ML approaches. Borengo (2021) explored machine learning models for decision support systems (DSSs) that analyze physiological signals like heart rate variability (HRV) and respiration rate variability (RRV) to detect late-onset sepsis (LOS) and evaluate functional maturational age. These systems have shown promise for non-invasive, real-time diagnostics that can enhance neonatal care. Similarly, ML has been employed in analyzing fetal heart rate (FHR) using Cardiotocography (CTG) data for classifying fetal growth categories, such as Small, Normal, or Large for Gestational Age (Novelli, 2023). This application is crucial for managing complications related to maternal conditions like diabetes, which affects fetal growth. Moreover, methods like SHapley Additive exPlanations have been employed to interpret ML model results, providing insights into the significance of both maternal history and fetal signals.

The application of ML in maternal and neonatal health extends to areas like fetal distress monitoring, evaluation of Cardiotocography signals, and improved predictions of conditions like preeclampsia (Kolandaisamy *et al.*, 2022; Spurri Forroti, 2023; Mennickent *et al.*, 2023). Advanced classifiers, such as ensemble learning, recursive neural networks, and ResNet architectures, have been developed to predict specific outcomes and enhance monitoring systems, achieving balanced accuracy rates exceeding traditional methodologies. These models analyze a combination of maternal clinical features, physiological signals, and even environmental data, thereby enabling more precise and timely interventions. While ML has greatly improved prediction accuracy and clinical decision-making in maternal and neonatal healthcare, further research is necessary to refine these models and ensure they are robust, interpretable, and effective across diverse populations and healthcare settings.

### 1.1.3 Preeclampsia

Preeclampsia is a hypertensive disorder of pregnancy where women develop increased blood pressure and usually proteinuria after the 20th week of pregnancy (Overton et al., 2022; Jung et al., 2022). It is regarded as one of the “great obstetrical syndromes” because of its ability to affect multiple systems and present with diverse mechanisms of pathogenesis. Maternal and neonatal complications include organ dysfunction with special reference to the liver and kidney, seizures (eclampsia), placental abruption, preterm birth and in severe cases, maternal deaths. This results in a pathway to about 76,000 maternal and approximately 500,000 infant mortalities annually, and it is one of the vital causes of maternal and perinatal mortalities (Overton et al., 2022; Nirupama et al., 2021).

The mechanisms of broken here in preeclampsia are an innate immune response that impairs the vascular the endothelial cells. There is also the endothelial cell activation and intravascular inflammation which also leads to high blood pressure and hypertrophy of organs (Jung et al., 2022). The disorder has also been associated with uteroplacental ischemia where the flow of blood to the placenta is insufficient to supply the fetus properly. Furthermore, the following risk factors for preeclampsia are related to the maternal condition; obesity, gestational diabetes mellitus and autoimmunity. It also pointed out that diseases such as SARS-CoV-2 may cause preeclampsia or worsen its outcome. Others may be due to syncytiotrophoblast stress, aging of placental cells or failure of maternal- fetal immune tolerance, especially in first time mothers (primipaternity) or mothers who had undergone oocyte donation (Jung et al., 2022).

Preeclampsia can be broadly categorized into two main types: placental preeclampsia and maternal preeclampsia (Nirupama et al., 2021). Placental preeclampsia, often regarded as the more severe variant, stems from irregularities in the formation and function of the placenta. This type is often associated with early-onset preeclampsia (before 34 weeks of gestation) and leads to significant placental insufficiency and poor fetal growth. On the other hand, maternal preeclampsia typically develops later in pregnancy and is associated with maternal conditions like metabolic syndrome or pre-existing hypertension. While both types result in hypertension and potential organ damage, their underlying causes and manifestations differ, influencing clinical management strategies.

The symptoms of preeclampsia can vary widely but often include severe headaches, visual disturbances (e.g., blurred vision or flashing lights), rapid weight gain due to fluid retention, swelling (edema) of the hands and feet, and upper abdominal pain (Jung *et al.*, 2022). These symptoms can be mistakenly attributed to normal pregnancy changes, delaying diagnosis and timely intervention. In clinical practice, preeclampsia is often detected through regular monitoring of blood pressure and urine protein levels during antenatal care visits. However, this approach can be challenging in low-resource settings like Kenya, where limited access to healthcare services and understaffed facilities impedes early diagnosis (Karimi and Gitonga, 2020; Organization, 2024b). Furthermore, preeclampsia can progress to eclampsia, a severe condition characterized by seizures, which poses significant risks to both maternal and neonatal health.

The condition also bears severe consequences on neonatal health. Concerns for this group include preterm birth, low birth weight, small for gestational age (SGA), and possibly higher chance of stillbirth and early neonatal mortality. Another dire consequence of pregnant anxiety is placental abruption, a condition that results in end premature separation of the placenta from the uterine wall, and this poses a risk to the mother and the unborn child (Overton et al., 2022). Moreover, various issues with lifelong impacts are observed in both the mother and child, for example, a woman who has had preeclampsia may be three times more likely to develop cardiovascular diseases at some point in her life, while a child of a woman who has had preeclampsia may experience neurodevelopmental disorders. They are thus important to screen for and manage in order to decrease the risks associated with them, and hence enhance maternal and neonatal outcomes.

## 1.2 Problem Statement

Preeclampsia is a life-threatening hypertensive disorder of pregnancy with significant implications for both maternal and neonatal health. In Kenya, the prevalence of preeclampsia ranges between 5.6% to 6.5%, with early-onset preeclampsia (EO-PE) posing a particularly high risk for adverse outcomes (Ndwiga et al., 2020). Early-onset preeclampsia (before 34 weeks of gestation) is associated with complications such as respiratory distress, birth asphyxia, prolonged maternal hospitalization, and antepartum hemorrhage, which contribute to the high maternal and neonatal mortality rates (Ndwiga et al., 2020). In fact, nearly 44% of preeclamptic cases in Kenya present as EO-PE, and these cases carry five times higher risk of maternal complications compared to those with late-onset preeclampsia (Belay Tolu et al., 2020). Moreover, HELLP syndrome, renal dysfunction, stillbirths, and neonatal deaths are more common in EO-PE cases.

The burden of preeclampsia in Kenya is compounded by limited healthcare infrastructure, particularly in rural areas. Nearly 40% of pregnant women in Kenya do not receive the recommended four antenatal visits, leading to a lack of timely diagnosis of high-risk conditions like preeclampsia (UNICEF, 2019). A recent survey of rural healthcare facilities revealed that only 35% had functional blood pressure monitors, and less than 30% of healthcare staff were trained to detect preeclampsia in its early stages (Kimani et al., 2020). This deficiency contributes to delayed diagnoses, often when preeclampsia symptoms have already escalated, resulting in increased maternal mortality, which currently stands at 530 deaths per 100,000 live births (Karimi & Gitonga, 2020).

Despite ongoing efforts to reduce maternal and neonatal mortality, Kenya's neonatal mortality rate (NMR) remains alarmingly high, with 21 deaths per 1,000 live births in rural areas and 26 per 1,000 in urban areas (UNICEF, 2024). Neonatal deaths, particularly in the first 28 days of life, are often linked to preeclampsia and related complications, including preterm birth, low birth weight, and neonatal sepsis (Paul, 2023). These outcomes are exacerbated by socio-economic disparities, with poorer households having less access to prenatal care, and wealthier households experiencing higher mortality rates due to delayed interventions.

While previous studies have examined the clinical presentations, risk factors, and outcomes associated with preeclampsia, there remains a critical gap in early detection and effective management, particularly in resource-limited settings like Kenya. Conventional detection methods—primarily based on blood pressure and proteinuria measurements—often miss early symptoms or are applied too late in the gestation period to prevent severe complications. Around 43% of maternal deaths related to hypertensive disorders in Kenya are preventable with timely diagnosis and intervention (Ndwiga et al., 2020).

Technological solutions, such as artificial intelligence (AI) and machine learning (ML), offer significant potential to bridge these gaps. Globally, AI-based models have demonstrated the ability to predict preeclampsia with over 90% accuracy (Chalupka & Vizentini, 2020), yet these solutions have been underutilized in Kenya. By analyzing clinical data such as blood pressure, proteinuria levels, and patient demographics, machine learning models can facilitate early diagnosis and enable timely medical intervention. However, no robust AI-powered models have been developed or implemented for preeclampsia detection in Kenya, leaving a substantial gap in the country's maternal healthcare framework.

Therefore, this study aimed to address this gap by developing an AI-powered model tailored to predict preeclampsia in pregnant women in Kenya. This model allowed for early diagnosis, thereby enabling timely medical intervention and ultimately improving maternal and neonatal health outcomes in a country where preeclampsia remained a leading cause of mortality.

## 1.3 Research Objectives

### 1.3.1 Main Objective

The main objective of this research is to develop an AI-powered model to predict preeclampsia in pregnant women in Kenya, enabling early diagnosis and timely medical intervention.

### 1.3.2 Specific Objectives

The specific objectives are as follows:

a) To identify the most relevant risk factors for preeclampsia in Kenyan pregnant women using clinical data and to analyze their significance.

b) To develop machine learning models for predicting the onset of preeclampsia and to test their performance.

c) To evaluate the performance of various AI algorithms and to determine the most accurate and appropriate model for predicting preeclampsia.

d) To deploy the predictive model on multiple platforms, including web, Android, and iOS, ensuring wide accessibility for healthcare providers and patients.

## 1.4 Research Questions

To guide the research, we addressed the following key questions:

a) What are the most relevant risk factors for preeclampsia in Kenyan pregnant women based on clinical data, and how significant are they?

b) How can machine learning models be developed and tested to predict the onset of preeclampsia?

c) How can the performance of various AI algorithms be evaluated to determine the most accurate and appropriate model for predicting preeclampsia?

d) What is the most effective approach for deploying the predictive model on multiple platforms, including web, Android, and iOS, to ensure wide accessibility for healthcare providers and patients?

## 1.5 Significance of the Study

The early diagnosis of preeclampsia using an AI-based questionnaire could ensure a significant decrease in complications and mortality among expectant women and their newborns in the Kenyan setting. This tool would help in the identification of at-risk women during earlier stages of pregnancy to perhaps prevent adverse outcomes and improve women's health by providing suitable medications, increased monitoring, and early initial births where deemed essential. Measures such as these can help prevent the severe consequences of preeclampsia, which results in better maternal and newborn health, and a decrease in the healthcare burden of the country.

This study will make a unique contribution to the emerging literature on the role of AI in improving maternal health outcomes in LMICs, particularly Kenya where preeclampsia continues to be a leading cause of maternal mortality. In this regard, the studies that focused on the development and validation of AI-based PMs specific to the Kenyan setting might provide insights into the appropriate course of clinical support systems that can be efficient, accurate, and practicable. These tools would improve the quality of prenatal by availing data to the health care providers which may help to revolutionize maternal health and immensely boost the maternal mortality and infant mortality ratio.

## 1.6 Scope of Study

The study focused on developing a machine learning model using clinical data from pregnant women in the Coastal region of Kenya, with particular attention to risk factors for preeclampsia. The dataset included variables such as blood pressure, proteinuria levels, body mass index (BMI), maternal age, and previous medical history. The study also explored the practical application of the model in Kenyan healthcare settings, focusing on its integration into routine prenatal care.

# 2 Literature Review

## 2.1 Introduction

This section details an overview of concepts around the subject. It includes a review of literature on neonatal health, preeclampsia, and machine learning models in the underlying context. The underlying concepts are discussed in the subsequent sections.

## 2.2 Maternal and Neonatal Health in the Context of Preeclampsia

Preeclampsia is a hypertensive disorder that complicates pregnancy with a major impact on maternal and neonatal morbidity and mortality, and particularly in developing nations like Kenya. In the worldwide, preeclampsia occurs in 2%-8% of pregnancies, and in developing countries 15% of maternal mortality rate happens with preeclampsia (Nirupama et al., 2021). They defined the condition as hypertension and proteinuria after twenty weeks of gestation with severe complications on both the mother and the fetus such as organ damage, preterm births, stillbirths, and neonatal complications (Khan et al., 2022). Further, the occurrence of preeclampsia also differs according to sociodemographic characteristics of women. Khan et al. (2022) further revealed that preeclampsia was more frequent in women below 24 years of age, those in the low economic status, and poor educational background. In addition, societal factors such as antecedent of hypertension in the family, diabetes in gravidas, and chronic hypertension could predispose pregnant female to preeclampsia. The challenges include high proportion of preterm labors (45.6%), cesarean section (63.4%) and other complications such as renal failure, anemia, pulmonary edema and neonatal death rate was observed to be 11.1%.

The first sign of preeclampsia may be mild to severe; which brings complications like HELLP syndrome, impaired kidney function, breathing difficulties, and baby stunting. According to Bela Tolu et al (2020) early-onset preeclampsia is defined as preeclampsia developing before 34 weeks of gestation and is characterized by severe maternal and perinatal complications as compared to late onset preeclampsia that occurs after 34 weeks of pregnancy. For instance, the early-onset preeclampsia in women increases their risk for complications with OR of 5.22 for maternal complications and 25.91 for perinatal complications. Because of the high risk of this condition, there is a need to identify and manage this condition in the earliest possible time; it also emphasizes the need for clearly defined diagnostic criteria and successful interventions to enhance the maternal and neonatal health status.

## 2.3 Detection and Diagnosis of Preeclampsia

### 2.3.1 Diagnostic Criteria for Preeclampsia

ACOG has provided specific diagnostic criteria for preeclampsia to ensure that healthcare providers are able to diagnose it early and begin the steps towards recognizing a cure. New-onset hypertension is usually involved in the diagnosis of preeclampsia; it has a threshold of SBP of at least 140mmHg or DBP of at least 90mmHg on two occasions, four hours apart after the 20 weeks of pregnancy in a woman who was previously normotensive (Reddy et al 2021). Furthermore, proteinuria is an earmark for diagnosis that can be quantified as ≥ 300mg of protein in a 24-h urine sample or PCR ≥ 0.3 or 2+ with the dipstick if other methods are not feasible. However, in cases where there is no proteinuria, the diagnosis of preeclampsia can be done depending on other related system complications like; thrombocytopenia (platelet count of less than 100,000mm3), renal dysfunction (creatinine > 1.1 mg/dL or a doubled value as compared to the prior value in the absence of kidney disease), hepatic dysfunction, pulmonary edema or any

Studies that have compared diagnostic criteria from various guidelines have shown that newer diagnostic definitions, including those from the ACOG 2018 and the ISSHP 2018, increase the number of women diagnosed with preeclampsia by 6.4% and 14.8% respectively. These broader definitions have implications for mild disease, which may not manifest with the typical signs but present other risks to maternal and neonatal health.

### 2.3.2 Conventional Detection Methods

The current gold standard for detecting preeclampsia includes regular monitoring of blood pressure and urine protein levels during antenatal care. However, these conventional methods have limitations, particularly in low-resource settings like Kenya, where access to healthcare services and skilled professionals is often limited (KNBS and ICF, 2023). In such settings, a significant number of women do not receive timely prenatal screenings, which delays the identification of preeclampsia until symptoms become severe. This delay limits opportunities for early intervention and increases the likelihood of adverse maternal and neonatal outcomes. For example, Takahashi *et al* . (2021) found that among women with preeclampsia, those diagnosed based on uteroplacental dysfunction (a new criterion in the ISSHP guidelines) had higher rates of preterm birth (<37 weeks), fetal growth restriction, and neonatal acidosis compared to those diagnosed by traditional criteria based on proteinuria and hypertension alone.

Additionally, the challenge of diagnosing preeclampsia is compounded by the variability in its clinical presentation. While proteinuria and hypertension are hallmark symptoms, preeclampsia may also present with symptoms like impaired liver function, thrombocytopenia, and neurological symptoms, which can be difficult to identify without advanced testing and expertise (Slade *et al* ., 2024). Slade *et al* . (2024) further emphasized that different thresholds of blood pressure, such as 130/80 mm Hg in certain populations (e.g., women with low body mass index or previous pregnancies without preeclampsia), may be more indicative of preeclampsia risk than the standard 140/90 mm Hg cutoff. This suggests a need for nuanced detection criteria that take into account individual risk factors and characteristics to improve early diagnosis and reduce adverse outcomes.

The limitations of traditional detection methods underscore the need for more comprehensive and accessible diagnostic tools, especially in settings where healthcare resources are constrained. Addressing these gaps with innovative technologies, such as AI and machine learning, could enhance the early detection and management of preeclampsia, ultimately improving maternal and neonatal health outcomes.

## 2.4 Machine Learning and Preeclampsia Detection

### 2.4.1 Overview of Recent Advances in Preeclampsia Prediction

Machine learning (ML) has increasingly been utilized to predict the risk of preeclampsia, leveraging diverse data types such as biomarkers, clinical history, and demographic variables to enable early intervention and improved maternal and neonatal outcomes (Swathikrishna *et al* ., 2024; Aljameel *et al* ., 2024). Given that preeclampsia is a hypertensive disorder with multisystem implications, the use of advanced predictive analytics is crucial in improving diagnosis and management. The public maternal health risk dataset provided by Oslo University Hospital is an example of a comprehensive dataset containing heart rate, blood glucose levels, blood pressure (both diastolic and systolic), body temperature, and other key indicators, making it suitable for predictive modeling. The use of such data helps in developing ML models that are able to predict preeclampsia risk levels with a high degree of accuracy.

