# Understanding Risks for Maternal Mortality in Rural Bangladesh Using XGBoost, Random Forest, and Decision Tree ML Models

MSc Research Project
Data Analytics

Temitope Oladimeji
Student ID: x23187204

School of Computing
National College of Ireland

Supervisor: Anderson Simiscuka

**National College of Ireland**
**Project Submission Sheet**
**School of Computing**

| | |
|---|---|
| **Student Name:** | Temitope Oladimeji |
| **Student ID:** | x23187204 |
| **Programme:** | Data Analytics |
| **Year:** | 2024 |
| **Module:** | MSc Research Project |
| **Supervisor:** | Anderson Simiscuka |
| **Submission Due Date:** | 16/09/2024 |
| **Project Title:** | Understanding Risks for Maternal Mortality in Rural Bangladesh Using XGBoost, Random Forest, and Decision Tree ML Models |
| **Word Count:** | 6884 |
| **Page Count:** | 26 |

I hereby certify that the information contained in this (my submission) is information pertaining to research I conducted for this project. All information other than my own contribution will be fully referenced and listed in the relevant bibliography section at the rear of the project.

**ALL** internet material must be referenced in the bibliography section. Students are required to use the Referencing Standard specified in the report template. To use other author's written or electronic work is illegal (plagiarism) and may result in disciplinary action.

| | |
|---|---|
| **Signature:** | Temitope Oladimeji |
| **Date:** | 14th September 2024 |

**PLEASE READ THE FOLLOWING INSTRUCTIONS AND CHECKLIST:**

| | |
|---|---|
| Attach a completed copy of this sheet to each project (including multiple copies). | ☐ |
| **Attach a Moodle submission receipt of the online project submission**, to each project (including multiple copies). | ☐ |
| **You must ensure that you retain a HARD COPY of the project**, both for your own reference and in case a project is lost or mislaid. It is not sufficient to keep a copy on computer. | ☐ |

Assignments that are submitted to the Programme Coordinator office must be placed into the assignment box located outside the office.

| **Office Use Only** | |
|---|---|
| Signature: | |
| Date: | |
| Penalty Applied (if applicable): | |

# Understanding Risks for Maternal Mortality in Rural Bangladesh Using XGBoost, Random Forest, and Decision Tree ML Models

Temitope Oladimeji
x23187204

**Abstract**

This paper explores the application of machine learning models for predicting pregnancy risks, focusing on the performance comparison of XGBoost, Random Forest, and Decision Tree classifiers. The motivation behind this research stems from the critical need for early identification of high-risk pregnancies to improve maternal health outcomes. Using a dataset consisting of anonymous information from pregnant women in rural Bangladesh, this study implements feature scaling, standardization, and encoding to prepare the data. Both pre- and post-hyperparameter tuning results are analysed, with additional focus on handling imbalanced data through the application of SMOTE (Synthetic Minority Oversampling Technique). The evaluation metrics include accuracy, precision, recall, F1-score, and ROC curves for each class. Key findings indicate that XGBoost outperforms the other models, particularly after hyperparameter tuning and SMOTE application, achieving an accuracy of 82%. The study emphasizes the importance of advanced machine learning techniques in healthcare, offering significant implications for early and accurate prediction of pregnancy-related risks.

# 1 Introduction

## 1.1 Background and Motivation

According to the World Health Organization [1], over 80% of 140 million annual births occur at proper medical centres, nearly double the rate from thirty years ago, reflecting significant progress. However, despite a 38% reduction in maternal mortality over the past two decades, the annual reduction rate is a slow 3%. At this pace, the Sustainable Development Goals (SDGs) 2030 target of reducing maternal deaths to fewer than 70 per 100,000 births is unlikely to be met. These statistics highlight the need for effective solutions to address this global health problem as soon as possible.

Perhaps the most widely discussed and often controversial solution to this problem is the use of Artificial Intelligence (AI). Machine Learning (ML), a subset of AI, has seen advancements in health domains. Besides its high diagnostic accuracy, AI offers other benefits. Complex algorithms can easily detect patterns and correlations in medical data that might be overlooked by human analysis, leading to new insights into diseases, drug

[1] https://www.who.int/health-topics/maternal-health#tab=tab_2

side effects, and interactions between compounds and the human body. AI systems also enable continuous patient monitoring, alerting doctors to any detected anomalies - a common feature during the COVID-19 pandemic to minimize physical contact (Khan et al., 2024).

The quality of, and access to healthcare services are major factors contributing to poor maternal health. Even well-developed countries like the United States struggle to provide equal access to all. Developing countries, such as Bangladesh, experience even bigger challenges, particularly in rural areas (Reza et al., 2024). AI can support health targets and help bridge gaps in healthcare access. By identifying high-risk patients with high precision, AI can significantly reduce maternal deaths. This study aims to contribute to these efforts by using ML algorithms to understand maternal mortality risk levels in rural Bangladesh, thereby supporting the SDGs' objectives of improving maternal health.

## 1.2 Research Importance

Understanding the interactions between different factors, and the impact each risk factor has on the target variable is important in order to develop an effective predictive model. Previous studies have identified different risk factors such as age, systolic and diastolic blood pressure, blood sugar levels, body temperature, and heart rate. This study focuses on a novel approach that integrates these factors using advanced ML techniques, to improve the prediction of maternal mortality risk.

## 1.3 Research Question and Objectives

*"How effective are XGBoost, Random Forest and Decision Tree compared to each other in terms of accuracy, recall, precision, and F1-score in understanding maternal mortality in rural Bangladesh based on risk factors such as age, hypertension, diabetes, temperature and heart rate?"*

To address this question, the study will:

1. Analyse various maternal health factors such as hypertension, diabetes, temperature, and age, to determine their individual and combined effects on maternal mortality risk in rural Bangladesh.
2. Develop and train XGBoost, Random Forest, and Decision Tree machine learning models using the identified factors, optimizing performance through data preprocessing and hyperparameter tuning.
3. Evaluate the predictive capabilities of the models using metrics such as accuracy, recall, precision, and F1 score to ensure a complete assessment of its performance.
4. Compare the performance of the models to validate its reliability and effectiveness in understanding maternal mortality risk factor.

## 1.4 Research Limitations, Assumptions and Plan

Several limitations are acknowledged in this study with the main one being the generalizability of the findings. The dataset used focuses on Bangladesh and includes variables commonly associated with maternal mortality in that region, potentially introducing a regional bias. Despite these limitations, the study assumes that the selected factors are sufficient enough to develop a good assessment model.

The report is organized into six sections: Section 2 reviews previous research on predictive modeling and maternal mortality risk factors. Section 3 provides a detailed description of the methodology. Section 4 presents the design specification. Section 5 describes the final stages of the implementation. Section 6 summarizes the findings and expands on the study's limitations. Section 7 provides recommendations for future research.

# 2 Review of Related Works

The related works are organized in a few subsections. 2.1 gives an overview of maternal health, including the major contributions to maternal mortality in Bangladesh. Then in 2.2, the various risk factors are critically examined based on previous work. 2.3 reviews traditional risk assessment methods and the use of AI in healthcare. 2.4 discusses the application of ML models in maternal mortality, providing a comparative analysis of XGBoost, Random Forest and Decision Tree.

## 2.1 Maternal Health in Bangladesh

In 2017, about 94% of maternal deaths occurred in low and middle-income countries due to limited medical facilities (Rumbeli et al., 2024). In Africa alone, more than half of these deaths were recorded. Although maternal mortality remains a crisis, Bangladesh has seen a decrease in rates since the 1990s. In rural areas, many women are forced to give birth at home due to poor conditions. Cost of care and shortage of equipment are significant barriers for women in these regions (Chowdhury et al., 2023). Those who can access health facilities often lack the means for additional services such as antenatal care. The COVID-19 pandemic prompted the introduction of remote antenatal care (ANC) services in many hospitals to support pregnant women and prevent the spread of the disease in medical centres, which has had a lasting impact by reducing delayed visits in rural areas (Islam et al., 2023; Hossain et al., 2024).

## 2.2 Maternal Health Risk Factors

This subsection critically reviews previous works that have discussed age-related risks, and the impact of physiological measurements in maternal health.

### 2.2.1 Impact of Age on Maternal Health

Londero et al. (2024)'s study, which relied on past datasets, demonstrated a significant correlation between increasing maternal age and the risk of developing breast cancer using a large study sample and kernel smoothing for density distribution, finding that pregnant women over 35 and teenagers are more susceptible to premature birth and mortality. With a bias towards Western population and varied sample sizes which complicated the identification of specific effects, Ahmad et al. (2024) highlighted the increasing chances of pregnancy complications with age, confirming higher rates of diabetes and hypertension in older mothers but noting improved emotional comfort for the child. In contrast, Wan et al. (2024)'s study focused on the impact of delayed pregnancy on cognitive development in rural areas, showing a correlation between older maternal age and better cognitive development using data from an antenatal trial in western China involving 1897 participants, though the study's limited inclusion of mothers over 35 and specific regional focus affect broader applicability.

### 2.2.2 The importance of monitoring blood pressure

A healthy blood pressure reading is considered to be 120/80 mm Hg. Even slight increases can be concerning, with readings of at least 140/90 mm Hg indicating hypertension. During pregnancy, high blood pressure can cause pre-eclampsia, a serious condition that can affect fetal development (Strahm et al., 2024).

Acherjya et al. (2023)'s study, involving 1812 participants from rural Bangladesh aged over 18, reported a hypertension frequency of 20.6% and a high normal blood pressure of 9%, using the STEPS surveillance tool for data collection. It found correlation between blood pressure and age, evaluating risk factors across different groups using statistical methods such as logistic regression and ANOVA test.

High blood-pressure related complications can be difficult to identify early. Currently, patients are treated through prolonged hospital stays and frequent clinical appointments, which may not be feasible in rural areas, leading to damaging consequences. Therefore, monitoring the condition is important. AI tools are under development for this purpose (Rosales et al., 2024), though existing tools can still assist doctors in their assessments. For example, D. Jones et al. (2023) measured the impact on twelve pregnant women of various demographics in rural areas and their overall satisfaction with a remote blood pressure monitoring device. Real-time data tracking provided reassurance that any disturbances would be noticed by an expert. Although there were concerns about the accuracy and frequency of calls, the use of AI to monitor these conditions is an emerging trend, given the capabilities and benefits it offers in healthcare.

### 2.2.3 Impact of Blood Sugar levels on Pregnancy Outcomes

The long-term effects of diabetes can potentially lead to the loss of bodily functions. Though it cannot be cured, lifelong medications slow its progression significantly (Chen et al., 2024).

Setiawan et al. (2023) conducted a study on predicting diabetes using machine learning models, emphasizing the high prevalence of diabetes in Southeast Asia and the increased risk of offspring developing the disease if the mother is affected. Using the PIMA Indian diabetes dataset from Kaggle, which includes 786 female patients above 21, the study employed a neural network with two hidden layers, recognized as a favourable approach for diabetes analysis. The dataset was balanced using SMOTE and tested using various machine learning algorithms including Random Forest, Logistic Regression, Support Vector Machines, Decision Trees, and Naive Bayes. Random Forest proved the most accurate, with evaluation metrics such as F1-score and accuracy at 82.9% and 87.9%, respectively. A key strength of the study was its detailed comparison of multiple algorithms, providing reproducibility of the findings. The methodology and findings are also relevant to understanding related chronic conditions, including hypertension, given the shared risk factors.

Pregnant women with diabetes face many health risks, including susceptibility to infections and the possibility of the offspring dying. Wierzchowska-Opoka et al. (2023)'s study demonstrated the protective role of vaccinations, showing an 80% improvement in health after 20 weeks. The COVID-19 lockdown led to unhealthy eating and physical habits, increasing the risk of high sugar intake and insulin resistance. The study's trial involving 127 pregnant women who received the vaccine showed no severe complications compared to those who did not receive the vaccine.

#### 2.2.4 Effect of Body Temperature on Maternal Health

The World Health Organization has identified climate change as the number one threat to human health this century. While developed countries experience very few direct deaths from extreme weather, the broader impact of global warming - including food and water scarcity - poses significant health risks worldwide, especially in lesser developed regions. High temperatures affect the body's ability to regulate its temperature, posing risks for vulnerable populations such as pregnant women. Pregnancy naturally raises body temperature, and more heat exposure can increase the risk of adverse outcomes like low birth weight. Mitigating the impact of these environmental changes on body temperature and health has become a priority, as climate changes worsens.

These concerns were examined more closely in a cohort study in Yunnan, China (Wu et al., 2023). Using data from the China National Meteorological Information Centre, the study focused on women aged 18 to 49 and found that high temperatures and high humidity increased the risk of preterm birth. The study suggested that high temperatures might raise cholesterol levels. Studies have also found that low temperatures could have similarly detrimental effects which can impact delivery (Min et al., 2024).

#### 2.2.5 Heart rate Monitoring on Maternal Health

The variability of the heart rate is another topic of concern when it comes to maternal health. Not only does it affect the mother, but research has shown that emotional stress can negatively impact the child's development too (Semeia et al., 2023).

In Bester et al. (2023)'s study, logistic regression models were used to analyse the impact of heart rate variability on pregnancy by comparing cardiac activity from wrist monitors in healthy pregnant and non-pregnant women. Correlation was found between pregnancy complications and an irregular heart rate, shown with an area under the receiver operating characteristic (AUROC) value of 0.825.

### 2.3 AI in Healthcare

The predictive capabilities of AI algorithms is a contributing factor to its use in the healthcare industry.

#### 2.3.1 Background

Over the years, various risk assessment techniques have been introduced to improve the health of mothers and infants. In rural areas especially, these techniques have been implemented to bridge the gap between healthcare needs and available resources (Hababa and Assarag, 2023). Training local health workers to provide pre and post-natal care have shown positive results. Medical case notes is highly advised by the WHO in order for health workers to have as much patient information as possible and give the best treatment (Mndala et al., 2024).

Traditionally, health risk assessment methods have relied mostly on clinical observations and patients' history. High risk individuals have been categorized based on factors such as pre-existing medical conditions or, in the case of maternal health, obstetric history. However, a major drawback of these methods is the delayed identification of risks and the inability to capture complex interactions between different risk factors (Doctor and MacEwan, 2017). Traditional methods also often require a high level of clinical expertise and are dependent on the availability of skilled health personnel, which is scarce in rural areas. To overcome these limitations, advancements have been made in the development of predictive tools that are able to identify complex patterns, detect symptoms early, and predict risks with high accuracy.

Machine learning models can analyse data with superior speed and accuracy compared to even an expert. Patterns that could go unnoticed from human analysis can now be easily identified, leading to more precise predictions and tailored patient plans if necessary. Not only does this save time but the high degree of accurate results can give comfort to patients. This, along with the discovery of drugs, is considered a breakthrough in the advancement of AI (Alowais et al., 2023). The ability to learn and adapt from new data is crucial for improved performance. The interaction of factors can be analysed to provide even more comprehensive analysis. A recent study by Ardeti et al. (2023) highlights the superiority of AI over traditional practices, demonstrating how machine learning models outperformed standard clinical assessments in detecting cardiovascular diseases.

### 2.3.2 Ethical Concerns Surrounding AI

Patient anonymity and consent are important things to consider when it comes to the ethical concerns surrounding the use of AI in the medical field. Many companies prioritize patients' privacy, and preach about transparency for the purpose of AI to be understood. Despite these practices however, many are still unsure about the complete takeover of AI especially in healthcare settings, with a vast majority demanding AI be used as an assistive tool rather than being in full control of the assessment (Hallowell et al., 2023). There is also a high chance of algorithmic bias which might result in patients suffering from racial discrimination. The collection of data primarily introduces bias into the model, and medical frameworks exist to overcome this issue. However, the implications it could have is still widely debated as many believe the benefits it has outweigh any existing partiality (Aquino et al., 2023).

### 2.3.3 AI-based Techniques in Healthcare

There are many AI proposed models used in healthcare. For example, Convolution Neural Network (CNN) models have been used for cancer detection (Haq et al., 2023) and Deep Learning models have been applied to detect cardiovascular diseases (Khanna et al., 2023). Machine Learning models have been used to process textual data due to their ability to achieve higher accuracy and analyse large amounts of data. Popular machine learning models include XGBoost, Random Forest, Naive Bayes, K-nearest neighbours (KNN) and Decision Tree.

## 2.4 Applying Machine Learning Models in Mortality Risk Analysis

As this study involves textual medical data, ML models have been chosen as the most suitable approach for processing. The three ML models to be evaluated for predicting maternal health risks - Decision Tree, Random Forest, and XGBoost - each have their own advantages. Decision Tree is ideal for use where interpretability and computation efficiency are prioritized. XGBoost is ideal for use where predictive accuracy is crucial. Random Forest offers a balanced approach of both. For rural areas in Bangladesh, a careful evaluation of performance metrics will guide the choice of the most appropriate model to improve maternal health.