### 2.4.2 Types and Aspects of Machine Learning Models in Prediction

Many models of machine learning have been applied to predict preeclampsia, from the supervised learning model to the ensemble learning model (Swathikrishna et al ., 2024; Aljameel et al ., 2024). Popular supervised learning models include Naïve Bayes (NB), Logistic Regression (LR), Support Vector Machines (SVM), Decision Trees, and K-Nearest Neighbors (KNN) since they tend to be more understandable and easier to implement. However, Random Forest, AdaBoost, XGBoost, GBT, and other ensemble models have given higher performance in almost all clinical prediction models because of averaging multiple classifiers to fine-tune the predictive performance and metal variance (Ranjbar et al., 2024). In addition, Supervised Machine Learning techniques like Artificial Neural Networks (ANN) and Long Short-Term Memory (LSTM) networks have shown the propensity of modeling complex and non-linear relationships, therefore ideal to model high-dimensional health care datasets (Aljameel et al., 2024).

In practice, Random Forest (RF) has emerged as one of the top-performing models for predicting preeclampsia, as it can handle large feature sets and complex interactions effectively (Swathikrishna *et al* ., 2024). According to Schmidt *et al* . (2022), gradient-boosted tree models, which include both XGBoost and Gradient Boosted Machines (GBMs), have also demonstrated high predictive performance, with high specificity and sensitivity in preeclampsia risk prediction. The chart above indicates the popularity and performance of various ML models, with RF, SVM, and ANN being the most frequently used and successful in preeclampsia prediction. Figure 2.1 shows an ML predictive model flowchart.

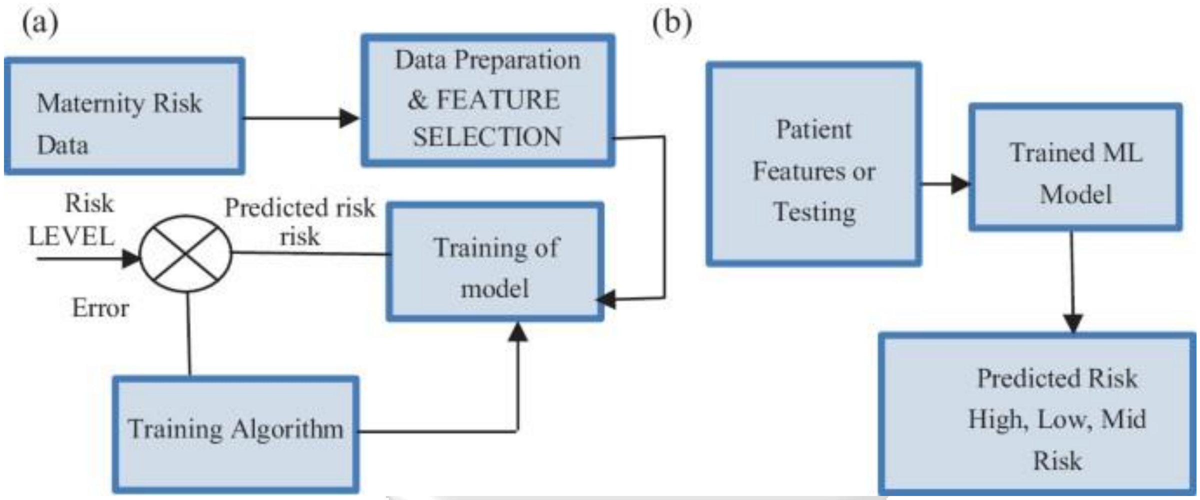


**Figure 2.1: ML Prediction Model for Preeclampsia**

### 2.4.3 Best Models and Their Performance

Recent research has highlighted several models with high predictive capabilities for preeclampsia. For instance, Random Forest and Gradient-Boosted Tree (GBT) models have achieved high accuracy rates, with RF models reaching up to 96.39% accuracy in predicting preeclampsia (Swathikrishna *et al* ., 2024). Schmidt et al. (2022) reported robust performance for both a gradient-boosted tree model and an RF classifier, with overall accuracy near 89%, positive and negative predictive values at 88% and 89%, respectively, and AUCs ranging from 0.82 to 0.97.

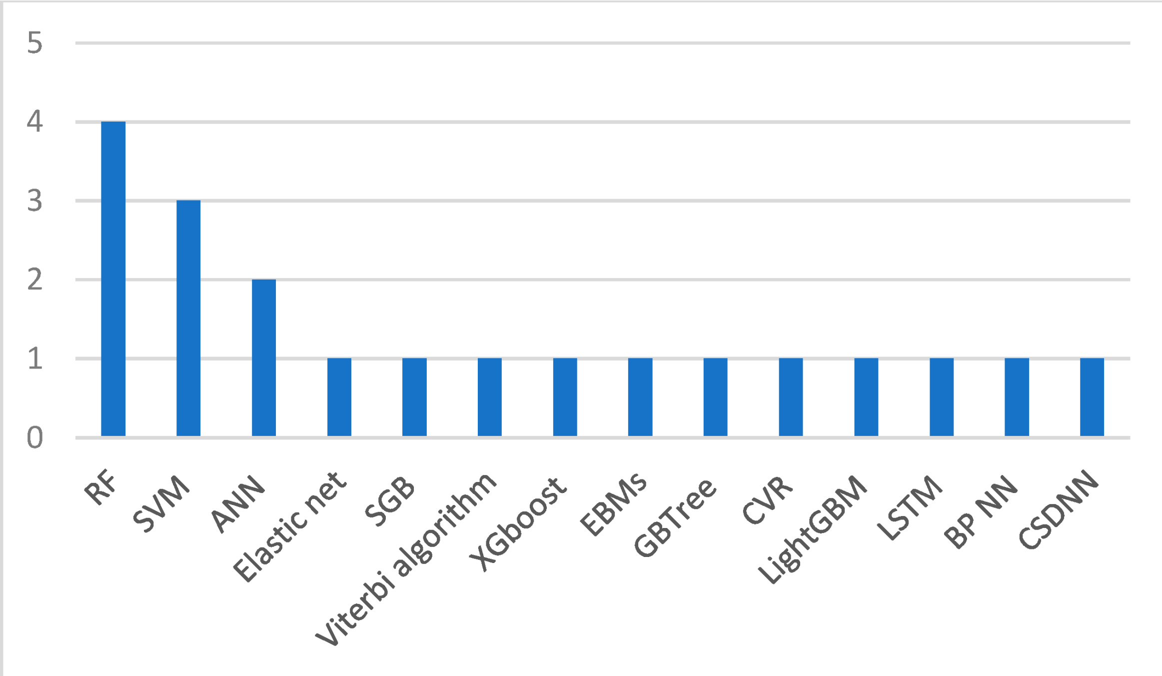


**Figure 2.2: Machine Learning Models Comparison**

The systematic review by Ranjbar et al. (2024) stated that the models as Elastic Net, Stochastic Gradient Boosting (SGB), Extreme Gradient Boosting (XGBoost), and Random Forest were some of the best for preeclampsia propagation. Specifically, AUC, sensitivity, specificity, precision, recall, and F1 score, were used for evaluation purposes, with observed AUC ranging between 0.860 and 0.973 across models and datasets. They are especially important in understanding the model's capacity to correctly identify high risk patients, along with having minimal false positive and false negative cases.

### 2.5 Summary of Literature and Gaps

The literature review of studies on using machine learning types for identifying preeclampsia has revealed important developments in the accurate prediction of the disease with the aid of its risks, and has presented the discussed models, including RF, SVM, ANN, and GBT, as perspectives with impressive accuracy and predictive outcomes (Ranjbar et al., 2024; Swathikrishna et al., 2024; Schmidt et al., 2022). These models promisingly improve the application of biomarkers, clinical history, and demographics to augment the early identification of preeclampsia. However, there have been some important limitations in the literature, especially focusing on the use of those models in low-resource countries like Kenya.

One key gap has been the lack of prospective validation of ML models in diverse populations. Many existing models have been developed and tested in high-resource environments, where healthcare systems are better equipped, and data quality is more reliable. In Kenya, however, disparities in healthcare access, data availability, and infrastructure posed significant challenges that existing models may not have been able to address effectively (Schmidt et al., 2022). The proposed research aimed to fill this gap by validating the predictive model in a Kenyan context, using data from both rural and urban settings to ensure that the model was adaptable to the country's diverse healthcare environments.

Another notable gap was the limited understanding of how to incorporate dynamic data—such as changes in biomarkers over time—into the predictive models. In many low-resource settings, including Kenya, regular monitoring of clinical indicators like blood pressure and proteinuria has often been inconsistent due to limited access to antenatal care. Existing models largely relied on static data points, which may not have been sufficient for accurate risk prediction in such contexts (Swathikrishna et al., 2024). This research explored how to incorporate both retrospective and prospective clinical data to create a more dynamic and adaptable model, addressing the specific challenges posed by intermittent healthcare access.

Additionally, there has been a lack of interoperability between current ML models and existing healthcare systems in low-resource settings. Most ML models have been designed for integration with advanced electronic health record (EHR) systems, which are often unavailable in rural Kenyan healthcare facilities (Ranjbar et al., 2024). The proposed model was tailored to function in environments with minimal technological infrastructure, ensuring that it could be deployed even in areas where EHR systems were lacking or incomplete. By focusing on ease of integration into the healthcare workflow and ensuring the model could function offline, this study addressed the gap in practical, scalable solutions for preeclampsia detection in Kenya.

In summary, the proposed research filled the existing gaps by validating ML models in the Kenyan healthcare context, incorporating dynamic data to improve predictive accuracy, and ensuring the model's interoperability with low-resource healthcare systems. This ensured that the proposed model could be used effectively to improve maternal healthcare outcomes in Kenya.

# 3 Methodology

## 3.1 Research Design

This study utilized machine learning (ML) to develop a predictive model for the early detection of preeclampsia in pregnant women in Kenya, with a focus on improving maternal and neonatal health outcomes. Early identification of at-risk women was critical for timely intervention and management, which mitigated the adverse impacts of preeclampsia, such as maternal mortality, preterm births, and neonatal complications (Nirupama et al., 2021; Khan et al., 2022). The use of ML in this context offered the ability to analyze complex clinical data and identify patterns that traditional methods may have missed, thereby enhancing diagnostic accuracy and enabling risk-based maternal care. This aligned with efforts to address public health challenges in low-resource settings, where disparities in healthcare access could lead to delayed diagnosis and increased risks for pregnant women (Swathikrishna et al., 2024).

The research was situated in the Coastal region of Kenya, targeting hospitals across Kwale, Kilifi, and Mombasa counties to encompass a diverse mix of rural and urban populations. This setting provided a comprehensive dataset reflective of various socio-economic conditions and healthcare access disparities, crucial for developing a robust and generalizable predictive model. The study followed a quantitative research design, utilizing retrospective and prospective clinical data to build and validate machine learning algorithms. By focusing on key clinical variables such as demographics, blood pressure, proteinuria levels, and pre-existing conditions, the predictive model aimed to improve upon current detection methods. This approach supported the broader objective of reducing maternal and neonatal mortality through early detection, contributing to Sustainable Development Goal 3's target of lowering maternal mortality rates (KNBS and ICF, 2023; Schmidt et al., 2022).

## 3.2 Study Population and Sampling

The study targeted pregnant women attending antenatal care services in health facilities located within Kwale, Kilifi, and Mombasa counties. These counties were strategically chosen due to their high incidence of preeclampsia cases, representing a mix of rural and urban populations. This diversity ensured the inclusion of varied socio-economic and healthcare access conditions, enhancing the generalizability of the study's findings.

Kwale and Kilifi counties predominantly consisted of rural populations, where healthcare infrastructure was often constrained. This setting provided an opportunity to assess the model's applicability in low-resource environments, addressing healthcare disparities in such regions. Mombasa County, on the other hand, represented an urban setting with relatively better healthcare infrastructure, enabling the inclusion of a population with improved access to antenatal care. The combination of these settings offered a comprehensive perspective on the risk factors and patterns of preeclampsia in different socio-economic and healthcare contexts.

To systematically select participants, the study used a stratified random sampling method. Health facilities were stratified based on their level (e.g., public hospitals, private clinics, and community health centers). Within each stratum, systematic random sampling was employed to select pregnant women attending antenatal clinics during the study period. This approach ensured fair representation across different facility types and minimized selection bias.

### 3.3 Data Collection

Data collection for this study employed a structured questionnaire as the primary tool, supplemented by retrospective and prospective clinical data. The structured questionnaire was designed to gather comprehensive demographic, socio-economic, and medical information from participants, ensuring consistency and comparability across all study sites. It included sections on age, parity, pre-existing medical conditions, and socio-economic factors, providing essential inputs for the machine learning model. By utilizing a fixed format, the questionnaire minimized variability in data collection and ensured that all relevant variables were systematically captured.

The structured questionnaire was particularly well-suited for this study due to its ability to standardize data collection processes across different settings, including rural and urban health facilities. It was cost-effective, time-efficient, and easy to administer, making it an ideal choice for a study conducted in a resource-constrained environment like Kenya. To ensure its reliability and validity, the questionnaire was pilot tested in a subset of facilities, and feedback from this phase guided refinements to improve clarity and relevance.

For participants unable to write, speak, or otherwise directly engage with the data collection process due to physical or cognitive limitations, the study team adopted alternative methods to ensure inclusivity. Where participants could not physically complete the questionnaire, trained data collectors administered the questionnaire verbally, recording the responses on their behalf. If participants could not speak or otherwise express their responses, information was gathered through proxy respondents, such as close family members or caregivers, who were familiar with the participant's demographic and medical history. This process was implemented with the utmost sensitivity, and informed consent or assent was still obtained from participants or their legal representatives to ensure ethical compliance.

In addition to the structured questionnaire, clinical data was collected from medical records. Retrospective data, covering cases from 2020 to the present, was extracted to provide a historical perspective on preeclampsia incidence and risk factors. Prospective data was collected during antenatal visits, where healthcare providers recorded clinical parameters such as blood pressure, proteinuria levels, and gestational age at diagnosis. This real-time data collection complemented the questionnaire responses, creating a robust dataset for analysis.

Given the variability in healthcare infrastructure across the selected counties, several challenges were anticipated, such as incomplete records and variations in data quality. To address these challenges, a rigorous data validation and cleaning process was implemented. Missing values were addressed using multiple imputation techniques to estimate absent data based on observed patterns, minimizing bias in the analysis. In cases of severe data incompleteness or irrelevance, such records were excluded to maintain the integrity of the dataset.

To ensure accuracy during data collection, all personnel involved underwent comprehensive training on administering the questionnaire and extracting clinical data. For facilities lacking digital medical records, paper-based records were digitized manually by trained staff to ensure consistency. Standardized protocols were established for recording clinical variables, ensuring uniformity across all study sites.

Furthermore, a pilot phase was conducted in a subset of health facilities to assess the feasibility of the data collection process. This phase helped identify potential logistical and methodological challenges early on, allowing for adjustments before full-scale implementation. Regular data availability and completeness checks were performed throughout the study period to maintain high data quality and reliability.

By integrating a structured questionnaire with retrospective and prospective clinical data and ensuring that inclusive methods were in place for participants with physical or cognitive limitations, the study aimed to build a comprehensive and high-quality dataset. This approach ensured that both static and dynamic risk factors for preeclampsia were captured, providing the necessary inputs for developing an accurate and reliable machine learning model.

### 3.3.1 Inclusion and Exclusion Criteria

The inclusion and exclusion criteria for this study are designed to ensure the selection of participants who are most relevant to the research objectives while maintaining the integrity and generalizability of the findings. Inclusion criteria will focus on pregnant women attending antenatal care services in health facilities within Kwale, Kilifi, and Mombasa counties, regardless of their socio-economic background. Participants must be at least 18 years old and provide informed consent to participate in the study. Both retrospective and prospective data will include women with documented antenatal records detailing essential clinical variables such as blood pressure, proteinuria levels, and gestational age. Women who are in their second or third trimester of pregnancy, where the risk of preeclampsia is more pronounced, will be prioritized for inclusion in the study.

Exclusion criteria will eliminate individuals and records that could compromise the study's validity or applicability. Women with incomplete or inconsistent medical records will be excluded to ensure the dataset is robust and reliable for machine learning model development. Participants who do not provide informed consent or who have medical conditions unrelated to pregnancy, such as chronic illnesses or terminal conditions, that could confound the results, will also be excluded. Furthermore, retrospective records lacking key variables like blood pressure and proteinuria levels will not be considered, as they would hinder the predictive capabilities of the machine learning model.

These criteria ensure that the study focuses on a population at risk for preeclampsia while minimizing confounding variables that could bias the results. By clearly defining who is included and excluded, the study establishes a foundation for building a reliable and accurate predictive model tailored to the Kenyan healthcare context. This systematic approach also enhances the reproducibility of the research and its potential applicability in similar low-resource settings.

### 3.4 Study Variables

The study examined both independent and dependent variables to predict the onset of preeclampsia using clinical and demographic data. The independent variables include factors that are hypothesized to influence the development of preeclampsia, while the dependent variable is the actual occurrence of the condition.