### 2.4.1 XGBoost

An ensemble learning algorithm, XGBoost (Extreme Gradient Boosting) is a type of decision tree model primarily used for regression and classification problems. It works by combining multiple ”weak” models into one to form a ”strong” model in such a way that errors are minimized (Price et al., 2022). Training multiple trees allows the next model to learn from the errors of the previous model, producing more accurate results and reducing the chances of over-fitting, which is when a model is unable to be used on any data besides the training data.

A recent study (Maheswari et al., 2024) applied XGBoost to predict maternal risk levels. The dataset contained 1014 records with six features: Age, systolic and diastolic blood pressure, blood sugar levels, heart rate and risk levels. To ensure a more balanced distribution of the data, the Synthetic Minority Over-Sampling (SMOTE) was employed. The data was split into an 80/20 ratio, resulting in 809 observations for training and 203 for testing. Out of 15 machine learning algorithms evaluated, XGBoost outperformed all others, achieving an accuracy of 88.9%. This high accuracy suggests XGBoost can provide reliable assessments, and its ability to handle large medical records makes it suitable for use in maternal health care.

### 2.4.2 Random Forest

Random Forest is an ensemble learning method that creates a ”forest” of decision trees. Similar to the XGBoost algorithm, Random Forest can be used for classification and regression tasks. The construct of multiple decision trees improves its accuracy and also reduces the chances of over-fitting (Mihali and Niță, 2024).

A study (Alamsyah et al., 2023) aimed at categorizing maternal health risks utilized Random Forest to analyse a dataset containing key risk factors including age, blood sugar, body temperature, heart rate, blood pressure and the risk level. To enhance the model’s performance, feature optimization was applied. The study compared the performance of five machine learning models: Decision Tree, Random Forest, Neural Network, K-Nearest Neighbours (KNN), and Naïve Bayes. Random Forest was the best-performing model, achieving an accuracy of 73.37% on a 90/10 training and testing split. Interestingly, the accuracy improved slightly to 74.07% on a 60/40 split, while KNN was the next best model with an accuracy of 71.99%.

### 2.4.3 Decision Tree

Decision Tree is a supervised learning algorithm that is suitable for classification and regression problems. As the name suggest, decision trees have "tree-like" structures beginning with a node at its base and outgoing branches called decision nodes that eventually feed into other branching nodes, called leaf nodes, that represents the final outcome (Hou et al., 2023). Decision trees are susceptible to over-fitting especially with increasing number of branches and nodes, and they usually perform slightly worse than ensemble methods such as XGBoost and Random Forest. However, it is easily interpretable and can be used to simplify the tasks for other machine learning algorithms.

In a study comparing the effectiveness of Decision Tree and K-Nearest Neighbours models in predicting maternal health risks, KNN was found to be more suitable with an accuracy of 71.5% while the Decision Tree model achieved a slightly lower accuracy of 70.71% (Fatmawati et al., 2023). KNN also outperformed Decision Tree in terms of precision and recall with values of 72.65% and 71.67% respectively compared to the Decision Tree's precision and recall of 71.59% and 70.40%. Despite KNN's better performance, Decision Tree's interpretability and minimal computational power makes them more suitable for use in areas with limited resources. Decision Tree also serves as a baseline model against which other models can be compared and can be used as a starting point for building more complex models that uses multiple decision trees to improve accuracy.

# 3 Methodology

The approach taken in this study was guided by the Knowledge Discovery in Database (KDD) methodology useful in analysing complex data, and detecting patterns in different contexts (Waseem and Abidin, 2023). Figure 1 details the steps of this methodology. The performance of the evaluated models is evaluated using appropriate metrics, and the results are interpreted to gain insights into the factors.

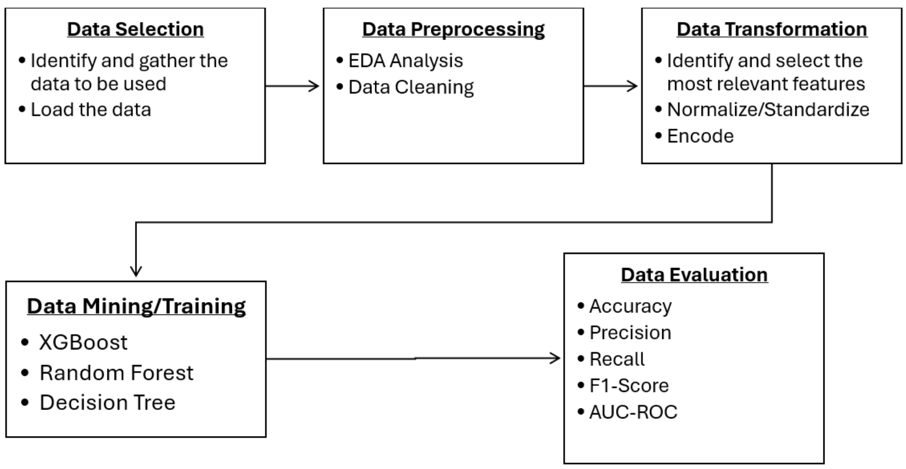


Figure 1: KDD Methodology Steps used in this study

## 3.1 Selection of Data

The dataset used for this research was a .CSV file sourced from the UCI Machine Learning Repository [2]. The dataset comprises of 1014 records and includes seven features as seen in table 1:

| **Attribute** | **Description** |
|---|---|
| Age | The age of the mother in years. |
| Systolic BP | The systolic blood pressure measurement in mmHg. |
| Diastolic BP | The diastolic blood pressure measurement in mmHg. |
| Blood Sugar | Blood sugar level in mmol/L. |
| Temperature | Body temperature in degrees Fahrenheit. |
| Heart Rate | The heart rate in beats per minute. |
| Risk Level | The target variable, categorized into three levels (Low, Moderate, High) indicating the risk of maternal health complications. |

Table 1: Description of Dataset Attributes

The data was collected and compiled from different hospital records in rural Bangladesh, including community clinics and maternal health cares, through a risk monitoring system. Each record represents an individual maternal health case with associated risk levels. The records include a range of physiological factors that are useful for this research.

## 3.2 Data Pre-processing

Though the full analysis is performed using Python, PowerBI was used for an initial exploratory analysis to visualize the distribution of risk levels and better understand the dataset before diving into the modelling phase.

[2] `https://archive.ics.uci.edu/dataset/863/maternal+health+risk`

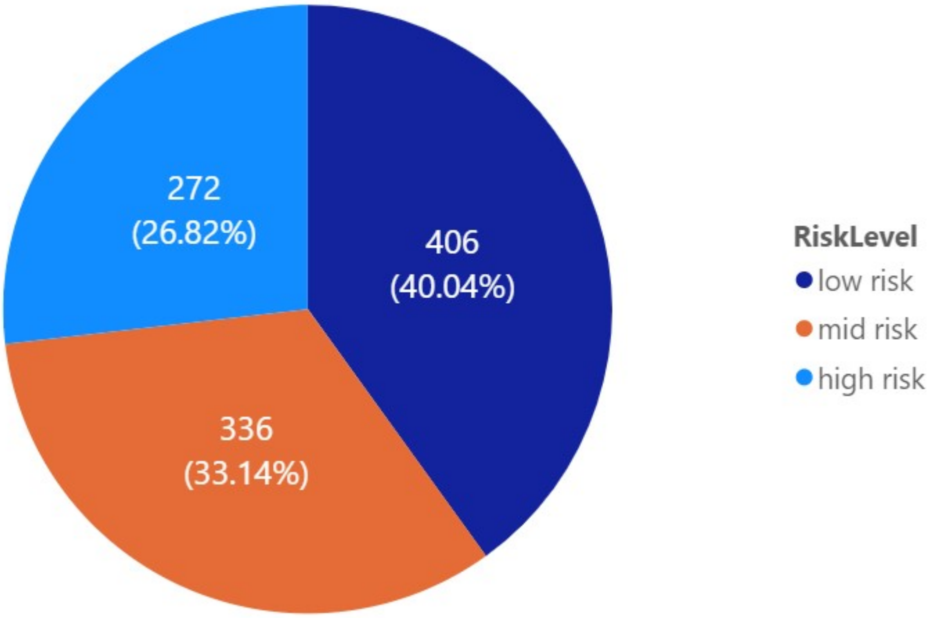


Figure 2: PowerBI Distribution of Risk Level

The initial step in the pre-processing stage done in Jupyter Notebook was to examine the basic information of the dataset in Python. This provided an overview of the dataset's structure, including the number of entries, column names, data types and non-null counts. Following this, an initial summary statistic table (table 2) was generated to understand the central tendencies and data distribution of the features.