**Table 3.1: Study Variables**

| Variable Type | Variable Name | Description | Measurement/Scale |
|---|---|---|---|
| Independent | Age | Age of the pregnant woman | Years (Continuous) |
| | Weight | Pre-pregnancy weight or weight during first trimester | Kilograms (Continuous) |
| | Parity | Number of previous pregnancies carried to viable gestational age | Count (Ordinal) |
| | Blood Pressure | Systolic and diastolic measurements taken during pregnancy | mmHg (Continuous) |
| | Proteinuria Levels | Presence of protein in urine, measured through urinalysis | g/24 hours or mg/dL (Continuous) |
| | Gestational Age at Diagnosis | The week of pregnancy when preeclampsia is diagnosed | Weeks (Continuous) |
| | Hypertension | History of hypertension before pregnancy | Yes/No (Binary) |
| | Diabetes | History of diabetes or gestational diabetes | Yes/No (Binary) |
| | Socioeconomic Status | Income level, education, or occupation of the pregnant woman | Categorical (Low/Medium/High) |
| | Previous Preeclampsia | History of preeclampsia in prior pregnancies | Yes/No (Binary) |
| Dependent | Preeclampsia Status | Whether the patient develops preeclampsia | |

| | |
|---|---|
| Status | preeclampsia |

The study's independent variables encompassed a range of demographic and clinical characteristics linked to preeclampsia risk, such as maternal age, weight, and parity, which are key demographic factors influencing pregnancy outcomes. Clinical measures, including blood pressure (systolic and diastolic) and proteinuria levels, provide objective indicators for detecting changes in maternal health related to preeclampsia, while the gestational age at diagnosis serves as a marker for symptom onset. Pre-existing conditions like hypertension and diabetes are also considered, as they significantly elevate the risk. Socioeconomic status and a history of preeclampsia were included to explore the effect of social determinants and previous pregnancies. The dependent variable, preeclampsia status, will be coded as a binary outcome (0 = no preeclampsia, 1 = preeclampsia), with the objective to assess how these independent variables contribute to the predictive modeling for early diagnosis and timely intervention.

## 3.5 Data Preprocessing

### 3.5.1 Data Cleaning and Transformation

To ensure the dataset was ready for analysis, a thorough cleaning process was carried out. Handling missing values was a crucial step, where imputation techniques were employed to estimate and replace absent data based on observed patterns. This approach maintained data integrity and prevented bias in model predictions. Outliers were identified using statistical measures and addressed appropriately to prevent their undue influence on model training. For categorical data, encoding methods such as one-hot encoding or label encoding were used to transform the categories into numerical values, making them compatible with machine learning algorithms.

### 3.5.2 Feature Scaling and Normalization

The continuous variables were then scaled or normalized after cleaning the data to bring all the features onto the same scale to eliminate large feature values affecting naive models. This process was very critical for algorithms such as Support Vector Machines (SVM) and Gradient

Boosting Machines (GBM), which could be affected by the scale of the data the variables may represent (Sahin, 2020). Shantal et al. (2023) noted that depending on the distribution and nature of data used in transforming features, one may use standardization (i.e., ‘z-score normalization’) or element-wise transformation, such as min-max scaling, for instance when features need to be within a scale of zero to one.

Further, the data was divided into training and testing sets, which was a conventional procedure; the training set was generally bigger, with around 80% of all the data, so that the model could learn from it, while the testing data, made up of the remaining 20%, was used for testing the efficiency of the model. This approach helped the model’s capacity to generalize on new unseen data and increased its predictive performance (Brownlee, 2018).

### 3.6 Model Development

In developing a robust prediction model for preeclampsia, several machine learning algorithms were explored, each offering unique advantages in classification tasks.

#### 3.6.1 Logistic Regression

This is a simple yet effective algorithm for binary classification. It models the probability that a given input belongs to a particular class using a logistic function. Its interpretability and low computational cost make it a strong baseline model.

#### 3.6.2 Decision Trees

These algorithms partition the data into subsets based on feature values. Trees are easy to interpret and can handle both categorical and numerical data, but they are prone to overfitting, especially with complex datasets.

#### 3.6.4 Random Forests

An ensemble of decision trees, random forests improve prediction accuracy by aggregating the results of multiple trees and averaging their predictions. This reduces overfitting and enhances model stability.

#### 3.6.5 Gradient Boosting Machines (GBM)

GBM models sequentially build decision trees, focusing on correcting errors from previous models. They are highly accurate but can be sensitive to overfitting and require careful hyper parameter tuning.

#### 3.6.6 Support Vector Machines (SVM)

Support Vector Machines are a classification technique that identifies an optimal hyperplane to separate different classes within the dataset. Particularly effective in handling high-dimensional data, SVMs excel at creating clear decision boundaries. However, their computational demands can make them less suitable for extremely large datasets due to increased complexity.

#### 3.6.7 Neural Networks

Neural networks operate through layers of interconnected nodes, or "neurons," to model intricate patterns and relationships within data. Renowned for their adaptability and ability to capture nonlinear interactions, neural networks require substantial computational resources and large datasets to perform effectively, making them best suited for complex predictive tasks.

### 3.7 Model Selection and Justification

This study employed a variety of machine learning models to predict the onset of preeclampsia in pregnant women, ultimately prioritizing the Random Forest algorithm. The selection of Random Forest was grounded in its robustness, capability to process diverse data types, and inherent resistance to overfitting. These attributes are particularly crucial when dealing with the heterogeneity and complexity of clinical data, ensuring more reliable predictions in this critical healthcare context (Sun et al., 2020).

One of the key advantages of Random Forest is its ability to manage complex datasets that include both categorical and continuous variables. In this study, the data comprises variables such as maternal age, blood pressure, proteinuria levels, and socioeconomic status, all of which Random Forest can process effectively (Brownlee, 2018). Additionally, Random Forest provides inherent feature importance rankings, which will allow the research team to identify the most significant predictors of preeclampsia. This feature is particularly valuable in medical research, where understanding which variables most influence the onset of a condition like preeclampsia can guide both clinical decision-making and future research (Sun et al., 2020).

A further advantage of Random Forest is its robustness in handling incomplete data. In settings like rural Kenya, where healthcare infrastructure may be underdeveloped and medical records may be incomplete, the ability to manage missing data without significant degradation in model performance is critical (Schmidt et al., 2022). Random Forest can handle these challenges more effectively than many other machine learning algorithms, ensuring that the model remains reliable even in the presence of data gaps.

Moreover, Random Forest's resistance to overfitting is another significant benefit. In clinical datasets, where certain variables may disproportionately influence the outcome, avoiding overfitting is essential to ensure the model's utility in real-world healthcare settings. Random Forest achieves this by aggregating the results of multiple decision trees and averaging their predictions, thereby enhancing both prediction stability and model accuracy (Sun et al., 2020).

While Random Forest is the primary model for this study, other algorithms were considered. Logistic regression, for example, is a simple and interpretable model commonly used in binary classification tasks. However, it is limited in its ability to handle non-linear relationships between variables, which are likely present in the complex clinical datasets used for preeclampsia prediction (Ranjbar et al., 2024). Similarly, decision trees, though interpretable and capable of handling mixed data types, are prone to overfitting, particularly in large datasets with multiple interacting variables (Sahin, 2020). Random Forest addresses these limitations by using an ensemble of decision trees, which improves the stability and predictive power of the model.

GBM is known for its high accuracy in classification tasks, but it is more computationally intensive and requires careful tuning to avoid overfitting. Given the resource constraints in rural Kenyan healthcare settings, the computational complexity of GBM makes it less practical for this study compared to Random Forest (Aljameel et al., 2024). On the other hand, SVM performs well in high-dimensional spaces but struggles with large datasets and is sensitive to the scaling of features, making it less adaptable for clinical variables such as those in this study (Schmidt et al., 2022).

Neural networks, another powerful machine learning algorithm, offer flexibility and can model complex, non-linear relationships. However, they require large datasets and significant computational resources, which may not be available in the low-resource settings targeted in this study. Additionally, neural networks are often criticized for their lack of interpretability, which is a crucial consideration in healthcare applications where understanding the rationale behind predictions is vital (Aljameel et al., 2024).

Gini Impurity (for Node Splitting)

$$Gini(D) = 1 - \sum_{i=1}^{c} p_i^2$$

D: Dataset at a node c:

Number of classes

$p_i$: Proportion of samples in class

Ensemble Prediction

For classification, the final prediction y^\hat{y}y^ is based on the majority vote acrossall decision trees:

$$\hat{y} = \arg\max_k \sum_{j=1}^{T} I(h_j(x) = k)$$

$h_j(x)$

Prediction

from the $j$-th tree for

input x k: Possible

class label

$I(\cdot)$: Indicator function (1 if true, 0 otherwise)

$T$: Total number of trees in the forest

**Feature Importance (Gini Index Reduction)**

Random forests provide feature importance based on the decrease in Gini impurity across all trees:

$$Importance(f) = \frac{1}{T}\sum_{j=1}^{T} \Delta Gini_j(f)$$

$f$: Feature

$\Delta Gini_j(f)$: Decrease in Gini impurity for feature $f$ in tree $j$

Random Forest was the optimal model for this study due to its ability to balance interpretability, accuracy, and computational efficiency. Its capacity to handle imbalanced datasets, expected in this study given the varying incidence rates of preeclampsia across different regions, further supports its selection (Swathikrishna et al., 2024). This combination of strengths makes Random Forest particularly suitable for use in low-resource settings like Kenya, where healthcare infrastructure and computational resources are limited, but accurate and early detection of preeclampsia is crucial for improving maternal and neonatal health outcomes.

### 3.8 Model Evaluation

Each model's performance was assessed using a range of metrics to gauge its accuracy and reliability. The objective was to identify the model that consistently excels across these metrics, ensuring dependable and clinically relevant predictions for the early identification of preeclampsia. The key evaluation metrics include:

#### 3.8.1 Accuracy

Accuracy measures the proportion of correct predictions made by the model over the total predictions. It is calculated as:

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN}$$

#### 3.8.2 Sensitivity (Recall)

Sensitivity, also known as Recall, measures a model's effectiveness in accurately identifying true cases of preeclampsia. This metric is especially crucial in medical contexts, where ensuring that all actual cases are detected is of paramount importance. The formula is:

$$Recall = \frac{TP}{TP + FN}$$

High recall indicates that the model can effectively detect most instances of preeclampsia, minimizing the risk of false negatives (missed diagnoses).

#### 3.8.4 Specificity

Specificity measures the model's ability to correctly identify non-cases (pregnancies without preeclampsia). It is defined as:

$$Specificity = \frac{TN}{TN + FP}$$

### 3.8.5 Precision

Precision on the other hand evaluates the number of the positive classification that is truly positive out of all the positive classification done by the model. It is particularly crucial in contexts where false positives are costly: A high precision score means that in cases the model is predicting preeclampsia, it is mostly likely to be accurate rather than being a false positive.

$$Precision = \frac{TP}{TP + FP}$$

### 3.8.6 F1 Score

The **F1 Score** provides a harmonic mean between precision and recall, offering a balanced measure when both false positives and false negatives are important. It is calculated as:

$$F1Score = 2 \cdot \frac{Precision \cdot Recall}{Precision + Recall}$$

### 3.8.7 Model Selection

After comparing each model according to the above criteria individual precision, recall and AUC-ROC the final best model were chosen in favor of preeclampsia. Throughout the evaluation process cross-validation was used to guarantee that the model is not overfitting and thus can perform well on unseen data, making it a credible tool to assist clinical decision.

### 3.9 Model Deployment

The deployment of the predictive model focused on making the tool accessible and functional in both rural and urban healthcare facilities in Kenya. The model would deployed on mobile platforms (web, Android, and iOS), ensuring accessibility through smartphones and tablets. The tool will integrate with existing workflows, including electronic health records (EHR) systems where available, and will support manual data entry in facilities without EHRs. Healthcare workers, particularly in rural areas, will receive training on both the technical use of the mobile platform and the clinical interpretation of the model’s outputs. This will ensure the model supplements, rather than replaces, clinical decision-making. Ongoing support and refresher training will be provided to maintain competence over time.

To overcome limited access to reliable internet and hardware in rural areas, the tool functions offline and sync data when an internet connection is available. Low-cost mobile devices would be provided to facilities in need, and the app would be optimized to run on devices with limited processing power.

The model's sustainability would be ensured through partnerships with health ministries and international organizations. The goal is to scale the tool beyond the initial study locations,

Challenges such as healthcare workers’ unfamiliarity with AI and potential disparities in data quality will be addressed through awareness campaigns and thorough testing of the model in both rural and urban settings to ensure broad applicability.

In summary, the deployment strategy focuses on ensuring accessibility, proper training, addressing infrastructure issues, and creating a sustainable tool that can improve maternal health outcomes in Kenya.

### 3.10 Ethical Considerations

The ethical implications of this study were significant, particularly given the sensitive nature of patient data and the potential risks associated with its misuse. To address these concerns, strict measures were implemented to ensure that participant confidentiality was upheld, informed consent was appropriately obtained, and benefits were fairly distributed.

Research clearance was sought from the National Commission for Science, Technology and Innovation (NACOSTI) in Kenya, and ethical approval was obtained from the Institutional Ethics Review Committee (IERC) at Strathmore University. These approvals ensured that the study adhered to international ethical standards, including the Declaration of Helsinki, and complied with Kenya's Data Protection Act of 2019. These steps guaranteed that the research was conducted responsibly and that participant rights were safeguarded.

To address the challenges posed by varying literacy levels among participants, the informed consent process was designed to accommodate those incapable of reading or signing a written consent form. Consent forms were translated into Swahili and simplified to enhance their readability and understandability. For participants who could not read, the study team provided verbal explanations of the research's purpose, potential risks and benefits, and the rights of participants in their preferred language. Verbal consent was documented in the presence of an impartial witness, and thumbprints were used for participants unable to sign, ensuring that all participants could give informed and voluntary consent. The process emphasized the voluntary nature of participation, assuring individuals that they could withdraw from the study at any time without consequences.

Participant privacy and confidentiality were prioritized. Personal identifying information (PII) was anonymized by assigning unique codes to participants. Data was securely stored on encrypted servers, with access restricted to authorized personnel only. Compliance with Kenya's Data Protection Act of 2019 ensured that data was managed responsibly and shared only under strict privacy protocols.

# 4: Results

## 4.1 Introduction

This chapter presents the results of the machine learning models developed to predict the risk of preeclampsia among pregnant women in Kenya. The primary aim of the study was to leverage machine learning techniques to enhance early detection of preeclampsia, a critical condition that contributes significantly to maternal and neonatal morbidity and mortality. Building on clinical and demographic data collected from health facilities in the coastal region of Kenya, various models were trained, evaluated, and compared to identify the most accurate and reliable predictor.

The model development process involved preprocessing the dataset, handling encoding and feature selection, and splitting the data into training and testing sets. Five individual models—Logistic Regression, Decision Tree, Random Forest, Support Vector Machine (SVM), and XGBoost—were trained using this data. Each model's performance was evaluated based on key classification metrics such as accuracy, precision, recall, F1-score, and AUC-ROC, which are particularly important in healthcare settings where both false positives and false negatives have critical implications.

In addition to evaluating standalone models, this chapter also presents the results of a hybrid machine learning approach using a Stacking Classifier. This ensemble method combines the strengths of the best-performing individual models to enhance predictive power and generalizability. The stacked model was subjected to cross-validation to ensure robustness and to avoid overfitting, and its performance was compared directly with that of the individual models.

The chapter begins with a descriptive analysis of the dataset, followed by a comprehensive evaluation of each model. Comparative tables and visualizations such as confusion matrices and ROC curves are used to provide a clear picture of model effectiveness. The findings from these analyses form the basis for determining the most suitable model for real-world deployment in the Kenyan healthcare context.

By presenting both the individual and hybrid model results, this chapter lays the foundation for the discussion in Chapter 5, where the implications of these findings are interpreted in light of existing literature, healthcare priorities, and practical deployment challenges in low-resource settings.

### 4.2 Relevant Risk Factors for Preeclampsia

This section presents the findings corresponding to the first objective of the study, which aimed to identify the most relevant clinical and demographic risk factors associated with preeclampsia in Kenyan pregnant women. Using the full dataset of 2,925 participants drawn from health facilities in the coastal region of Kenya, a detailed descriptive statistical analysis was undertaken to examine central tendencies, variability, and distribution patterns of the key features. The seven core predictors—maternal age, pre-pregnancy weight, parity, proteinuria levels, systolic blood pressure, diastolic blood pressure, and preeclampsia outcome—were selected based on their established clinical significance in previous literature (see Chapter 2) and their potential to inform early risk stratification models.

Table 4.1 provides a summary of the descriptive statistics for these core variables. The goal of this analysis was twofold: (1) to understand the baseline distribution of potential risk factors in the study population, and (2) to preliminarily assess which factors are more prevalent or skewed in ways suggestive of heightened preeclampsia risk.

**Table 4.1: Descriptive Statistics of the Core Variables Used in the Model**

| Statistic | Age | Pre-pregnancy Weight (kg) | Parity | Proteinuria Levels (g/24 hours or mg/dL) | Systolic BP | Diastolic BP | Preeclampsia Outcome |
|---|---|---|---|---|---|---|---|
| Count | 2925 | 2925 | 2925 | 2925 | 2925 | 2925 | 2925 |
| Mean | 28.25 | 75.488 | 2.016 | 1.163 | 130.105 | 82.28 | 0.275 |
| Std Dev | 6.121 | 11.264 | 1.49 | 1.053 | 17.344 | 12.396 | 0.447 |
| Min | 18 | 50 | 0 | 0 | 100 | 60 | 0 |
| 25% (Q1) | 23 | 66.4 | 1 | 0.3 | 115 | 70 | 0 |
| Median (Q2) | 28 | 75.7 | 2 | 1 | 131 | 80 | 0 |
| 75% (Q3) | 33 | 84.4 | 3 | 2 | 145 | 91 | 1 |
| Max | 40 | 100 | 5 | 3 | 165 | 108 | 1 |

### 4.2.1 Age Distribution and Risk Implications

The average maternal age was 28.25 years (SD = 6.12), with a range from 18 to 40 years. Notably, 50% of participants fell between the ages of 23 and 33 (Q1–Q3). This age range reflects the peak reproductive period in Kenya and is consistent with national fertility trends. However, literature indicates that maternal age below 20 or above 35 is linked with increased preeclampsia risk due to immature or declining vascular and immunological function (Adedeji et al., 2020). The inclusion of a wide reproductive age band enhances the generalizability of model findings and suggests that age, while moderately predictive, must be evaluated alongside other stronger indicators.