| | **Age** | **SystolicBP** | **DiastolicBP** | **BS** | **BodyTemp** | **HeartRate** |
|---|---|---|---|---|---|---|
| **count** | 1014.000 | 1014.000 | 1014.000 | 1014.000 | 1014.000 | 1014.000 |
| **mean** | 29.872 | 113.198 | 76.461 | 8.726 | 98.665 | 74.302 |
| **std** | 13.474 | 18.404 | 13.886 | 3.294 | 1.371 | 8.089 |
| **min** | 10.000 | 70.000 | 49.000 | 6.000 | 98.000 | 7.000 |
| **25%** | 19.000 | 100.000 | 65.000 | 6.900 | 98.000 | 70.000 |
| **50%** | 26.000 | 120.000 | 80.000 | 7.500 | 98.000 | 76.000 |
| **75%** | 39.000 | 120.000 | 90.000 | 8.000 | 98.000 | 80.000 |
| **max** | 70.000 | 160.000 | 100.000 | 19.000 | 103.000 | 90.000 |

Table 2: Summary Statistics of Predictor Variables

Next, the dataset was checked for missing values. It was noted that there were no missing values which meant no removal of entries or imputation was needed. Outliers were identified and addressed, beginning with visual inspections through box plots for numerical features. Box plots are effective for spotting outliers, which are defined as values lying beyond 1.5 times the interquartile range (IQR) above the third quartile and below the first quartile. Despite tree-based models being less sensitive to outliers, handling them ensures cleaner data and potentially better model performance. The IQR test was applied, and the findings revealed that Age, Systolic BP, and Heart Rate contained outliers that were deemed to be valid information points rather than errors, hence they were untouched, while Blood Sugar and Body Temperature showed a very small range of values. Outliers in these columns were capped and removed. The body temperature column, in particular, was found to be irrelevant to the analysis and was therefore removed entirely from the dataset.

### 3.3 Feature Scaling

Feature scaling ensures that features contribute equally to the model's performance, preventing features with larger scaled from dominating the analysis. Various methods are used e.g. Normalization, which rescales the features to a fixed range typically from 0 to 1, and Standardization, which transforms the features to a mean of 0 and a standard deviation of 1. Tree-based models generally do not require feature scaling compared to linear models. As such, they do not need variables to be normalized, and can still produce good results (de Amorim et al., 2023).

### 3.4 Experimental Design

The experimental design was structured around the evaluation of three different ML models: XGBoost, Decision Tree, and Random Forest. Each model's performance was assessed using the same methodology for a fair comparison. Since ML models require numerical input, categorical labels in the target variable were converted into numerical forms using label encoding so that the models could perform the analysis appropriately.

### 3.5 Model Evaluation

Training the model helps in identifying patterns in the dataset, and observing relationships between the features and the target variable. If necessary, the model's parameters might need refinements to improve its predictive accuracy. During this phase, the dataset is split into two subsets (training and testing). The model is trained on the training set and makes predictions on new instances using the testing set. A validation subset might also be needed to fine tune the model and prevent over-fitting.

Model evaluation assesses the performance of the trained model using appropriate metrics such as accuracy, precision, recall, F1-score, and AUC-ROC which measures the performance quantitatively. These metrics help in understanding how well the model predicts outcomes and handles different classes in the dataset. Also, in this context, the evaluation of different models using the same metrics gives researchers the option of selecting the best model for any given task. Evaluation also helps in understanding the practicality of the model. A model that performs only well on training data might not be useful especially in the real-world. A table summarizing the accuracy, precision, recall, and F1-score for each model will be displayed in the coming sections. The interpretation of the results, as well as visualizations such as confusion matrices and ROC curves will also be shown.

## 4 Design Specification

The design specification outlines the techniques, architecture, and frameworks used in this study to understand maternal health risks using Decision Tree, Random Forest, and XGBoost models.

## 4.1 Model Techniques

Three machine learning models were evaluated: Decision Tree, Random Forest, and XGBoost. Each model was chosen for its potential advantages in handling healthcare data, particularly in the context of rural Bangladesh.

- Decision Tree: A simple, interpretable, and widely-used model that splits the data into subsets, forming a tree-like structure. Beyond classification and regression problems, Decision trees can also be used in clustering analysis, and can be used as a hybrid model (Blockeel et al., 2023). It was implemented using the 'scikit-learn' library, with hyper-parameter tuning to prevent over-fitting.
- Random Forest: An ensemble model that builds upon multiple decision trees to improve accuracy and reduce over-fitting. As a result, it works well with complex, high-dimensional data (Hu and Szymczak, 2023). It was also implemented using the 'scikit-learn' library with parameters such as the number of trees maximum features to optimize performance.
- XGBoost: A gradient boosting framework that is an enhanced version of a decision tree model. It builds the model sequentially, correcting errors of previous models to optimize performance. It also has the benefit of having far better accuracy than both random forest and decision tree models (Niazkar et al., 2024).

## 4.2 Data Preprocessing Techniques

To ensure reliability of the input data used for model training, a comprehensive preprocessing is required. The following steps were taken:

- Summary Statistics: The data structure was assessed to understand distributions and detect anomalies.
- Data Exploration: Verified that there were no missing values in the dataset. Box plots were generated to identify outliers. The IQR test was used to cap extreme values, particularly for blood sugar and body temperature, while valid outliers in age, systolic BP, and heart rate were retained.
- Feature Selection: Based on initial EDA, relevant features with significant predictive power were chosen. Principal Component Analysis (PCA) was also done to address multi-collinearity between variables.

## 4.3 Evaluation Techniques

- Accuracy: The ratio of correctly predicted instances to the total number of instances.
- Precision: The ratio of true positive predictions to the total positive predictions made by the model.
- Recall (Sensitivity): The ratio of true positive predictions to the actual number of positive instances in the data.
- F1-Score: The mean of precision and recall.

- Confusion Matric: Used to visualize the performance of the models, such as the true positive, true negative, false positive, and false negative rates.

# 5 Implementation

This section describes the final stage of the implementation of the proposed solution, including the tools and languages used, the processes involved in developing the models and transforming the data, and the outputs produced.

## 5.1 Tools and Languages Used

The primary programming language used for implementation was Python due to its extensive libraries, and Jupyter Notebook was used as the development environment. Key libraries used included 'pandas' for data manipulation and preprocessing tasks, 'scikit-learn' for model development, training, and evlauation, 'matplotlib' and 'seaborn' for data visualization.

## 5.2 Data Transformation

For the preprocessing outputs, the data was cleaned, transformed, and prepared for analysis. The key steps involved were:

- Initial Inspection: Using Python to inspect the data and check for basic information about the dataset.
- Missing Values: No missing values were found, confirmed using 'data.info()' and 'data.isnull().sum()'.
- Outlier Detection and Handling: Outliers in numerical features were detected using box plots and handled using the IQR method. Only the most extreme outliers are removed.
- Feature Removal: The 'BodyTemp' column was removed due to irrelevance.
- Final Dataset: The cleaned and transformed dataset was prepared for modeling, ensuring it was free from any inconsistencies and ready for analysis.

Next, the target variable was encoded numerically. A collinearity heatmap was created to check for collinearity between variables. Two variables (SystolicBP and DiastolicBP) were highly correlated with a value of 0.74. This was addressed by applying PCA, creating and adding a new principal component to the dataframe called 'PCA' which was a combination of the two variables.