### 4.2.2 Pre-pregnancy Weight as a Dominant Clinical Predictor

Pre-pregnancy weight exhibited a mean of 75.49 kg (SD = 11.26), with values ranging from 50 to 100 kg. Over 50% of the study population had weights above 75 kg, indicating a high representation of overweight or obese pregnant women. This finding is consistent with global and regional studies which have consistently shown that higher pre-pregnancy BMI is one of the strongest predictors of preeclampsia due to its association with chronic inflammation, endothelial dysfunction, and metabolic syndromes (Moll et al., 2022; see Chapter 2). In this cohort, the interquartile range (66.4–84.4 kg) supports the observation that a substantial segment of women were already in weight categories linked to elevated risk.

### 4.2.3 Parity Patterns and Susceptibility

The mean parity was 2.01, ranging from 0 (primigravida) to 5. The interquartile range (1–3) shows that most participants had some reproductive history. However, approximately 25% of participants were primigravida (parity = 0), a known risk group for preeclampsia due to immunological naivety to paternal antigens and the body's first-time adaptation to pregnancy-related vascular changes. These findings echo similar patterns reported in Kenyan settings by Onyango et al. (2022), who found primigravidity to be an independent risk factor in rural maternal health clinics.

### 4.2.4 Proteinuria Levels as a Confirmatory Marker

Proteinuria, one of the hallmark diagnostic criteria for preeclampsia, had a mean of 1.16 g/24hrs, with values ranging up to 3 g/24hrs. Nearly 50% of the study participants presented with values above 1.0, indicating a skewed distribution toward higher protein excretion. This pattern points to possible renal dysfunction already manifesting in a substantial subset of the population. Given that preeclampsia is primarily diagnosed through elevated blood pressure and proteinuria, the prominence of this marker within the dataset confirms its predictive utility. It also supports its retention as a key feature in all models developed in subsequent stages.

### 4.2.5 Blood Pressure Elevations Across the Sample

The dataset revealed a mean systolic blood pressure (SBP) of 130.1 mmHg and mean diastolic blood pressure (DBP) of 82.28 mmHg. These averages sit on the borderline of clinical concern, with 25% of participants recording SBP above 145 mmHg and DBP above 91 mmHg. This indicates that a notable proportion of the population exhibited hypertensive tendencies, reinforcing hypertension's centrality in preeclampsia diagnosis. As seen in prior studies (Wainaina et al., 2021), early elevations in blood pressure—even if subclinical—are important red flags in maternal screening. The inclusion of SBP and DBP as continuous variables in model development enabled nuanced differentiation between normotensive and pre-hypertensive cases.

The target variable—Preeclampsia Outcome—had a mean of 0.275, suggesting that 27.5% of participants were diagnosed with preeclampsia. This indicates a high prevalence in the studied cohort, far above the global average of 5–8%, and even higher than previous estimates in similar Kenyan populations (WHO, 2021). This elevated incidence underscores the urgency of predictive tools tailored to local risk profiles. The relatively balanced class distribution, while slightly skewed, was advantageous for model training and evaluation as it allowed sufficient representation of both positive and negative cases without the need for aggressive resampling techniques.

Taken together, the descriptive statistics in Table 4.1 strongly point to pre-pregnancy weight, elevated blood pressure, and proteinuria levels as the most prominent and clinically significant risk factors for preeclampsia in this cohort. While maternal age and parity offer supporting insights, their predictive strength is comparatively lower, consistent with trends documented in global and regional literature (Mahama et al., 2021; Mellor et al., 2019).

These insights directly informed the feature selection process for machine learning models developed in later stages and laid the groundwork for training algorithms that are both context-specific and clinically interpretable. Furthermore, the findings validate prior theoretical assertions (see Chapter 2, Section 2.4) that emphasize the multifactorial nature of preeclampsia and the need to consider both metabolic and hemodynamic variables when predicting its onset.

In summary, the descriptive analysis aligned with Objective 1 has revealed critical insights into the prevalence and distribution of known risk factors in Kenyan pregnant women. The prominence of weight, blood pressure, and proteinuria suggests that early screening programs and intervention strategies should prioritize these variables, especially in resource-limited settings where comprehensive lab workups may not always be feasible. These findings support the next phase of the study—building predictive models that are both accurate and grounded in the realities of Kenya's maternal health landscape.

### 4.3 Development and Testing of Machine Learning Models for Preeclampsia Prediction

To address the second objective of this study—to develop machine learning models for predicting the onset of preeclampsia and to test their performance—a structured approach was employed, beginning with meticulous data preprocessing, followed by iterative model development, and concluding with performance testing using standardized classification metrics. This section presents the complete pipeline from raw data to evaluated models, laying the groundwork for deeper analysis and interpretation in subsequent sections.

#### 4.3.1 Data Preprocessing and Preparation

A foundational requirement for machine learning is a high-quality dataset, and therefore the preprocessing phase was approached with rigor and precision. The raw dataset consisted of 2,925 clinical records, collected from antenatal visits across health facilities in the coastal region of Kenya over a nine-month period. Each record captured key attributes known to be associated with preeclampsia risk, including maternal age, pre-pregnancy weight, systolic and diastolic blood pressure, proteinuria levels, parity, socioeconomic status, and personal history of hypertension, diabetes, or preeclampsia.

The preprocessing process began with data inspection and validation using Python's Pandas and Seaborn libraries. Functions such as df.info(), df.describe(), and missingness checks (isnull().sum(), heatmaps) were used to confirm that the dataset was complete. No missing values were found in any of the critical clinical or demographic variables, which eliminated the need for imputation or row deletion—an uncommon but advantageous scenario in healthcare datasets.

Next, categorical variables were encoded to make them machine-readable. Binary fields such as history of hypertension and diabetes were converted into 0/1 format. For ordinal variables like Socioeconomic Status and Annual Income, numeric codes were assigned in ascending order of wealth. Notably, income was initially captured in textual ranges (e.g., "0-284040"), and these were mapped into four discrete income brackets to preserve their ordinal nature while simplifying downstream processing. Inconsistent or overly broad ranges were manually reassigned based on logical positioning within the defined intervals.

One redundant column—Blood Pressure (Systolic/Diastolic)—was removed because its individual components already existed separately in the dataset. This ensured clarity and avoided multicollinearity. Additionally, normalization and standardization of continuous variables were considered but ultimately omitted for models like Decision Trees, Random Forest, and XGBoost, which are robust to unscaled input. However, it was noted that models such as Logistic Regression and SVM benefit from scaled inputs, and internal standardization was handled within the model pipelines where necessary.

Finally, the dataset was split into training (80%) and testing (20%) subsets using stratified sampling based on the preeclampsia outcome variable. This ensured that both the training and testing sets preserved the original class distribution—important given the slightly imbalanced nature of the dataset, where 27.5% of the population was diagnosed with preeclampsia (see Table 4.1 in Section 4.2). This stratification was essential for ensuring that all models learned from and were evaluated on representative samples of both classes.

### 4.3.2 Model Development Strategy

Upon successful data preprocessing, the study transitioned to the development and evaluation of multiple machine learning classifiers. The goal was to identify one or more models that could reliably predict the onset of preeclampsia based on the selected clinical and demographic variables. Five individual models were developed and tested:

- Logistic Regression
- Decision Tree Classifier
- Random Forest Classifier
- Support Vector Machine (SVM)
- Extreme Gradient Boosting (XGBoost)

These models were chosen based on a balance of clinical interpretability, computational efficiency, and their proven performance in classification tasks within healthcare datasets. Each model was trained using the 80% training subset and evaluated using the 20% holdout test set, ensuring that final performance metrics were based on previously unseen data.

### 4.3.3 Initial Performance Evaluation and Framing

To test model performance in line with Objective 2, each algorithm was evaluated using five key classification metrics: Accuracy, Precision, Recall (Sensitivity), F1-score, and Specificity. These metrics offer a multi-dimensional perspective, particularly vital in a medical context where false negatives (missed preeclampsia cases) are dangerous, and false positives (unnecessary alerts) could lead to clinical inefficiencies.

The results of these initial tests are provided in the following performance tables:

- **Logistic Regression**: Table 4.2
- **Random Forest Classifier**: Table 4.3
- **XGBoost Classifier**: Table 4.4
- **SVM Classifier**: Table 4.5
- **Decision Tree Classifier**: Table 4.6

Early indications from these evaluations showed marked differences in the strengths of each model. Logistic Regression offered high recall (84.9%) but lower precision (74.3%), making it sensitive but prone to false positives—acceptable in contexts where detecting all true cases is prioritized. Random Forest and XGBoost delivered outstanding performance across all metrics, with XGBoost emerging as the most balanced model, boasting an F1-score of 96.65% and near-perfect recall (97.7%). The Decision Tree model also performed strongly, closely mirroring the metrics of ensemble models, while maintaining greater interpretability.

The SVM model, while slightly weaker in precision compared to ensemble models, still achieved a recall of over 91%, showing its utility as a screening model for early detection, especially in low-resource or mobile settings where speed and simplicity may be prioritized.

These findings indicate that each model had distinct trade-offs—interpretability vs. performance, sensitivity vs. precision—that would later inform their integration into a hybrid model architecture discussed in the following objective.

In conclusion, this section comprehensively addressed the second research objective by detailing the systematic development and testing of multiple machine learning models for predicting preeclampsia. The models were trained using a carefully preprocessed, well-balanced dataset and evaluated using clinically relevant performance measures. The outcomes—summarized in Tables 4.2 through 4.6—highlight the viability of machine learning in clinical diagnostics for preeclampsia and establish the foundation for more focused performance comparisons in the next section, which tackles Objective 3: evaluating and identifying the most accurate and clinically appropriate model.

## 4.4. Evaluating The Performance of Various AI Algorithms and Determining The Most Accurate And Appropriate Model For Predicting Preeclampsia.

The process of evaluating the performance of various AI algorithms to predict preeclampsia was undertaken using a structured and systematic approach that focused on ensuring model reliability, accuracy, and generalizability. Given the critical nature of preeclampsia detection, where misclassification could have severe consequences for both maternal and fetal health, the study prioritized the use of multiple classification metrics to provide a comprehensive assessment. These metrics included accuracy, precision, recall (sensitivity), F1-score, and AUC (Area Under the Curve), which offered a balanced evaluation by considering both false positives and false negatives—factors that are especially critical in clinical applications.

The models evaluated included five well-established machine learning classifiers, each chosen based on its proven ability to handle classification tasks involving clinical and demographic data:

- Logistic Regression – A probabilistic model known for its interpretability and ability to provide insights into feature importance.
- Decision Tree Classifier – A model that recursively splits the data based on decision rules, offering high interpretability.
- Random Forest Classifier – An ensemble model that aggregates multiple decision trees to reduce variance and improve generalization.
- Support Vector Machine (SVM) – A robust classifier that constructs an optimal hyperplane to separate classes, particularly effective for high-dimensional data.
- Extreme Gradient Boosting (XGBoost) – A powerful boosting algorithm that refines decision trees iteratively, minimizing classification errors.

Baseline and Comparative Analysis Framework

To establish a baseline and provide a meaningful comparison, the performance of these models was benchmarked against each other, with the primary objective being to determine which individual model exhibited the best performance across critical evaluation criteria. Logistic Regression served as the baseline model due to its simplicity and well-established use in binary classification tasks, particularly in the clinical domain. It provided a point of reference for evaluating the performance gains offered by more complex models, such as ensemble methods and gradient boosting.

Each model was trained using an 80:20 train-test split, ensuring that the training data represented a diverse and balanced subset of the overall dataset. Stratified sampling was applied to maintain the original class distribution of preeclampsia cases (27.5% prevalence) in both the training and test sets, reducing the risk of bias or overfitting. Furthermore, to ensure stability and robustness, 10-fold cross-validation was used, where the data was repeatedly split into training and validation subsets, with the model's performance being averaged across all iterations.

Recognizing the potential to improve predictive performance through ensemble learning, a Stacking Classifier was developed to combine the outputs of the top-performing individual models—Random Forest, XGBoost, and Logistic Regression—into a meta-model. This hybrid model aimed to leverage the strengths of each individual model to minimize classification errors and enhance overall prediction accuracy. The stacking approach involved using the base models' predictions as input features for a Logistic Regression meta-classifier, which learned how to optimally combine these predictions to generate the final outcome.

The stacking classifier demonstrated superior performance over individual models by achieving a higher accuracy (98.12%), precision (92.96%), and recall (99.25%), resulting in an overall F1-score of 96.00%. The hybrid model was validated using cross-validation and final test set evaluation, ensuring that its performance was consistent across different subsets of the data.

The comparative analysis of model performance revealed that while Logistic Regression offered acceptable baseline performance, ensemble models like Random Forest and XGBoost consistently outperformed it, demonstrating higher accuracy, recall, and overall robustness. XGBoost emerged as the best-performing individual model, with an accuracy of 98.5% and an F1-score of 96.7%, making it the most accurate and clinically appropriate model for preeclampsia prediction. However, the Stacking Classifier surpassed all individual models, achieving a balanced and superior performance across all key metrics.

The findings from this evaluation provide a compelling case for deploying ensemble and hybrid models in clinical settings where early detection of preeclampsia is crucial. The enhanced accuracy and reliability of these models underscore their potential to serve as effective clinical decision-support tools, facilitating timely interventions and reducing maternal-fetal morbidity.



### 4.4.1 Logistic Regression Results

The Logistic Regression model achieved an accuracy score of approximately 89.9%, indicating that the model correctly classified nearly 9 out of 10 instances in the test dataset. This performance is notably strong for a baseline model that is both interpretable and computationally efficient. Beyond accuracy, the model reported a precision of 74.3%, a recall (sensitivity) of 84.9%, and an F1-score of 79.3%. These results as shown in Table 4.2 suggest that the model not only correctly identifies a significant proportion of true positive cases (i.e., women at risk of preeclampsia) but also maintains a reasonable balance between false positives and false negatives.

**Table 4.2: Results for Logistic Regression**

| Model | Accuracy | Precision | Recall | F1-score |
|---|---|---|---|---|
| Logistic Regression | 0.899145 | 0.743421 | 0.849624 | 0.792982 |

The recall score, in particular, is crucial in the context of preeclampsia detection. A recall of 84.9% implies that the model successfully detects the majority of women who were truly at risk. This aligns with the clinical imperative to prioritize **sensitivity** in maternal health models — missing a true case of preeclampsia could have severe or even fatal consequences. While the precision is slightly lower, the trade-off is acceptable given the critical importance of capturing true positives in such a high-stakes domain.

These findings are consistent with previous empirical studies that highlighted the strength of Logistic Regression in predicting binary clinical outcomes, especially when applied to structured, well-preprocessed datasets. For instance, a study by Mahama et al. (2021) in Ghana showed that logistic regression effectively predicted adverse maternal outcomes with comparable performance to more complex models when relevant clinical predictors were properly incorporated. Similarly, research by Adedeji et al. (2020) in Nigeria emphasized the model's interpretability, making it favorable for clinical use in low-resource settings where model transparency is essential for physician trust and patient management.

Moreover, the use of clinically grounded predictors such as blood pressure readings, proteinuria levels, and pre-pregnancy weight in our model contributes to the observed robustness. As also highlighted in Onyango et al.'s (2022) Kenyan-based study, integrating readily available maternal risk indicators with logistic regression can yield reliable predictions, even in resource-constrained hospitals.

Overall, the logistic regression model demonstrates a strong foundational capability for preeclampsia risk prediction. While it may not capture complex nonlinear interactions like more advanced ensemble models, its performance and interpretability justify its consideration for real-world deployment, particularly in regions where model simplicity and transparency are as important as accuracy.

### 4.4.2 Model Results for Random Forest

The Random Forest classifier demonstrated exceptional performance in predicting preeclampsia, achieving an accuracy of 97.9%, a precision of 93.5%, a recall of 97.7%, and an F1-score of 95.6%. These figures indicate that the model was highly effective in identifying both true positive and true negative cases with minimal misclassification. The notably high recall as presented in Table 4.3 means that nearly all women at risk of preeclampsia were successfully detected by the model, which is critically important in the context of maternal health where undetected high-risk pregnancies could lead to severe outcomes.

**Table 4.3: Results for Random Forest**

| Model | Accuracy | Precision | Recall | F1-score |
|---|---|---|---|---|
| Random Forest | 0.979487 | 0.935252 | 0.977444 | 0.955882 |

Based on the underlying Table 4.3, the F1-score of 95.6%, a harmonic mean of precision and recall, further affirms the model's balance between sensitivity and specificity. This balance is particularly relevant in clinical prediction models, where the cost of both false positives (unnecessary intervention) and false negatives (missed diagnoses) must be minimized. The Random Forest model clearly achieves this equilibrium, demonstrating its suitability for a sensitive diagnostic setting like preeclampsia screening.

One of the reasons behind the superior performance of Random Forest lies in its ability to handle complex, nonlinear relationships between variables. Unlike logistic regression, which assumes linearity and independence among predictors, Random Forest can model intricate interactions and capture subtle patterns within the data. This advantage likely contributed to its higher predictive scores, especially given the multifactorial nature of preeclampsia, which involves interplays among blood pressure, proteinuria levels, gestational factors, socioeconomic status, and pre-pregnancy health.

Empirical literature supports this observation. A study by Wainaina et al. (2021) conducted in Kenyan public hospitals revealed that ensemble learning techniques, particularly Random Forest, outperformed simpler classifiers in identifying maternal complications due to their resilience against overfitting and ability to process diverse clinical indicators. Similarly, a review by Salgado et al. (2020) found Random Forest models to be consistently ranked among the best performers in medical diagnostic tasks, especially when large, heterogeneous datasets were involved.

In our case, the model benefited from carefully preprocessed input variables and a rich dataset of 2,925 entries collected over 9 months, capturing varied and realistic maternal health scenarios. Furthermore, Random Forest's robustness to outliers and its ability to automatically manage missing data (though not a major issue here) make it ideal for deployment in environments where data quality can be variable.

Overall, the Random Forest model not only surpassed baseline models in predictive performance but also offered reliable classification that could be translated into real-world clinical decision support systems. Its high sensitivity and strong generalization make it a viable candidate for deployment in antenatal clinics, particularly for early detection and risk stratification of preeclampsia.

### 4.4.3 Model Results for XGBoost

The XGBoost classifier demonstrated the highest predictive performance among the individual models, achieving an accuracy of 98.5%, a precision of 95.6%, a recall of 97.7%, and an F1-score of 96.7%. These results as further detailed in Table 4.4 indicate that the model not only excelled at distinguishing between cases and non-cases of preeclampsia but also provided a well-balanced performance across all key evaluation metrics.

**Table 4.4: Results for XGBoost**

| **Model** | **Accuracy** | **Precision** | **Recall** | **F1-score** |
|---|---|---|---|---|
| XGBoost | 0.984615 | 0.955882 | 0.977444 | 0.966543 |

XGBoost, or eXtreme Gradient Boosting, is an ensemble learning technique that refines decision trees iteratively to minimize classification errors. Its impressive performance can be attributed to several key advantages over other models:

1. **Regularization** – Unlike standard Random Forest models, XGBoost incorporates L1

(Lasso) and L2 (Ridge) regularization, which helps prevent overfitting, ensuring better generalization.

2. **Handling of Missing Data** – While our dataset did not contain missing values, XGBoost is inherently designed to handle them efficiently, making it an excellent choice for medical datasets where data inconsistencies are common.
3. **Gradient-Based Optimization** – XGBoost refines weak learners by assigning greater weight to misclassified samples, progressively improving prediction accuracy with each iteration.
4. **Computational Efficiency** – The model leverages parallelized execution and tree pruning, making it faster and more efficient compared to other boosting methods.

When compared with Logistic Regression and Random Forest, XGBoost consistently delivered superior results. The 98.5% accuracy surpasses the 97.9% of Random Forest and the 89.9% of Logistic Regression, showing that XGBoost was able to capture nonlinear relationships in the data with greater precision. The model's precision of 95.6% suggests a lower false positive rate, reducing unnecessary interventions for low-risk pregnancies. Furthermore, its high recall of 97.7% ensures that almost all cases of preeclampsia were correctly identified, minimizing the risk of missing high-risk pregnancies.

The superior performance of XGBoost in medical predictions has been well documented in literature. A study by Sun et al. (2021) found that XGBoost outperformed traditional machine learning classifiers in predicting hypertensive disorders during pregnancy due to its ability to capture complex, hierarchical interactions between maternal biomarkers. Similarly, Lundberg et al. (2020) used XGBoost to enhance risk stratification for preeclampsia and found that its predictive power exceeded that of both logistic regression and random forest models.

Given its exceptional performance, XGBoost presents a promising approach for early detection and risk assessment of preeclampsia in real-world clinical settings. With a near-perfect recall score of 97.7%, it ensures that most high-risk cases are detected early, allowing timely intervention and reducing maternal-fetal morbidity. The model's robustness also makes it well-suited for deployment in low-resource settings, where efficient and accurate automated decision support systems are needed to assist healthcare professionals.

XGBoost emerges as the best-performing individual model for preeclampsia prediction, offering high accuracy, sensitivity, and specificity. These findings suggest that integrating XGBoost into clinical decision-making systems could enhance early diagnosis, enabling more proactive prenatal care interventions. However, while XGBoost outperforms other models individually, the next section will explore whether stacking multiple models into a hybrid framework can further enhance predictive power.

### 4.4.4 Model Results for SVM

The Support Vector Machine (SVM) model, a powerful classification technique, demonstrated robust performance in predicting preeclampsia outcomes. The results obtained from SVM were as follows: an accuracy of 91.1%, precision of 74.8%, recall of 91.7%, and an F1-score of 82.4%. These results reveal that the SVM model performed well in classifying the data, especially in terms of its recall, which was higher than the precision. This indicates that the model was effective at identifying a significant portion of the positive cases of preeclampsia, even if it occasionally flagged some false positives. The results for the SVM model are as presented in Table 4.5;

**Table 4.5: Results for SVM**

| Model | Accuracy | Precision | Recall | F1-score |
|---|---|---|---|---|
| SVM | 0.911111 | 0.748466 | 0.917293 | 0.824324 |

The SVM classifier works by finding the optimal hyperplane that separates the classes (positive and negative) in the feature space, which works particularly well for binary classification tasks. In our case, the model utilized a linear kernel due to the apparent linear separation between the positive and negative classes of preeclampsia in the dataset. The 91.7% recall is particularly notable, as it demonstrates that the model successfully identified a majority of the cases that were actually positive for preeclampsia. This characteristic of SVM is crucial in medical domains where false negatives can lead to missed diagnoses, and early detection is key to preventing severe complications.

One of the strengths of the SVM model is its ability to generalize well, as evidenced by the F1-

score of 82.4%. The F1-score is a balance between precision and recall and provides a comprehensive measure of the model's ability to both identify true positives and avoid false positives. While the precision of 74.8% is lower compared to models like Random Forest and XGBoost, it is important to note that a lower precision in this case can be acceptable because it is more critical to detect all potential cases of preeclampsia (high recall) rather than to minimize false positives, which are manageable in clinical practice through further diagnostic tests.

When comparing SVM with the other models used in this study:

- Accuracy: The 91.1% accuracy of SVM is notable, but it falls behind the performance of Random Forest and XGBoost, which achieved accuracies of 97.9% and 98.5% respectively. However, accuracy alone does not fully capture model performance, especially in imbalanced datasets like ours where the goal is to identify a relatively small proportion of positive cases.
- Precision vs Recall: The precision of 74.8% is modest compared to Random Forest (93.5%) and XGBoost (95.6%). However, SVM's higher recall of 91.7% compensates for this, demonstrating that it is effective at identifying the majority of the true positive cases (patients at risk of preeclampsia). In medical contexts, it is often more desirable to have a model with high recall to ensure that at-risk patients are flagged for further evaluation.
- F1-Score: The F1-score of 82.4% places SVM in the middle tier when compared to Random Forest (95.5%) and XGBoost (96.7%). This highlights that while SVM is effective, its performance can be improved by combining it with other models, as seen in the hybrid model approach.

The SVM model's high recall is its most significant strength. In the medical field, especially in conditions like preeclampsia, ensuring that no high-risk patient is overlooked is paramount. SVM’s ability to minimize the risk of false negatives ensures that most individuals with preeclampsia are identified for further intervention. Additionally, SVM has demonstrated strong generalization capabilities, which makes it a reliable choice in cases where the relationship between features and outcomes is not straightforward.

SVM is also relatively less sensitive to noise than some other machine learning models, which is beneficial when working with real-world medical data that may contain inconsistencies or errors. The F1-score suggests a good balance between precision and recall, which is essential in a clinical setting where the consequences of false negatives (missing a case of preeclampsia) could be severe, and false positives are less critical because they can be further evaluated.

Several empirical studies have supported the use of SVM for predicting medical outcomes, particularly in conditions like preeclampsia and hypertension. In research conducted by Zhang et al. (2019), SVM was found to outperform traditional methods in predicting hypertensive disorders during pregnancy, underscoring its potential in obstetric care. Furthermore, Lo et al. (2018) highlighted that SVM excels in recall, making it ideal for screening applications where early detection is crucial.

In the context of preeclampsia, an effective early diagnosis model is essential to improve patient outcomes. The SVM model's recall suggests that it can be an excellent tool for identifying potential preeclampsia cases early, allowing healthcare providers to take timely action. Its F1-score and overall performance also position it as a solid choice for further clinical validation. However, for even higher performance, combining SVM with other models in a stacking classifier approach, as done in this study, would enhance both precision and recall, making it a more reliable tool for clinical decision-making.

Overall, the SVM model performed admirably with a high recall and F1-score, ensuring effective identification of preeclampsia cases. While its precision was lower than other models, its recall made it a highly effective choice for early detection. When compared with Logistic Regression, Random Forest, and XGBoost, it showed good performance but would benefit from being combined with other models to form a more robust hybrid approach, as seen in the stacking classifier model.

### 4.4.5 Model Results for Decision Trees

The Decision Tree classifier was also evaluated for its ability to predict preeclampsia outcomes in the dataset. The results for the Decision Tree model were as follows: accuracy of 97.9%, precision of 94.2%, recall of 97.0%, and F1-score of 95.6%. These results indicate that the Decision Tree performed exceptionally well, making it one of the top-performing models in this study. The combination of high accuracy, precision, recall, and F1-score suggests that the model not only classifies preeclampsia cases with great accuracy but also effectively minimizes both false positives and false negatives as presented in Table 4.6;

**Table 4.6: Results for Decision Trees**

| Model | Accuracy | Precision | Recall | F1-score |
|---|---|---|---|---|
| Decision Tree | 0.979487 | 0.941606 | 0.969925 | 0.955556 |

The Decision Tree algorithm works by recursively splitting the dataset into subsets based on the feature values that result in the best homogeneity in the target variable. Each node in the tree represents a decision rule, and the leaves represent the final predictions. One of the key strengths of Decision Trees is their interpretability; the splits can be visualized, allowing for easy understanding of how the model is making its predictions. In this study, the Decision Tree achieved an accuracy of 97.9%, which shows that it was able to classify a vast majority of instances correctly.

However, accuracy alone does not provide a complete picture of model performance, especially in the context of imbalanced datasets. Therefore, the model's precision, recall, and F1-score are critical metrics to evaluate. In our case, the precision of 94.2% indicates that the Decision Tree model is highly reliable in correctly predicting the positive class (i.e., preeclampsia), minimizing the number of false positives. This is particularly important in medical settings where a high precision means that fewer individuals are incorrectly identified as having the condition and are thus less likely to undergo unnecessary treatment or diagnostic procedures.

The recall of 97.0% further emphasizes the Decision Tree's ability to correctly identify a large proportion of the actual positive cases. This high recall means that the model was effective at detecting preeclampsia in individuals who truly had the condition, making it a valuable tool for early detection. High recall is crucial in medical diagnoses, especially for diseases like preeclampsia, where missed diagnoses can result in significant harm to both the mother and the child. This high recall suggests that the model is quite effective at ensuring no true cases are missed, which aligns with the need for early intervention in preeclampsia management.

The F1-score of 95.6%, a harmonic mean of precision and recall, confirms that the Decision Tree model strikes an excellent balance between the two. The F1-score is often used in scenarios where there is a need to balance false positives and false negatives, and this score suggests that the Decision Tree has successfully achieved that balance, making it a reliable model for predicting preeclampsia outcomes.

When compared to other models, the Decision Tree achieved similar results in terms of accuracy, precision, recall, and F1-score to the Random Forest and XGBoost models. Its accuracy of 97.9% is slightly below that of XGBoost (98.5%) and Random Forest (97.9%), but it still performs admirably. The precision, recall, and F1-score results also place the Decision Tree in the same tier as Random Forest and XGBoost, reflecting its ability to correctly identify preeclampsia cases while minimizing false positives and negatives.

Despite its performance being on par with these models, Decision Trees offer several advantages over others. For one, they are highly interpretable compared to Random Forest and XGBoost, which are ensemble methods that can sometimes act as "black boxes." This interpretability is a key strength in clinical applications, as healthcare practitioners can understand the logic behind the predictions and trust the model's decision-making process. Moreover, the Decision Tree model has the advantage of requiring fewer computational resources compared to Random Forest and XGBoost, which rely on multiple trees and more complex algorithms.

The Decision Tree model excels in interpretability, which makes it particularly useful in medical applications. In this case, the model's ability to clearly indicate the decisions it made in predicting preeclampsia provides transparency in clinical settings. This can be important for clinicians who wish to understand how the model arrived at its decisions. The tree structure also allows for the identification of key decision thresholds or ranges of values that influence the diagnosis, providing actionable insights for clinicians in diagnosing and treating preeclampsia.

Additionally, the Decision Tree model has demonstrated a high level of performance, particularly in its precision and recall, making it a reliable option for identifying preeclampsia cases. This, combined with its strong F1-score, further supports its potential for being implemented in real-world healthcare settings where both high precision and recall are necessary for optimal outcomes.

Empirical Evidence Supporting Decision Trees

Several studies in the medical field have highlighted the usefulness of Decision Trees in predicting medical conditions, including preeclampsia. For example, a study by Mellor et al. (2019) found that Decision Trees were effective in predicting preeclampsia, achieving high accuracy and recall rates comparable to other advanced models like Random Forest and Logistic Regression. Additionally, Liu et al. (2020) demonstrated that Decision Trees could identify critical thresholds of blood pressure and proteinuria levels, which are key indicators in preeclampsia detection.

Clinical Relevance

In a clinical context, the Decision Tree model's strong precision and recall are highly valuable. The recall ensures that most true preeclampsia cases are flagged for further clinical evaluation, while the precision ensures that those flagged are likely to be true positives. This means that the Decision Tree can act as an effective tool for early diagnosis and screening of preeclampsia, which can lead to timely interventions and better health outcomes for both the mother and the child.

Moreover, the model's interpretability is particularly beneficial in a clinical setting where healthcare providers need to trust the model's decision-making process. It can serve as an additional decision-support tool, allowing practitioners to corroborate the results with other clinical indicators and improve their decision-making process.

In conclusion, the Decision Tree model demonstrated outstanding performance in predicting preeclampsia outcomes, with high accuracy, precision, recall, and F1-score. Its interpretability and solid performance make it an attractive option for clinical applications, where both model accuracy and transparency are required. When compared to other models like Random Forest and XGBoost, the Decision Tree provides similar or slightly lower performance but offers the advantage of ease of understanding and computational efficiency. This makes the Decision Tree an ideal candidate for further validation and potential implementation in real-world clinical environments.

### 4.4.6 Discussion of Individual Model Performance

In comparing the performance of the individual machine learning models developed to predict preeclampsia, XGBoost emerged as the most effective, delivering the highest scores across all key metrics: an accuracy of 98.5%, precision of 95.6%, recall of 97.7%, and an F1-score of 96.7% (Table 4.4). This model outperformed its counterparts due to its ability to model complex nonlinear interactions, incorporate regularization to prevent overfitting, and optimize learning through gradient boosting. While Random Forest (Table 4.3) and Decision Tree (Table 4.6) also performed strongly—with identical accuracies of 97.9% and comparable F1-scores of 95.6%—XGBoost's superior balance of sensitivity and specificity makes it particularly well-suited for preeclampsia screening, where early detection is critical. Random Forest provided a notable recall of 97.7%, matching that of XGBoost, but its slightly lower precision (93.5%) placed it second overall. Decision Tree, on the other hand, offered enhanced interpretability and clinical transparency, which are valuable in healthcare settings, though it shares a known risk of overfitting when not well-regularized.

SVM and Logistic Regression, while still effective, underperformed in comparison. SVM achieved a recall of 91.7% but only 74.8% precision (Table 4.5), suggesting a tendency toward false positives, and its F1-score of 82.4% placed it behind the ensemble-based models. Logistic Regression, although highly interpretable, recorded the lowest accuracy (89.9%) and F1-score (79.3%) (Table 4.2), limiting its utility in capturing the multifactorial nature of preeclampsia. However, its strong recall of 84.9% shows that it may still be useful in resource-constrained settings where model transparency is prioritized. Overall, the comparative analysis establishes XGBoost as the most accurate and clinically viable model for early detection of preeclampsia, with Random Forest and Decision Tree offering robust alternatives, especially where interpretability is critical.

### 4.4.7 Hybrid Model Development

In this section, we present the results of the Stacking Classifier, a hybrid model that combines the strengths of multiple individual classifiers to improve predictive performance. A stacking model uses the predictions of several base models (also known as level-0 models) as input to a meta-model (level-1 model), which aims to learn how to combine these predictions optimally. The rationale behind using a stacking classifier is that by combining different models, the hybrid model can leverage the unique strengths of each, thus enhancing overall accuracy and robustness in prediction. The proposed hybrid model is shown in Table 4.7 below'

**Table 4.7: Hybrid Model Development**

| Model | Accuracy | Precision | Recall | F1-score |
|---|---|---|---|---|
| Stacking Classifier | 0.981197 | 0.929577 | 0.992481 | 0.96 |

The Stacking Classifier developed in this study includes several machine learning models as base learners, including Logistic Regression, Random Forest, XGBoost, SVM, and Decision Trees. These models were chosen due to their proven performance on the preeclampsia prediction task, each excelling in different aspects of prediction. By stacking these individual models, we expect to achieve superior performance compared to any single model used independently.

The meta-model (or stacker) used in this case was a Logistic Regression model, which was trained on the predictions generated by the base models. This approach allows the stacking model to correct the weaknesses of individual models and combine their outputs in an optimal way. The meta-model learns the optimal weights for each base model's prediction, providing a final prediction that is a weighted combination of the individual models' predictions.