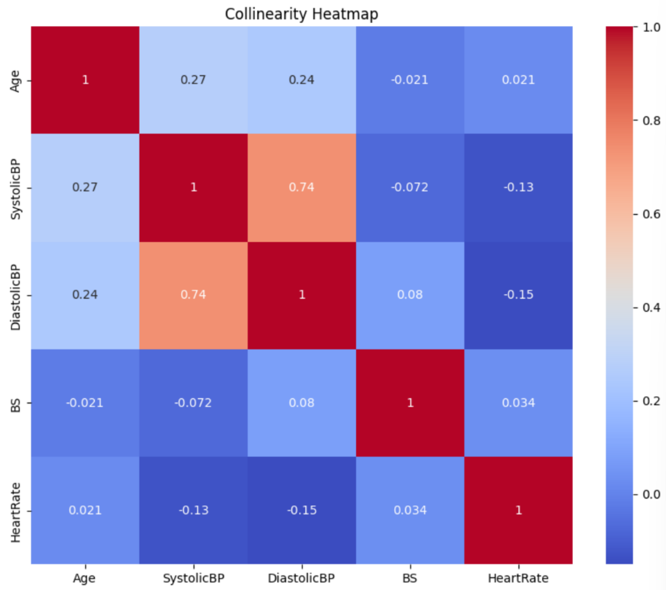


Figure 3: Heatmap before PCA

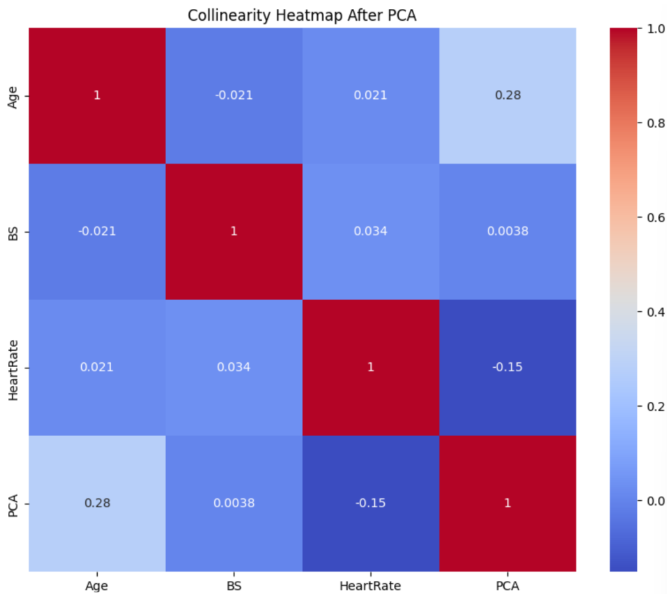


Figure 4: Heatmap after PCA

After, the dataset was split into training and testing sets with an 80/20 ratio using `train_test_split` from 'scikit-learn'. In building the model, the predictions and true labels were converted back to their original labels via inverse transform.

## 5.3 Model Implementation

Each model was initialized in their respective ways. To have better understanding of the precision, recall, f1-score, and accuracy of each model, a classification report was generated which also highlighted the significance of each class for each model. This was visually represented via a confusion matrix too which highlighted the number of true instances for both true and predicted labels. An ROC (receiver operating characteristic) curve showing the three model's performance with their true and false positive rates parameters was also plotted.

### 5.3.1 Implementing XGBoost

The XGBoost model was initialized with multiclass objective for classification by setting the objective to 'multi:softmax', which is designed for problems where the target variable can take on multiple discrete values (Shaheed et al., 2023). With its initial parameters, this classifier was fitted to the training data. Post-training, the model's prediction on the testing set was transformed back to their original labels as mentioned in 5.2. The model was then evaluated appropriately using the metrics specified in 5.3. To improve the performance, Hyperparameter tuning was done using RandomSearchCV to explore a parameter grid, including `n_estimators`, `max_depth`, `learning_rate`, and subsample. This random search helps identify the best combination of hyperparameters by evaluating the model's performance across different parameter settings. The RandomSearchCV was fitted to the training data, then evaluated on the test set using the same metrics as before.

### 5.3.2 Implementing Random Forest

For the Random Forest model, a RandomForestClassifier() was fitted to the training data. The same metrics stated in 5.3, including ROC curves, were used to observe the model's performance. For hyperparameter tuning, RandomizedSearchCV explored different range of parameters such as `n_estimators`, `max_depths`, `min_samples_split`, `min_samples_leaf` and bootstrap. Due to the large number of trees used, the process was extensive. After, the best set of hyperparameters were used to evaluate the model using the same metrics as before.

### 5.3.3 Implementing Decision Tree

For the Decision Tree model, a DecisionTreeClassifier was instantiated and fitted to the training data. Post-training, the model was evaluated by predicting on the test set. The same metrics stated in 5.3 were used to evaluate the model. Hyperparameter tuning was done using RandomizedSearchCV, which searches over a specified parameter grid to find the best combination of hyperparameters. Parameters such as `max_depth`, `min_samples_split`, and `min_samples_leaf` were tuned. The RandomizedSearchCV was fitted to the training data, and the best estimator was obtained. This best model was then evaluated on the test set using the same metrics as the initial model to ensure the model was optimized as much as possible.

# 6 Evaluation

The evaluation section provides analysis of the results obtained from implementing and testing three machine learning models. The study aims to compare these models in classifying and predicting maternal health risks. The primary objectives of this evaluation are to identify the most effective model, analyse the performance metrics, and discuss the implications of these findings from both an academic and practitioner perspective. For the context of this research, an effective predictive model is needed to improve maternal health care, particularly in areas where resources are limited. The models were developed and tested using a dataset that includes major health indicators. This evaluation not only assesses the performance and reliability of these models but also considers their practical applicability in real-world settings.

From figure 2, the chart shows that out of 1,014 records in the dataset, there were higher low-risk instances compared to the other levels. As the dataset contained an imbalance in class distribution, it was balanced using the Synthetic Minority Over-Sampling Technique (SMOTE), which generates synthetic samples in the minority classes to balance the distribution (Siti et al., 2024). Part of the discussion in this section is the comparative analysis of all three models, with and without SMOTE.

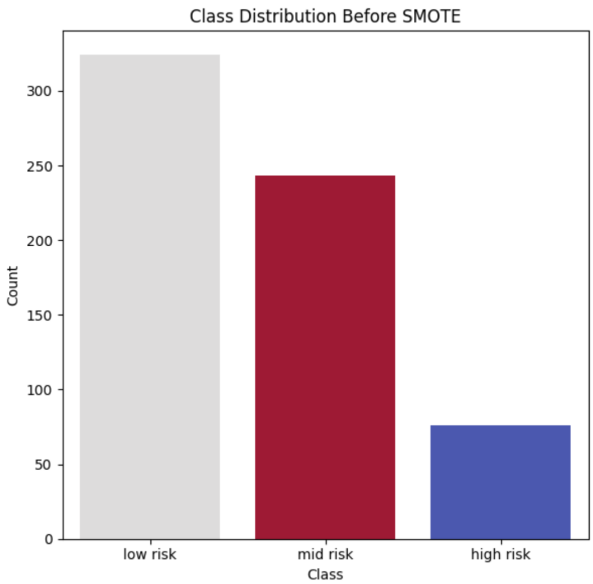


Figure 5: Class Distribution before SMOTE

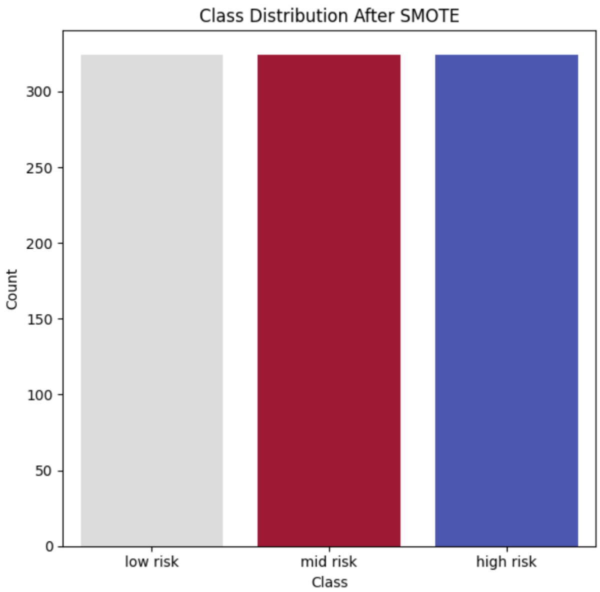


Figure 6: Class Distribution after SMOTE

## 6.1 Case Study 1: Comparative analysis with SMOTE

For the first case study, the focus is on the analysis of model performance after applying SMOTE. The objective is to evaluate the impact of SMOTE on the classification accuracy, precision, recall and F1-score of XGBoost, Random Forest, and Decision Tree, before and after hyperparameter optimization.

| Model | Accuracy | Precision | Recall | F1-Score |
|---|---|---|---|---|
| **Pre Hyperparameter-Tuning** | | | | |
| XGBoost | 0.82 | 0.84 | 0.83 | 0.83 |
| Random Forest | 0.80 | 0.81 | 0.81 | 0.81 |
| Decision Tree | 0.78 | 0.77 | 0.80 | 0.78 |
| **Post Hyperparameter-Tuning** | | | | |
| XGBoost | 0.81 | 0.82 | 0.82 | 0.82 |
| Random Forest | 0.74 | 0.73 | 0.74 | 0.73 |
| Decision Tree | 0.73 | 0.71 | 0.72 | 0.71 |

Table 3: Classification Report Summary [i] for Three Models

- XGBoost shows good performance with SMOTE applied. The high precision and recall values indicate that XGBoost handles the class imbalance well, reducing false positives and false negatives. After hyperparameter tuning, XGBoost shows a slight decrease in performance, which could be due to overfitting on the training data during tuning, leading to lower generalization performance. However, the model still maintains a strong overall performance.
- Random Forest shows good performance, with metrics slightly lower than XGBoost but still demonstrating its ability to handle the imbalanced data well. The balance between precision and recall indicates that Random Forest is effective in identifying both minority and majority class instances. There is a noticeable drop after hyperparameter tuning, suggesting it might not be optimal.
- Decision Tree has a moderate performance. The precision is lower, suggesting a higher rate of false positives, but with a relatively high recall it shows it can correctly identify minority classes. Post hyperparameter tuning, there is a significant decrease in performance the most compared to the other two models. The drop in precision, recall, and F1-score indicate a lower predictive power possibly due to suboptimal hyperparameter selection too.