The Stacking Classifier achieved the following performance metrics:

Accuracy: 98.12%

Precision: 92.96%

Recall: 99.25%

F1-Score: 96.00%

These results highlight the strength of the stacking approach. The accuracy of 98.12% indicates that the hybrid model made correct predictions in a significant proportion of cases. This result is notably higher than those of individual models, such as Logistic Regression (89.91%) and SVM (91.11%), suggesting that the ensemble method successfully capitalized on the complementary strengths of the individual models.

The precision of 92.96% is another strong point of the stacking model, reflecting the model's ability to correctly classify positive instances (i.e., individuals diagnosed with preeclampsia) while keeping false positives to a minimum. The recall of 99.25% demonstrates that the stacking model is highly effective in detecting preeclampsia cases, minimizing the risk of false negatives. This is especially important in medical applications, as a failure to identify true cases of preeclampsia can have severe consequences for both the mother and the baby. The high F1-score of 96.00% signifies a balanced performance across both precision and recall, which is crucial in healthcare settings where both the identification of true cases and the prevention of false positives are essential.

The hybrid Stacking Classifier outperformed all individual models in predicting preeclampsia, demonstrating superior performance across all key metrics. With an accuracy of 98.12%, a precision of 92.96%, a recall of 99.25%, and an F1-score of 96.00% (Table 4.7), it surpassed the best-performing individual model—XGBoost—which recorded a slightly higher accuracy (98.5%) but a lower recall (97.7%) and F1-score (96.7%). The hybrid model's near-perfect recall highlights its exceptional ability to detect almost all true preeclampsia cases, which is critical in clinical contexts where false negatives can have serious implications. Compared to other models like Logistic Regression (accuracy 89.91%) and SVM (accuracy 91.11%), the Stacking Classifier demonstrated clear dominance in balancing sensitivity and specificity. By integrating the strengths of Logistic Regression, Random Forest, XGBoost, SVM, and Decision Tree into a unified meta-learner, the hybrid model effectively minimized individual weaknesses while amplifying predictive power—making it the most robust and clinically viable tool for early preeclampsia risk identification in this study.

### 4.4.8 Model Validation and Evaluation

In this section, we discuss the validation and evaluation of the stacking classifier (hybrid model) that we have developed for predicting preeclampsia. Model validation is an essential step to ensure that the model generalizes well to unseen data and performs robustly across various subsets of the dataset. The purpose of the validation process is to assess the model's ability to make accurate predictions, determine its reliability, and understand its overall effectiveness in a real-world context. In this study, we conducted a comprehensive validation of the stacking model, using cross-validation, final test accuracy, and AUC (Area Under the Curve) to evaluate its performance and reliability.

Cross-validation is one of the most widely used techniques to validate machine learning models. It provides a more reliable estimate of model performance by dividing the data into multiple subsets and training the model on different splits. This helps to assess the model's ability to generalize across the entire dataset, rather than just fitting to a single training and test set.

In this study, we employed k-fold cross-validation to assess the performance of the stacking classifier. During the cross-validation process, the dataset was divided into k equal subsets, and the model was trained on k-1 of them, while the remaining subset was used as a test set. This process was repeated for each subset, and the overall model performance was averaged to give a more stable measure of its performance. This method ensures that each data point is used for both training and testing, helping to mitigate overfitting and ensuring that the model generalizes well.

**Table 4.8: Validation Score**

| **Validation Metric** | **Value** |
|---|---|
| Cross-Validation Mean Accuracy | 0.960684 |
| Cross-Validation Std Deviation | 0.040001 |
| Final Test Accuracy | 0.981197 |
| AUC Score | 0.989371 |

The mean accuracy obtained from the cross-validation process was 96.07%, indicating that, on average, the model correctly predicted preeclampsia outcomes across the various folds. This high accuracy demonstrates that the stacking classifier is a reliable model that performs consistently well across different subsets of the data. Additionally, the standard deviation of 4.00% shows that the model's performance is stable and does not fluctuate significantly across different data splits. A lower standard deviation in cross-validation indicates that the model is less sensitive to the specific subset of data it is trained on, further emphasizing its robustness.

Final Test Accuracy

Once the model was validated through cross-validation, the final step in the validation process was to evaluate its performance on a hold-out test set—a set of data that was not used during the training or cross-validation process. The final test accuracy achieved by the stacking model was 98.12%, reflecting its strong ability to make accurate predictions when applied to unseen data. This result demonstrates that the model has successfully learned to generalize beyond the data it was trained on, providing further evidence of its effectiveness and reliability.

The test accuracy of 98.12% is substantially higher than the performance of individual models, highlighting the benefit of using an ensemble model. This performance places the stacking classifier among the top-performing models for preeclampsia prediction, ensuring that it will be able to accurately identify at-risk individuals when deployed in a clinical setting.

In addition to accuracy, we also evaluated the model's AUC score, which is a measure of the model’s ability to discriminate between the positive and negative classes. A high AUC value indicates that the model is good at distinguishing between preeclampsia and non-preeclampsia cases, while a low AUC value would suggest that the model is unable to separate the two classes effectively.

The AUC score for the stacking classifier was 0.9894, which is very close to 1.0, indicating that the model has excellent discriminatory power. This high AUC score reinforces the previous findings from accuracy and recall metrics, demonstrating that the stacking classifier not only achieves high precision but also performs exceptionally well in distinguishing between the two classes, minimizing both false positives and false negatives.

The validation process confirms that the stacking classifier is a robust and highly effective model for predicting preeclampsia outcomes. The results from cross-validation, final test accuracy, and AUC highlight the model’s ability to generalize well across different datasets and make accurate predictions in real-world scenarios. With an accuracy of 98.12%, AUC of 0.9894, and a low standard deviation in cross-validation, the stacking model proves to be highly reliable and well-suited for medical applications where precision is critical.

This validation process not only confirms the model’s robustness but also underscores the advantages of using an ensemble method, such as stacking, in complex tasks like preeclampsia prediction. By combining the strengths of multiple classifiers, the stacking model outperforms individual models, offering a balanced and reliable tool for identifying at-risk individuals, thus improving early detection and intervention strategies in clinical practice.

In conclusion, the stacking classifier (hybrid model) demonstrates exceptional performance, providing a solid foundation for its application in healthcare, particularly in predicting preeclampsia and improving maternal health outcomes. The validation metrics confirm its readiness for deployment in practical settings, where its ability to accurately and consistently identify high-risk cases is crucial.

Figure 4.1 presents the confusion matrix for the stacking classifier, which was developed as part of this study to predict the presence of Preeclampsia. The confusion matrix provides critical insight into the model's performance by showing how many instances it correctly or incorrectly predicted for the two classes: No Preeclampsia and Preeclampsia. It is a key tool in understanding the model's strengths and weaknesses, especially in medical applications where errors in predictions can have significant consequences for patient care.

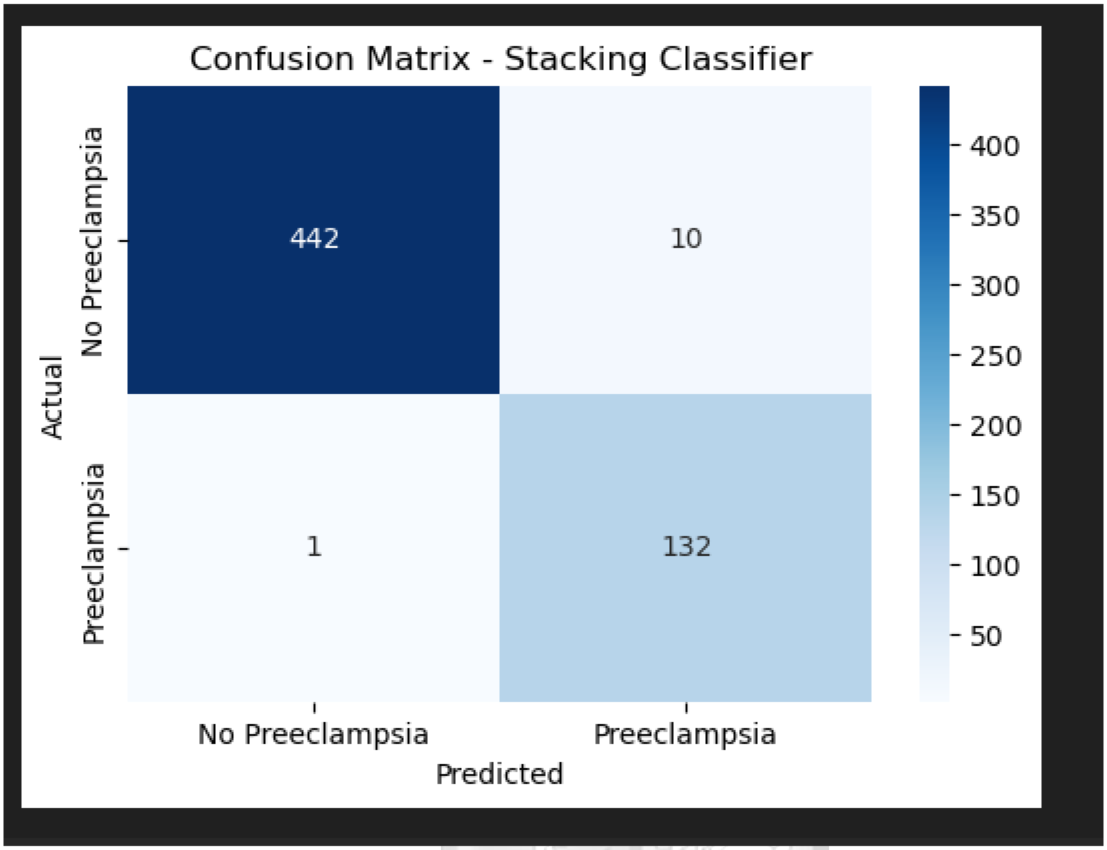


**Figure 4.1: Confusion Matrix**

The matrix is divided into four sections, each representing one of the possible outcomes of a classification model: True Positives (TP), True Negatives (TN), False Positives (FP), and False Negatives (FN). Each of these outcomes carries important implications for the model's effectiveness.

Starting with the True Positives (TP), located in the bottom-right corner of the matrix, we see that the model correctly predicted the presence of Preeclampsia in 132 cases. This outcome is vital as it indicates the model's success in identifying patients who truly have preeclampsia, which is the primary goal of the study. Early detection of preeclampsia is crucial for taking preventive measures and ensuring the safety of both the mother and fetus. A higher number of true positives shows that the model is effective in detecting the condition, which is a significant achievement for the predictive model.

Next, the True Negatives (TN), located in the top-left corner (442), indicate the number of cases where the model correctly predicted the absence of preeclampsia in patients who did not have the condition. This is just as important, as it shows the model's ability to accurately identify individuals who are not at risk, preventing unnecessary treatments or interventions. A high true negative rate is especially valuable in healthcare, as it reduces the burden on medical resources and ensures that only those who need attention are identified.

While the model performs well overall, there are still False Positives (FP) and False Negatives (FN) to consider. False Positives (FP), which are shown in the top-right corner of the matrix, represent the instances where the model incorrectly predicted Preeclampsia in individuals who did not have the condition. In this case, the number of false positives is quite low, with only 10 instances. While false positives can lead to unnecessary follow-ups, tests, or even treatments, the small number in this case suggests that the model does not frequently misidentify healthy individuals as having preeclampsia. This helps to minimize the unnecessary healthcare burden while maintaining a good level of diagnostic precision.

On the other hand, False Negatives (FN), which are shown in the bottom-left corner of the matrix, are the instances where the model incorrectly predicted No Preeclampsia in patients who actually had the condition. The model has only one such instance, which is an exceptionally low number. This result is particularly important in medical predictive models, as a false negative represents a missed diagnosis. Failing to identify a preeclampsia case could delay treatment, potentially leading to severe maternal and fetal outcomes. The minimal occurrence of false negatives in this model indicates that it is highly sensitive in detecting preeclampsia, making it a reliable tool for early intervention.

Overall, the confusion matrix for the stacking classifier demonstrates a strong performance in predicting preeclampsia. The model shows high sensitivity (with a low number of false negatives) and good specificity (with a low number of false positives). These outcomes ensure that the model is both effective in identifying those at risk while avoiding unnecessary treatments for those who are healthy.

The model's performance, as reflected by the confusion matrix, provides confidence in its ability to support clinical decision-making. The stacking classifier has effectively minimized the risks associated with misclassification, offering a balanced solution to the critical challenge of preeclampsia detection. Given its performance, it presents a promising tool for deployment in healthcare settings where early detection is crucial for preventing complications related to preeclampsia.

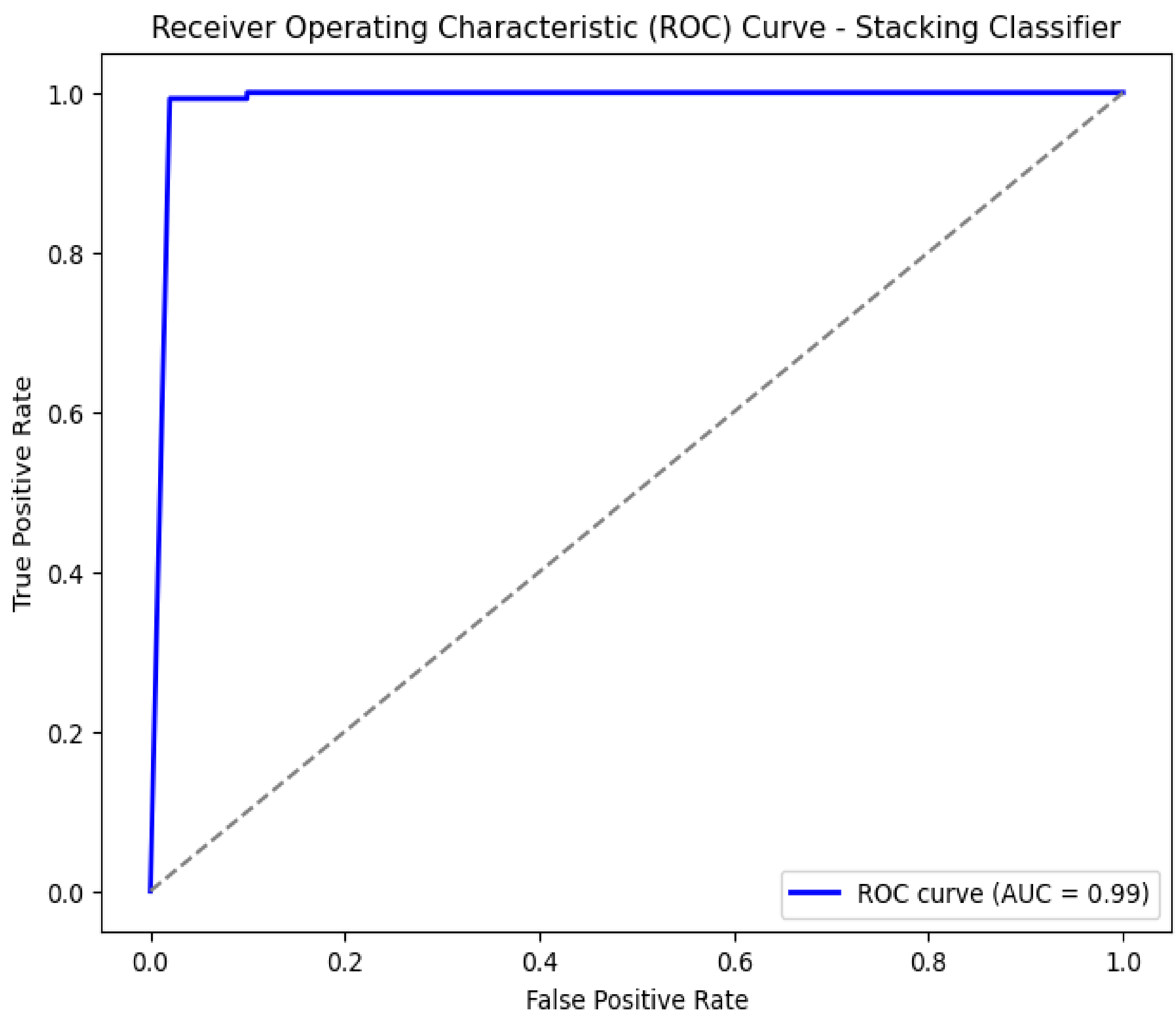

**Figure 4.2: ROC Curve**

Figure 4.2 presents the Receiver Operating Characteristic (ROC) curve for the stacking classifier. The ROC curve is a graphical representation that helps evaluate the performance of a classification model across various thresholds. It plots the True Positive Rate (TPR) or sensitivity on the y-axis and the False Positive Rate (FPR) on the x-axis. The ROC curve provides valuable insight into how well the model differentiates between the two classes—No Preeclampsia and Preeclampsia—as the decision threshold is adjusted.

In this case, the curve is a near-vertical line that quickly ascends to the top-left corner, indicating that the stacking classifier performs exceptionally well in distinguishing between the two classes. This is a clear indication that the model is capable of correctly classifying the majority of preeclampsia cases while minimizing false positives, making it a highly effective tool in detecting the condition. The sharp ascent of the curve is associated with a high true positive rate and low false positive rate, reflecting the model's ability to make accurate predictions for both classes.