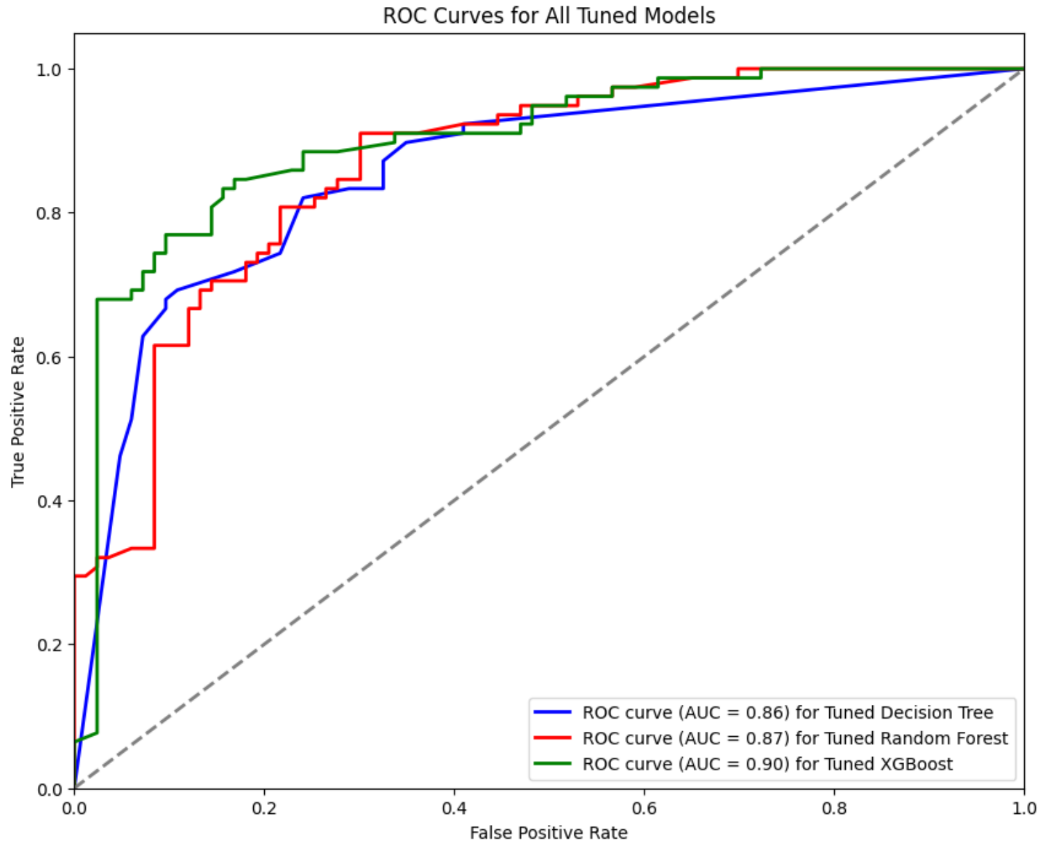


Figure 7: ROC Curve with SMOTE (Hyperparameter Tuning)

From table 4, a precision of 0.90 indicates that 90% of the instances labelled as high risk are correctly identified. The recall of 0.86 suggests that 86% of actual high risk instances are accurately captures. The F1-score of 0.88, which balances precision and recall, demonstrates that the model is highly effective in identifying high-risk cases with minimal false positives and false negatives.

| Class | Precision | Recall | F1-Score |
|---|---|---|---|
| Low-Risk | 0.83 | 0.83 | 0.83 |
| Mid-Risk | 0.77 | 0.79 | 0.78 |
| High-Risk | 0.90 | 0.86 | 0.88 |
| Macro Avg | 0.84 | 0.83 | 0.83 |
| Weighted Avg | 0.82 | 0.82 | 0.82 |
| **Overall Accuracy: 0.82** | | | |

Table 4: XGBoost Model Classification Report

For the low-risk category, the precision and recall are both 0.83, resulting in an F1-score of 0.83. This balance indicates that the model is equally good at identifying low-risk instances and minimizing false results. In practical applications, this reliability ensures that resources are not wasted on unnecessary interventions for low-risk cases.

The mid-risk category shows slightly lower performance with a precision of 0.77 and a recall of 0.79. The F1-score of 0.78 indicates that while the model is reasonably effective, there is room for improvement. This category often includes instances that may not be as clear-cut as high or low risk, but the model still performs well in maintaining a balance between precision and recall. The overall accuracy of 0.82 signifies that 82% of all instances are correctly classified. The macro and weighted averages of precision, recall, and F1-score provide a broader view of the model's performance across all classes. The macro average treats all classes equally, while the weighted average accounts for the class distribution, both indicating strong and balanced performance.

## 6.2 Case Study 2: Comparative analysis without SMOTE

This case study focuses on the analysis of model performance without the application of SMOTE. Evaluating the models on the imbalanced dataset provides insights into the ability to handle class imbalance and their overall performance.

| Model | Accuracy | Precision | Recall | F1-Score |
|---|---|---|---|---|
| **Pre Hyperparameter-Tuning** | | | | |
| XGBoost | 0.80 | 0.82 | 0.80 | 0.81 |
| Random Forest | 0.81 | 0.83 | 0.81 | 0.82 |
| Decision Tree | 0.78 | 0.79 | 0.79 | 0.79 |
| **Post Hyperparameter-Tuning** | | | | |
| XGBoost | 0.81 | 0.83 | 0.81 | 0.82 |
| Random Forest | 0.71 | 0.79 | 0.65 | 0.69 |
| Decision Tree | 0.68 | 0.67 | 0.67 | 0.67 |

Table 5: Classification Report Summary [ii] for Three Models (Before and After SMOTE)

- With an F1-score of 0.81, there was a good balance between precision and recall for the XGBoost model, suggesting it was able to identify positive instances while maintaining a low false positive rate. After hyperparameter tuning, there was improvement in all metrics. Fine tuning the model improved its generalization capabilities.
- Random Forest slightly outperformed XGBoost. It also had fewer false positives. Post-tuning, the performance dropped significantly, with a recall value of 0.65. There were higher false positives and false negatives captured in the model, reducing its precision by 10%.
- The Decision Tree was the most inferior model with an accuracy of 78%. Its F1-score reflects a slightly less balanced performance potentially due to its susceptibility to overfitting. It also declined in performance after tuning, failing to capture complex patterns in the data especially without SMOTE to address class imbalance.

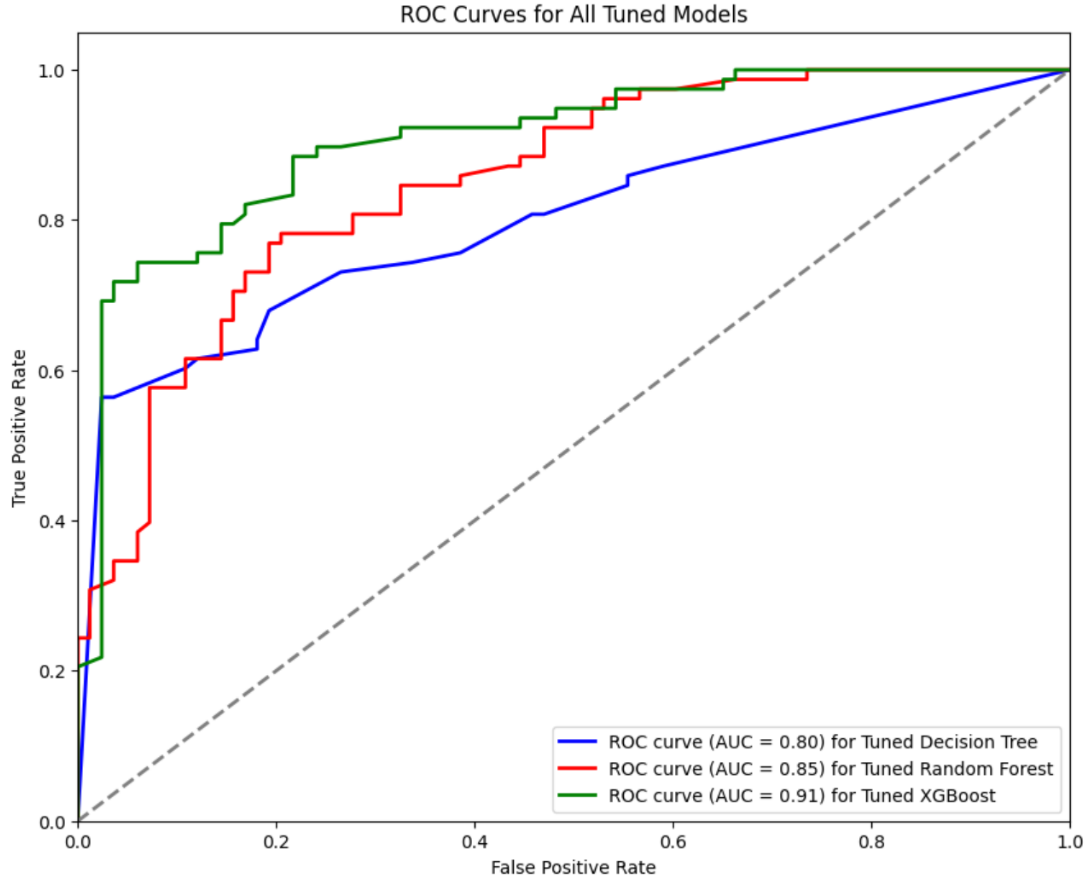


Figure 8: ROC Curve without SMOTE (Hyperparameter Tuning)

Though hyperparameter tuning helps in improving the model's predictive capability by optimizing the necessary parameters, it can still lead to overfitting if not done carefully. While it searches for the best possible combinations across all its parameter values, the outcome it provides might not be the most ideal hyperparameter combination (Christioko et al., 2023). This might have been the case for both the random forest and decision trees models, especially with the depth and number of trees, leading to poor performance on the testing sets and increasing the complexity of the models beyond their optimal level.

| Class | Precision | Recall | F1-Score |
|---|---|---|---|
| Low-Risk | 0.84 | 0.81 | 0.82 |
| Mid-Risk | 0.74 | 0.80 | 0.77 |
| High-Risk | 0.90 | 0.82 | 0.86 |
| Macro Avg | 0.83 | 0.81 | 0.82 |
| Weighted Avg | 0.81 | 0.81 | 0.81 |
| **Overall Accuracy: 0.81** | | | |

Table 6: XGBoost Model Classification Report (without SMOTE)

## 6.3 Case Study 3 - Analysis in underlying issues related to high risk pregnancies

XGBoost proved to be the best performing model in this study, with and without the application of SMOTE. Without SMOTE, XGBoost achieved an accuracy of 80%, precision of 82%, recall of 80%, and F1-score of 81%. After hyperparameter tuning, these metrics slightly improved, showcasing the model's ability to generalize well across different datasets. When SMOTE was applied, the performance metrics improved, with the model achieving an accuracy of 82%, precision of 84%, recall of 83%, and F1-score of 83%. This improvement highlights the importance of addressing class imbalance in healthcare data, where the minority class might represent cases that require accurate identification. Within the context of healthcare, the precision and recall trade-off is important. High precision ensures that fewer healthy individuals are incorrectly diagnosed as high-risk, minimizing unnecessary stress and medical demands. High recall guarantees that most at-risk individuals are correctly identified, which is essential for urgent medical treatment. In a scenario where false positives and false negatives might have serious consequences, XGBoost's performance makes it a suitable model for healthcare application.

XGBoost also showed stability and was reliable even after refining its parameters, which are important factors in real-world cases where data variability is common, justifying the use of ML models in such a domain (2.4.1). This can lead to better patient outcomes, an efficient allocation of resources, and overall improvements in the quality of care. The identification of contributing factors was also a key part of this analysis. For that, permutation importance analysis was done to measure the contribution of each feature to the model's predictive performance when shuffled, and how influential they were in determining high-risk pregnancies. In the context of high-risk pregnancies, these insights justifies the need for interventions.

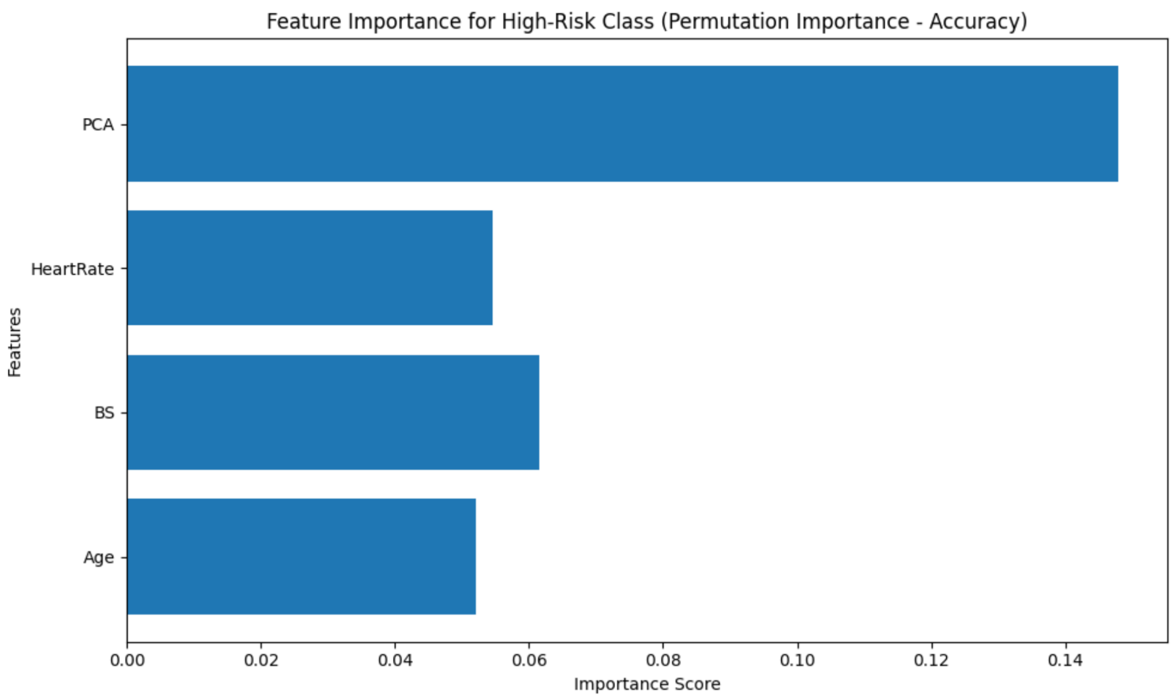


Figure 9: Feature Importance Plot - High Risk

In the permutation analysis, the Principal Component Analysis (PCA) feature had the highest importance score for the high-risk class, with an importance of approximately 0.15, suggesting that the combined effects of systolic and diastolic blood pressure are critical in predicting high-risk pregnancies. Medical knowledge (2.2.2) confirms this finding. Despite its importance, other factors such as heart-rate, blood sugar, and age, had relatively lower importance scores. However, their contribution is still relevant as these features play a role in assessing overall maternal risk, particularly when considered together. In the real world, this analysis can inform decision-making in clinical centres by highlighting specific indicators that should be closely monitored in pregnant women. For example, prioritizing blood pressure control and looking for fluctuations can directly address one of the most impactful predictors of high-risk pregnancies. Recognizing that heart rate, blood sugar, and age—though lower in importance—are still relevant, healthcare providers can adopt a more holistic approach, addressing multiple risk factors to reduce maternal mortality.

## 6.4 Analysis of results using Cross-Validation

The initial analysis was dependent on a single train-test split and a single training run, which can sometimes lead to misleading estimates of model performance. To ensure that the performance of the model is not overly reliant on any particular split of the data, cross-validation was applied. This technique is useful for evaluating the generalizability of the model by testing it across multiple subsets of the data, providing a more reliable estimate of how the model would perform on unseen data In this case, 5-fold cross-validation was applied, dividing the training data into five equal parts (folds). In each iteration, four of the folds were used for training the model, while the remaining fold was set aside as a validation set to test the model's performance. This process was repeated five times, with each fold serving as the validation set exactly once. By averaging the accuracy scores from each of these five iterations, the final cross-validation accuracy score was obtained.