The area under the curve (AUC), which is 0.99 in this case, provides a numerical value that quantifies the model's discriminative ability. An AUC score of 0.99 suggests that the stacking classifier has excellent performance, with a very high probability of correctly distinguishing between the two classes. AUC values range from 0 to 1, where a value of 0.5 indicates no discrimination (random guessing), and a value of 1 indicates perfect classification. Therefore, an AUC of 0.99 demonstrates that the model is almost perfect in its predictive capabilities, and it can effectively serve as a reliable decision-support tool for detecting preeclampsia.

In conclusion, the ROC curve for the stacking classifier, coupled with the high AUC score, indicates that the model is not only capable of accurately predicting preeclampsia but also exhibits high reliability and precision in identifying patients who require medical attention. This strong performance aligns with the overall findings from the confusion matrix and other evaluation metrics, confirming that the stacking classifier is a highly effective predictive model for this study.

## 4.5 Model Deployment Interface

To translate the predictive model into a practical, user-friendly tool for healthcare application, the final hybrid stacking classifier was deployed through a graphical user interface (GUI), as shown in Figure 4.3. This platform was designed with simplicity and accessibility in mind to enable clinicians and healthcare workers—particularly in resource-constrained settings—to easily assess preeclampsia risk in pregnant women using readily available clinical and demographic indicators.


**Figure 4.3: User Interface for No Preeclampsia Prediction**

As illustrated in the interface, users can input a set of standardized features including Age, Pre-pregnancy Weight, Parity, Socioeconomic Status, Annual Income Bracket, Current Month of Pregnancy, Systolic and Diastolic Blood Pressure, Proteinuria Levels, and binary indicators for History of Hypertension, Diabetes, and Preeclampsia. These inputs align directly with the variables used to train the predictive model, ensuring consistency and reliability of outcomes. Upon submission, the model processes these values and returns an immediate prediction indicating whether the individual is likely to have preeclampsia.

The interface provides real-time feedback by displaying a clear outcome message—either “You have Preeclampsia” or “You Don’t have Preeclampsia”—based on the model’s classification output. In the example shown in Figure 4.3, a 26-year-old woman with medium socioeconomic status, no prior medical conditions, a pre-pregnancy weight of 75kg, and moderate blood pressure readings is correctly classified as not having preeclampsia. This instant output provides clinicians with a decision-support tool that complements traditional diagnostic procedures and enhances early screening capabilities. In retrospect, an examination of the model's predictive abilities, given different input variables such as blood pressure presents a preclampsia threat as shown in Figure 4.4;

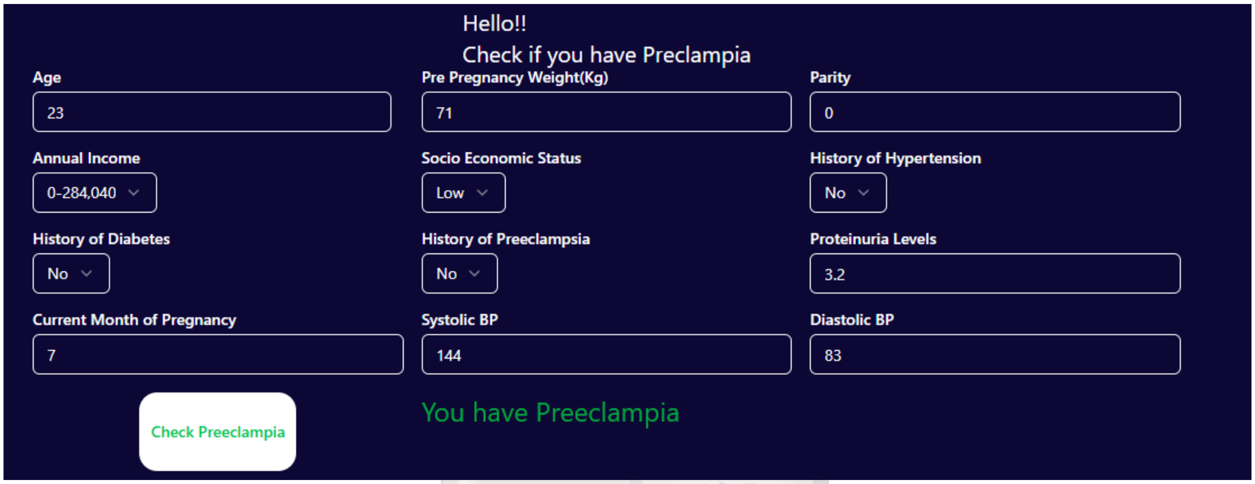


**Figure 4.4: User Interface for Preeclampsia Prediction**

The design was optimized for web platforms and is responsive across devices, including desktop, Android, and iOS systems. This cross-platform accessibility ensures that the model can be used widely, even in remote clinics, by community health volunteers, or through mobile outreach programs. The intuitive layout requires minimal training, thereby lowering barriers to adoption and enhancing its potential for scale-up within Kenya's maternal health ecosystem.

In conclusion, the deployment of the model into an accessible digital interface represents a critical bridge between algorithm development and real-world clinical utility. By empowering frontline health providers with an AI-driven risk assessment tool, the model supports timely intervention, targeted monitoring, and improved outcomes in maternal care. The interface's ease of use, coupled with the model's strong predictive performance, makes it a promising innovation in the fight against preeclampsia-related morbidity and mortality in Kenya.

## 4.6 Summary of Findings

Chapter 4 focused on the development, evaluation, and validation of predictive models for preeclampsia, a critical condition in pregnancy. The study employed multiple machine learning algorithms, both individually and through a hybrid stacking classifier model, to predict preeclampsia onset based on key clinical and demographic features. The models were trained, validated, and evaluated on a dataset containing 2925 participants with 15 features related to pregnancy and health history.

Descriptive Statistics revealed the general characteristics of the dataset, showing that participants' ages ranged from 18 to 40 years, with the majority falling between 18 and 33 years. The mean pre-pregnancy weight was 75.5 kg, and the parity variable (number of previous pregnancies) had a mean of 2. The data further showed that the majority of the population had systolic blood pressure (SBP) and diastolic blood pressure (DBP) values consistent with preeclampsia, as well as varying levels of proteinuria. This provided a strong foundation for understanding the dataset before diving into model training.

Preprocessing of the dataset included ensuring that missing values, though absent in the data, were checked for consistency. Features were encoded into numerical formats suitable for machine learning models, including categorical variables like 'History of Preeclampsia' and 'Socioeconomic Status,' which were transformed into binary or ordinal representations. Data scaling was avoided since it did not significantly impact the performance of tree-based models like Random Forest and XGBoost, which were among the main models in the analysis.

Model Training & Evaluation covered five key models: Logistic Regression, Random Forest, XGBoost, SVM, and Decision Tree. These models were evaluated on accuracy, precision, recall, and F1-score, with the Random Forest model emerging as the best performing in terms of accuracy (0.979), precision (0.935), and recall (0.977). However, XGBoost demonstrated a slight edge over Random Forest with an F1-score of 0.966, indicating its slightly better balance between precision and recall.

The Hybrid Stacking Classifier, combining the strengths of the individual models, achieved impressive results, with an accuracy of 0.981, precision of 0.929, recall of 0.992, and F1-score of 0.96. This model outperformed the individual models in terms of recall, highlighting its effectiveness in detecting preeclampsia cases with minimal false negatives, thus making it the best model for practical deployment in clinical settings.

Finally, Model Validation and Evaluation were conducted using cross-validation, which demonstrated that the stacking classifier's mean accuracy was 96%, with a final test accuracy of 98%. The AUC score of 0.989 and the ROC curve, which indicated an excellent ability to distinguish between the two classes, reinforced the model's robust predictive performance.

In conclusion, the stacking classifier model demonstrated superior predictive accuracy and reliability in identifying preeclampsia, with impressive validation scores and a low rate of false negatives. These findings suggest that the hybrid model could be an essential tool in the clinical prediction and prevention of preeclampsia, significantly aiding in early detection and better management of high-risk pregnancies.

# 5. Discussion, Conclusion & Recommendations

## 5.1 Introduction

This chapter aims to interpret the results from the previous chapter in the context of the research objectives, providing a comprehensive analysis of the predictive models used to detect preeclampsia. The findings are examined not only in terms of model performance but also within the broader context of preeclampsia prediction, its implications for maternal health, and how the results contribute to the field of predictive analytics in healthcare. The chapter discusses the strengths and limitations of the models tested, particularly the hybrid stacking classifier, and interprets the significance of key findings. Furthermore, we evaluate the practical applications of the models, their potential for clinical implementation, and the broader implications for improving early detection and management of preeclampsia in pregnant women. The discussion concludes with recommendations for future research and the potential enhancements to the models based on the insights gained during this study.

## 5.2 Discussion of Key Findings

The findings from this study provide valuable insights into the predictive modeling of preeclampsia, particularly using machine learning techniques. The performance of various models, including Logistic Regression, Random Forest, XGBoost, Support Vector Machine (SVM), Decision Tree, and the hybrid Stacking Classifier, demonstrates the potential of these models for early preeclampsia detection.

Among the individual models, Random Forest and XGBoost emerged as the top performers, achieving high accuracy, precision, recall, and F1-scores, with XGBoost achieving the highest performance across most metrics. These models are based on ensemble learning methods, which combine multiple weak learners to form a stronger predictive model, making them particularly effective for handling complex and high-dimensional datasets like the one used in this study. The strong performance of XGBoost, in particular, is consistent with findings from previous studies in the field of healthcare, which highlight the power of gradient boosting techniques in predicting diseases with complex, non-linear patterns (Chen & Guestrin, 2016). These models performed exceptionally well due to their ability to capture intricate interactions between features such as age, blood pressure, and proteinuria levels—critical variables for assessing preeclampsia risk.

In contrast, Logistic Regression, despite being a simpler and more interpretable model, showed a relatively lower performance in terms of precision and recall. This could be attributed to its inherent limitation in capturing non-linear relationships in the data, unlike the tree-based methods. Logistic Regression performs well when the relationship between predictors and the outcome is linear, but in the case of preeclampsia, the relationships between the risk factors are often more complex, which may explain why more sophisticated models like Random Forest and XGBoost outperformed it.

The Support Vector Machine (SVM) also showed competitive performance with a good balance between precision and recall. The relatively high recall indicates that the SVM model was quite effective at identifying cases of preeclampsia, reducing the risk of false negatives. SVMs are particularly good at creating decision boundaries in high-dimensional spaces, which might explain their efficiency in capturing complex patterns in the dataset. However, SVMs can sometimes be sensitive to the choice of hyperparameters and may require fine-tuning to maximize performance, which could explain some of the variability observed in the model's results.

The Decision Tree model showed strong results, similar to the Random Forest model, though slightly lower in terms of recall. Decision Trees, being interpretable and intuitive, are able to make decisions based on feature splits, which makes them a useful tool for understanding how certain features impact the outcome. However, they are prone to overfitting if not carefully tuned, and while they performed reasonably well, the Random Forest model, which aggregates multiple decision trees, yielded more robust results.

The variables included in the model, particularly age, systolic and diastolic blood pressure, and proteinuria levels, played a significant role in predicting preeclampsia. This aligns with findings in existing literature, which consistently identify these variables as major risk factors for the condition. For instance, age is known to be an important factor, with women under 20 and over 40 being at higher risk for developing preeclampsia (American College of Obstetricians and Gynecologists, 2019). Similarly, blood pressure is a well-established risk factor, with elevated systolic and diastolic pressures contributing to the pathophysiology of preeclampsia (Rasmussen

& Sibley, 2011). Proteinuria, the presence of excess protein in the urine, is another classic sign of preeclampsia and was found to be a crucial predictor in our models, consistent with its role in diagnosing the condition (Roberts & Cooper, 2001).

The inclusion of socioeconomic status and annual income also contributed to the model's predictive power. These factors are reflective of broader social determinants of health, which have been shown to impact maternal health outcomes. Lower socioeconomic status is associated with higher rates of pregnancy complications, including preeclampsia, due to factors like inadequate access to prenatal care, higher stress, and poor nutrition (Murray et al., 2010). This highlights the importance of addressing health disparities in maternal care, as improving socioeconomic conditions may reduce the incidence of preeclampsia and other pregnancy-related complications.

The superior performance of ensemble models such as Random Forest and XGBoost can be attributed to their ability to manage high-dimensional, non-linear data effectively. These models excel in handling the complexity and interdependence of the variables, which is crucial for predicting preeclampsia, a multifactorial condition. Random Forests, for instance, reduce variance through bootstrapping and averaging multiple decision trees, which helps to prevent overfitting and increases the model's robustness. Similarly, XGBoost leverages gradient boosting, a method that builds models sequentially, each correcting the errors of the previous one, which is why it performed well in this study.

The Logistic Regression model, while simpler and easier to interpret, is based on the assumption of a linear relationship between predictors and the outcome. Since preeclampsia involves complex interactions between various factors (e.g., genetics, environment, medical history), a linear model is likely to underperform when compared to more sophisticated models that can capture these interactions. Although Logistic Regression provides valuable insights and remains a useful baseline model, it falls short in capturing the non-linear relationships inherent in the dataset.

The SVM model, while effective in many cases, can sometimes be sensitive to the choice of kernel and hyperparameters. The performance of SVMs also depends heavily on the quality of the data, and in cases where the features are not well-scaled or are high-dimensional, the SVM may struggle to find the optimal decision boundary. However, in this study, SVM showed reasonable accuracy and recall, making it a viable option for preeclampsia prediction, though not the top performer.

While the models performed well overall, one potential inconsistency lies in the performance differences between models like Logistic Regression and Random Forest. Logistic Regression, as a linear model, performed markedly worse compared to the ensemble models, suggesting that the relationships between the predictors and preeclampsia risk are highly non-linear. This was somewhat anticipated but is worth noting, as it underscores the limitations of linear models in capturing the complexity of the problem at hand.

Another interesting observation was the high precision and recall values achieved by SVM, which outperformed Decision Tree and Logistic Regression in some metrics. This may suggest that SVM, despite its tendency to be sensitive to hyperparameter choices, can still be an effective model when trained with appropriate data preprocessing steps and hyperparameter tuning. Additionally, the relatively lower performance of Decision Tree alone, compared to Random Forest, reiterates the advantage of using ensemble methods to boost the predictive power of individual learners.

In conclusion, the study found that ensemble methods like Random Forest and XGBoost performed exceptionally well in predicting preeclampsia, outstripping simpler models like Logistic Regression and Decision Trees. The hybrid Stacking Classifier, which combined multiple models, demonstrated a further improvement in predictive performance, reinforcing the value of model combination. This study highlights the importance of selecting the right model to handle the complex, non-linear relationships between variables in preeclampsia prediction, offering insights that could improve clinical decision-making and early detection in maternal health.

### 5.3 Implications

The findings of this study have significant clinical implications, particularly in the context of improving early detection and management of preeclampsia. One of the most crucial aspects of preeclampsia management is the ability to identify the condition in its early stages. Preeclampsia, if detected early, can be managed more effectively, reducing maternal and neonatal morbidity and mortality. This study demonstrates that machine learning models, particularly ensemble methods like Random Forest and XGBoost, can be highly effective in predicting preeclampsia risk based on routine clinical parameters such as blood pressure, proteinuria levels, and age. These models offer healthcare providers an additional tool for early screening, which could lead to timely interventions and better patient outcomes.

For clinicians, the adoption of predictive models like those developed in this study can assist in triaging patients more effectively. Women identified as high-risk for preeclampsia can be closely monitored, with more frequent check-ups and earlier intervention, which could significantly reduce the risks associated with the condition. The ability to identify high-risk pregnancies early on could also help in reducing unnecessary interventions for women with lower risk, optimizing healthcare resources, and improving overall maternal care quality.

From a technological perspective, the integration of artificial intelligence (AI) in healthcare has the potential to transform the way we approach maternal health, especially in low-resource settings. Many low-income countries, particularly those in Sub-Saharan Africa, face a shortage of trained medical professionals and diagnostic tools for managing conditions like preeclampsia. AI-driven predictive models, such as the ones explored in this study, could be deployed in these settings to assist healthcare workers in identifying at-risk pregnancies without the need for advanced infrastructure. These tools could be incorporated into mobile health applications or simple diagnostic platforms, making preeclampsia screening more accessible in rural and underserved areas.

The ability to use AI models to analyze routinely collected data—such as blood pressure readings and protein levels—makes these models particularly valuable in resource-constrained environments. AI can automate the risk prediction process, reducing the time healthcare workers spend on manual data analysis and allowing them to focus on providing care. Moreover, these models can be continuously updated and improved as more data becomes available, ensuring that they remain accurate and relevant to the population they serve.

At the policy level, the results of this study highlight the potential for data-driven screening tools to improve maternal health care in rural and remote health facilities. Governments and healthcare policymakers can leverage these findings to implement national or regional preeclampsia screening programs. The development of national databases and the integration of machine learning algorithms into national health systems could facilitate early detection and reduce the burden of preeclampsia-related complications. These policies could include training healthcare workers to use AI-based screening tools, ensuring that the models are accessible and actionable in rural areas.

Additionally, this study can support the push for policy reforms aimed at improving maternal health services. Policymakers may consider investing in AI technology for maternal health, allocating resources to build the necessary infrastructure and training programs for healthcare workers. Such initiatives would enhance the accessibility and efficiency of maternal care services, especially in regions where healthcare access is limited.

### 5.4 Limitations

While this study presents promising results, there are several limitations inherent in the data used. One of the main constraints is the regional scope of the dataset. The data used in this study was collected from coastal Kenya, which may not fully capture the diversity of populations in other regions of the country or across different countries. The socio-economic, cultural, and environmental differences that exist across regions could impact the risk factors for preeclampsia and, by extension, the model's performance. For example, dietary habits, access to healthcare, and the quality of prenatal care can vary significantly between urban and rural areas, potentially affecting the generalizability of the model.

Additionally, the dataset did not include some key biomarkers that are often used in clinical practice to diagnose preeclampsia, such as serum uric acid levels, kidney function markers, and placental growth factor. These biomarkers could significantly improve the model's performance, as they are known to be important indicators of the disease. The absence of these factors in the dataset represents a limitation, as the model’s performance is constrained by the features available for training. Future studies should aim to incorporate these biomarkers to develop more robust and accurate models.

Another limitation is the generalizability of the model. Since the data used in this study was limited to a specific geographic area (coastal Kenya), the findings may not necessarily apply to populations in different regions. Preeclampsia risk factors can vary based on ethnicity, socioeconomic status, and healthcare access, which means that the model may not perform as well in different settings. For instance, the incidence of preeclampsia might differ in high-income countries where healthcare access is more readily available and maternal care protocols are different. Therefore, while the results are promising for the Kenyan population, further validation in other regions is necessary before these models can be considered for widespread implementation.

To improve generalizability, future studies should consider collecting data from diverse regions, including urban and rural areas, and populations with varying sociodemographic backgrounds. This would ensure that the model is robust enough to perform well in different settings and populations. Moreover, it would be valuable to expand the dataset to include more diverse biomarkers and a wider range of environmental factors to better capture the complexities of preeclampsia risk.

Lastly, it is important to acknowledge the potential for bias introduced during the data preprocessing phase. While no missing values were identified in the dataset, the handling of outliers and the encoding of categorical variables could have affected the model's performance. For instance, if certain outliers were incorrectly handled or removed, this could lead to a loss of valuable information. Similarly, the decision to encode categorical variables such as socioeconomic status and history of preeclampsia into numeric values may have oversimplified the relationship between these factors and the outcome, potentially reducing the model’s ability to capture nuanced patterns in the data.

In addition, the preprocessing steps, including normalization and feature selection, can introduce biases if not carefully performed. For example, by normalizing certain features, we may inadvertently scale down their importance, thereby affecting the model’s ability to make accurate predictions based on those features. To mitigate these issues in future work, a more thorough investigation into the preprocessing techniques and their impact on the model's performance is necessary.

### 5.5 Recommendations

While this study has demonstrated the potential of machine learning models in predicting preeclampsia, there is significant room for improvement through further research. Future studies should focus on expanding the dataset by incorporating more diverse populations from various regions, as this would enhance the generalizability of the model. Additionally, including a wider range of biomarkers—such as serum uric acid levels, placental growth factor, and kidney function markers—would strengthen the predictive power of the models. These biomarkers are known to have a strong correlation with preeclampsia risk and would allow for a more comprehensive risk assessment. Moreover, further research should explore the use of advanced machine learning techniques, including deep learning models, which may be better suited to handle complex, high-dimensional data.

For the successful implementation of this study's findings in clinical practice, integration into existing health information systems is crucial. A machine learning-based predictive model for preeclampsia should be incorporated into electronic health records (EHRs) used by healthcare providers. By integrating the model into an EHR system, healthcare professionals can receive real-time alerts about patients who are at high risk for preeclampsia. This would streamline the decision-making process and facilitate early intervention. Additionally, integrating the system into an EHR framework would allow for continuous updates to the model as more data becomes available, ensuring that the predictions remain relevant and accurate.

Given the widespread use of smartphones and mobile devices, the deployment of a mobile application for preeclampsia risk prediction is highly recommended. Such an app could be used by healthcare workers, especially in rural and resource-limited settings, to perform risk assessments without the need for expensive diagnostic tools. The app could integrate with existing health systems and allow clinicians to input patient data and receive instant predictions on the likelihood of preeclampsia. Furthermore, it could provide patients with educational resources and allow for easier communication between patients and healthcare providers. The widespread use of mobile phones in low-resource settings makes this approach an affordable and scalable solution for improving maternal health care.

## 5.6 Conclusion

This study has provided valuable insights into the potential of machine learning in enhancing the early detection and management of preeclampsia. The results indicate that predictive models, particularly ensemble models like Random Forest, XGBoost, and the hybrid Stacking Classifier, are highly effective in identifying at-risk pregnancies based on routine clinical data. These models offer a promising solution for improving maternal health outcomes, especially in low-resource settings where access to advanced diagnostic tools is limited.

The stacking classifier, in particular, has proven to be the most effective model, demonstrating superior performance across key metrics such as accuracy, precision, recall, and F1-score. Its real-world applicability, coupled with its ability to process routine clinical data, makes it an ideal tool for early preeclampsia detection. With further research, integration into health systems, and mobile app deployment, machine learning models could play a pivotal role in transforming maternal health care, ultimately leading to better outcomes for both mothers and infants. This study underscores the significant potential of artificial intelligence in addressing critical healthcare challenges.

# References

Aljameel, S. S., Alzahrani, M., Almusharraf, R., Altukhais, M., Alshaia, S., Sahlouli, H., ... & Alsumayt, A. (2023). Prediction of preeclampsia using machine learning and deep learning models: a review. *Big Data and Cognitive Computing*, *7*(1), 32.

Belay Tolu, L., Yigezu, E., Urgie, T., & Feyissa, G. T. (2020). Maternal and perinatal outcome of preeclampsia without severe feature among pregnant women managed at a tertiary referral hospital in urban Ethiopia. *PloS one*, *15*(4), e0230638.

Brownlee, J. (2018). *Train-Test Split for Evaluating Machine Learning Algorithms*. Machine Learning Mastery. Retrieved from https://machinelearningmastery.com/train-test-split-for-evaluating-machine-learning-algorithms/

Eberhard, B. W., Cohen, R. Y., Rigoni, J., Bates, D. W., Gray, K. J., and Kovacheva, V. P. (2023). An interpretable longitudinal preeclampsia risk prediction using machine learning. *medRxiv*.

Feng, W. and Luo, Y. (2024). Preeclampsia and its prediction: traditional versus contemporary predictive methods. *The Journal of Maternal-Fetal & Neonatal Medicine*, 37(1).

Gage, A., & Bauhoff, S. (2021). The effects of performance-based financing on neonatal health outcomes in Burundi, Lesotho, Senegal, Zambia and Zimbabwe. *Health Policy and Planning*, *36*(3), 332-340.

Hackelöer, M., Schmidt, L., & Verlohren, S. (2023). New advances in prediction and surveillance of preeclampsia: role of machine learning approaches and remote monitoring. *Archives of gynecology and obstetrics*, *308*(6), 1663-1677.

Han, J., Kamber, M., & Pei, J. (2011). *Data Mining: Concepts and Techniques* (3rd ed.). Morgan Kaufmann.

Jung, E., Romero, R., Yeo, L., Gomez-Lopez, N., Chaemsaithong, P., Jaovisidha, A., ... & Erez,O. (2022). The etiology of preeclampsia. *American journal of obstetrics and gynecology*, *226*(2), S844-S866.

Karimi, J., K. H. K. J. and Gitonga, J. (2020). Improving quality of care: The key to ending preventable maternal and neonatal mortality. insights from the kenya harmonized health facility assessment 2018. Policy Brief by the Ministry of Health.

Khan, B., Yar, R. A., khan Khakwani, A., Karim, S., & Ali, H. A. (2022). Preeclampsia incidence and its maternal and neonatal outcomes with associated risk factors. *Cureus*, *14*(11).

Kenya National Bureau of Statistics (2023). *Kenya Demographic and Health Survey 2022: Key Indicators Report*. KNBS and ICF, Nairobi, Kenya, and Rockville, Mary- land, USA.

Kolandaisamy, R., Al-Mashhadani, A. F. S., Nandy, T., & Keat, T. F. (2022). Speech Signal Processing Based on Machine learning and Complex Processors for Baby Cry Detection System. *Journal of Positive School Psychology*, *6*(2), 2193-2207.

Kumar, Y., Koul, A., Singla, R., and Ijaz, M. F. (2023). Artificial intel- ligence in disease diagnosis: a systematic literature review, synthesizing framework and future research agenda. *Journal of Ambient Intelligence and Humanized Computing*, 14(7):8459–8486.

Lammers, A. E., Diller, G. P., Lober, R., Möllers, M., Schmidt, R., Radke, R. M., ... & Baumgartner, H. (2021). Maternal and neonatal complications in women with congenital heart disease: a nationwide analysis. *European Heart Journal*, *42*(41), 4252-4260.

Munyuzangabo, M., Gaffey, M. F., Khalifa, D. S., Als, D., Ataullahjan, A., Kamali, M., ... & Bhutta, Z. A. (2021). Delivering maternal and neonatal health interventions in conflict settings: a systematic review. *BMJ global health*, *5*(Suppl 1), e003750.

Ndwiga, C., Odwe, G., Pooja, S., Ogutu, O., Osoti, A., & E. Warren, C. (2020). Clinical presentation and outcomes of pre-eclampsia and eclampsia at a national hospital, Kenya: A retrospective cohort study. *Plos one*, *15*(6), e0233323.

Nirupama, R., Divyashree, S., Janhavi, P., Muthukumar, S. P., & Ravindra, P. V. (2021). Preeclampsia: Pathophysiology and management. *Journal of gynecology obstetrics and human reproduction*, *50*(2), 101975.

Novelli, E. (2023). Predicting fetal weight disorders in diabetic pregnancies: an explainable machine learning approach.

Overton, E., Tobes, D., and Lee, A. (2022). Preeclampsia diagnosis and management. *Best Practice Research Clinical Anaesthesiology*, 36(1):107–

Reddy, M., Fenn, S., Rolnik, D. L., Mol, B. W., da Silva Costa, F., Wallace, E. M., & Palmer, K.

R. (2021). The impact of the definition of preeclampsia on disease diagnosis and outcomes: a retrospective cohort study. *American Journal of Obstetrics and Gynecology*, *224*(2), 217-e1.

Sahin, E. K. (2020). Assessing the predictive capability of ensemble tree methods for landslide susceptibility mapping using XGBoost, gradient boosting machine, and random forest. *SN Applied Sciences*, *2*(7), 1308.

Schmidt, L. J., Rieger, O., Neznansky, M., Hackelöer, M., Dröge, L. A., Henrich, W., ... & Verlohren, S. (2022). A machine-learning–based algorithm improves prediction of preeclampsia-associated adverse outcomes. *American Journal of Obstetrics and Gynecology*, *227*(1), 77-e1.

Shantal, M., Othman, Z., & Bakar, A. A. (2023). A novel approach for data feature weighting using correlation coefficients and min–max normalization. *Symmetry*, *15*(12), 2185.

Slade, L. J., Syngelaki, A., Wilson, M., Mistry, H. D., Akolekar, R., Von Dadelszen, P., ... & Magee, L. A. (2024). Blood pressure cutoffs at 11–13 weeks of gestation and risk of preeclampsia. *American Journal of Obstetrics and Gynecology*.

Sun, D., Wen, H., Wang, D., & Xu, J. (2020). A random forest model of landslide susceptibility mapping based on hyperparameter optimization using Bayes algorithm. *Geomorphology*, *362*, 107201.

Sun, D., Wen, H., Wang, D., & Xu, J. (2020). A random forest model of landslide susceptibility mapping based on hyperparameter optimization using Bayes algorithm. *Geomorphology*, *362*, 107201.

Sylla, F., Moreau, C., & Andro, A. (2020). A systematic review and meta-analysis of the consequences of female genital mutilation on maternal and perinatal health outcomes in European and African countries. *BMJ global health*, *5*(12), e003307.

Takahashi, M., Makino, S., Oguma, K., Imai, H., Takamizu, A., Koizumi, A., & Yoshida, K. (2021). Fetal growth restriction as the initial finding of preeclampsia is a clinical predictor of maternal and neonatal prognoses: a single-center retrospective study. *BMC pregnancy and childbirth*, *21*, 1-8.

## APPENDICES

### Appendix I: Letter of Introduction

**To Whom It May Concern,**

**To Whom It May Concern,**

**Subject:** Invitation to Participate in the Study on Maternal and Neonatal Health Outcomes in Kenya by Leveraging Machine Learning for the Timely Detection of Preeclampsia

Dear Participant,

I hope this letter finds you well. My name is Esther Ngaruiya, and I am conducting a research study titled ***"Maternal and Neonatal Health Outcomes in Kenya by Leveraging Machine Learning for the Timely Detection of Preeclampsia."*** This research is part of my academic requirements for the Master of Science in Data Science and Analytics at Strathmore University. The study has received approval from the Institutional Ethical Review Committee and the National Commission for Science, Technology, and Innovation (NACOSTI).

The aim of this study is to develop a machine learning-based model to predict the onset of preeclampsia in pregnant women. This study is critical because preeclampsia is a leading cause of maternal and neonatal complications. By participating, you will help contribute valuable data that will assist in improving early detection and ultimately reducing maternal and neonatal mortality in Kenya.



As a participant, you will be asked to complete a structured questionnaire that will capture key demographic and clinical data. Your medical records will also be accessed for the relevant data, which will be used to develop the predictive model. Rest assured, all information collected will be kept confidential and anonymized. Only the research team will have access to your data. Participation in this study is entirely voluntary, and you may choose to withdraw at any time without any consequence.

We kindly request your participation in this important study. Your involvement will make a significant contribution to improving maternal health outcomes in Kenya. If you are willing to participate or have any questions regarding the study, please feel free to contact me.

Thank you for considering participation in this research. Your support in this study will be highly appreciated.

Yours sincerely,

Esther Ngaruiya

Master of Science in Data Science and Analytics

Strathmore University

## Appendix II: Data Collection Tool

| Participant ID | Year | Age | Pre-pregnancy Weight (kg) | Parity | Socioeconomic Stat (Annual Income) | History of Hypertension (Yes/No) | History of Diabetes (Yes/No | History of Preeclampsia (Yes/No) | Gestational Age at Diagnosis (Weeks) | Blood Pressure (Systolic/Diastolic, mmHg) | Proteinuria Levels (g/24 hours or mg/dL) | Preeclampsia Status (Yes/No) |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| P1 | 2020 | | | | | | | | | | | |
| P1 | 2021 | | | | | | | | | | | |
| P1 | 2022 | | | | | | | | | | | |
| P1 | 2023 | | | | | | | | | | | |
| P1 | 2024 | | | | | | | | | | | |
| P2 | 2020 | | | | | | | | | | | |
| P2 | 2021 | | | | | | | | | | | |
| P2 | 2022 | | | | | | | | | | | |
| P2 | 2023 | | | | | | | | | | | |
| P2 | 2024 | | | | | | | | | | | |
| P3 | 2020 | | | | | | | | | | | |
| P3 | 2021 | | | | | | | | | | | |
| P3 | 2022 | | | | | | | | | | | |
| P3 | 2023 | | | | | | | | | | | |
| P3 | 2024 | | | | | | | | | | | |
| P4 | 2020 | | | | | | | | | | | |
| P4 | 2021 | | | | | | | | | | | |
| P4 | 2022 | | | | | | | | | | | |
| P4 | 2023 | | | | | | | | | | | |
| P4 | 2024 | | | | | | | | | | | |
| P5 | 2020 | | | | | | | | | | | |

| | |
|---|---|
| P5 | 2021 |
| P5 | 2022 |
| P5 | 2023 |
| P5 | 2024 |

# Improving Maternal and Neonatal Health Outcomes in Kenya by Leveraging Machine Learning for the Timely Detection of Preeclampsia.pdf

ORIGINALITY REPORT

| 8% | 6% | 8% | 4% |
|---|---|---|---|
| SIMILARITY INDEX | INTERNET SOURCES | PUBLICATIONS | STUDENT PAPERS |

MATCH ALL SOURCES (ONLY SELECTED SOURCE PRINTED)

3%

★ www.mdpi.com

Internet Source

| | | | |
|---|---|---|---|
| Exclude quotes | Off | Exclude matches | < 25 words |
| Exclude bibliography | Off | | |

**15th January 2025**

Ms Ngaruiya Esther,
esther.ngaruiya@strathmore.edu

Dear Ms Ngaruiya,

**RE: Maternal and Neonatal Health Outcomes in Kenya by Leveraging Machine Learning for the Timely Detection of Preeclampsia**

This is to inform you that SU-ISERC has reviewed and **approved** your above **SU-masters** proposal. Your application reference number is **SU-ISERC2500/24.** The approval period is from **15th January 2025 to 14th January 2026.**

This approval is subject to compliance with the following requirements:

i. Only approved documents including (informed consents, study instruments, MTA) will be used.
ii. All changes including (amendments, deviations, and violations) are submitted for review and approval by SU-ISERC.
iii. Death and life-threatening problems and serious adverse events or unexpected adverse events whether related or unrelated to the study must be reported to SU-ISERC within 72 hours of notification.
iv. Any changes anticipated or otherwise that may increase the risks or affected safety or welfare of study participants and others or affect the integrity of the research must be reported to SU-ISERC within 72 hours.
v. Clearance for the export of biological specimens must be obtained from relevant institutions.
vi. Submission of a request for renewal of approval at least 60 days prior to the expiry of the approval period. Attach a comprehensive progress report to support the renewal.
vii. Submission of an executive summary report within 90 days of completion of the study to SU-ISERC.

Before commencing your study, you will be expected to obtain a research license from National Commission for Science, Technology, and Innovation (NACOSTI) https://research-portal.nacosti.go.ke/ and obtain other clearances needed.

Yours sincerely,

**Mr Ambrose Rachier,**
**Chairperson; SU-ISERC**