The importance of this step not only increases the validity of the model, but also balances the bias-variance trade-off by preventing overfitting and underfitting. Cross-validation also offers a more reliable measure of model performance, compared to a single train-test split, where the result may be heavily influenced by how the data happens to be split. For this analysis, the cross-validation results for three tuned models—XGBoost, Random Forest, and Decision Tree—were calculated.

| **Model** | **Train-Test Split Accuracy (Without SMOTE/With SMOTE)** | **Cross-Validation Accuracy** |
|---|---|---|
| XGBoost | 0.81 / 0.82 | 0.85 |
| Random Forest | 0.71 / 0.75 | 0.78 |
| Decision Tree | 0.68 / 0.73 | 0.77 |

Table 7: Model Performance Summary with Cross-Validation

# 7 Conclusion and Future Work

Future work on this project should primarily aim at improving the practicality of the models. Firstly, exploring alternative algorithms such as Neural Networks or other ensemble methods could provide improvements in predictive accuracy. Other resampling techniques beyond SMOTE, such as ADASYN, could be implemented to better address class imbalance. A deeper investigation into feature interactions via Partial Dependence Plots (PDPs) can provide clearer insights into how different feature interacts and their potential impact for each risk class. On the practical side, based on clinical expertise, collaborating with healthcare professionals might be useful to validate the model's predictions and refine ts features if needed. This also includes building real-time risk assessment techniques to integrate these models into healthcare systems. Complex methods like an autonomous model which learns from new data, ensuring it remains accurate over time, might be a good idea. There is also the potential for commercialization, as such a tool can be marketed to hospitals and clinics worldwide. Regulatory approvals can be secured to give more credibility and encourage a wider adoption of the tool beyond even rural areas.

# References


Acherjya, G. K., Tarafder, K., Dutta, D., Mortuza, M. G., Sarkar, A. K., Das, N. L. and Ali, M. (2023). Frequency and risk factors stratification of hypertension among the rural population of bangladesh, *Journal of Family Medicine and Primary Care* **12**(10): 2488–2495.

Ahmad, M., Sechi, C. and Vismara, L. (2024). Advanced maternal age: A scoping review about the psychological impact on mothers, infants, and their relationship, *Behavioral Sciences* **14**(3): 147.

Alamsyah, D. P., Ramdhani, Y. and Susanti, L. (2023). Maternal health risk classification: Random forest and evolutionary algorithms, *2023 10th International Conference on ICT for Smart Society (ICISS)*, pp. 1–6.

Alowais, S. A., Alghamdi, S. S., Alsuhebany, N., Alqahtani, T., Alshaya, A. I., Almohareb, S. N., Aldairem, A., Alrashed, M., Bin Saleh, K., Badreldin, H. A. et al. (2023). Revolutionizing healthcare: the role of artificial intelligence in clinical practice, *BMC medical education* **23**(1): 689.

Aquino, Y. S. J., Carter, S. M., Houssami, N., Braunack-Mayer, A., Win, K. T., Degeling, C., Wang, L. and Rogers, W. A. (2023). Practical, epistemic and normative implications of algorithmic bias in healthcare artificial intelligence: a qualitative study of multidisciplinary expert perspectives, *Journal of Medical Ethics* .

Ardeti, V. A., Kolluru, V. R., Varghese, G. T. and Patjoshi, R. K. (2023). An overview on state-of-the-art electrocardiogram signal processing methods: Traditional to ai-based approaches, *Expert Systems with Applications* **217**: 119561.

Bester, M., Almario Escorcia, M., Fonseca, P., Mollura, M., van Gilst, M., Barbieri, R., Mischi, M., van Laar, J., Vullings, R. and Joshi, R. (2023). The impact of healthy pregnancy on features of heart rate variability and pulse wave morphology derived from wrist-worn photoplethysmography, *Scientific Reports* **13**(1): 21100.

Blockeel, H., Devos, L., Frénay, B., Nanfack, G. and Nijssen, S. (2023). Decision trees: from efficient prediction to responsible ai, *Frontiers in Artificial Intelligence* **6**: 1124553.

Chen, J., Wen, L., Fu, G., Bai, C., Lei, X. and Zhang, Y. (2024). The relationship between health literacy and blood sugar control in rural areas among diabetes patients, *Frontiers in Endocrinology* **15**: 1334100.

Chowdhury, A. H., Hanifi, S. M. A., Iqbal, M., Hossain, A., Stones, W., Amos, M., Pallikadavath, S., Bhuiya, A. and Mahmood, S. S. (2023). Does maternal health voucher scheme have association with distance inequality in maternal and newborn care utilization? evidence from rural bangladesh, *Plos one* **18**(12): e0295306.

Christioko, B. V., Khoirudin and Daru, A. F. (2023). A best exponential smoothing method with hyperparameter tuning to predict the number of pandemic cases, *2023 International Conference on Technology, Engineering, and Computing Applications (ICTECA)*, pp. 1–6.

D. Jones, R., Allison, M. K., Moody, H., Peng, C. and Eswaran, H. (2023). Use of cellular-enabled remote patient monitoring device for hypertension management in pregnant women: A feasibility study, *Maternal and Child Health Journal* **27**(7): 1191–1198.

de Amorim, L. B., Cavalcanti, G. D. and Cruz, R. M. (2023). The choice of scaling technique matters for classification performance, *Applied Soft Computing* **133**: 109924.

Doctor, J. and MacEwan, J. P. (2017). Limitations of traditional health technology assessment methods and implications for the evaluation of novel therapies, *Current Medical Research and Opinion* **33**(9): 1635–1642.

Fatmawati, M., Novita, M., Saputro, N. D., Chauhan, A. S., Nada, N. Q., Sinha, A., Herlambang, B. A. and Gupta, S. (2023). Maternal health risk classification - a comparison between algorithm decision tree and k-nearest neighbor (knn), *2023 International Conference on Sustainable Emerging Innovations in Engineering and Technology (ICSEIET)*, pp. 507–511.

Hababa, H. and Assarag, B. (2023). Overall maternal morbidity during pregnancy using the who-woice tools, *Plos one* **18**(8): e0275882.

Hallowell, N., Badger, S., McKay, F., Kerasidou, A. and Nellåker, C. (2023). Democratising or disrupting diagnosis? ethical issues raised by the use of ai tools for rare disease diagnosis, *SSM-Qualitative Research in Health* **3**: 100240.

Haq, A. U., Li, J. P., Kumar, R., Ali, Z., Khan, I., Uddin, M. I. and Agbley, B. L. Y. (2023). Mcnn: a multi-level cnn model for the classification of brain tumors in iot-healthcare system, *Journal of Ambient Intelligence and Humanized Computing* **14**(5): 4695–4706.

Hossain, M. I., Rahman, T., Sadia, T. S., Saleheen, A. A. S., Sarkar, S., Khan, M., Ohi, T. F. and Haq, I. (2024). Survival analysis of early intention of antenatal care among women in bangladesh, *Scientific Reports* **14**(1): 4738.

Hou, X., Du, D., Wang, Y., Song, Y. and Qian, J. (2023). Decision tree model-based flight phase classification method, *2023 International Conference on Computer Applications Technology (CCAT)*, pp. 13–16.

Hu, J. and Szymczak, S. (2023). A review on longitudinal data analysis with random forest, *Briefings in Bioinformatics* **24**(2): bbad002.

Islam, A., Begum, F., Williams, A., Basri, R., Ara, R. and Anderson, R. (2023). Midwife-led pandemic telemedicine services for maternal health and gender-based violence screening in bangladesh: an implementation research case study, *Reproductive health* **20**(1): 128.

Khan, M., Gupta, G., Shervani, K. I. et al. (2024). Predicting heartbeats in real-time: A continuous monitoring approach with face recognition algorithm, *2024 11th International Conference on Signal Processing and Integrated Networks (SPIN)*, IEEE, pp. 399–402.

Khanna, A., Selvaraj, P., Gupta, D., Sheikh, T. H., Pareek, P. K. and Shankar, V. (2023). Internet of things and deep learning enabled healthcare disease diagnosis using biomedical electrocardiogram signals, *Expert Systems* **40**(4): e12864.

Londero, A. P., Bertozzi, S., Xholli, A., Cedolini, C. and Cagnacci, A. (2024). Breast cancer and the steadily increasing maternal age: are they colliding?, *BMC Women's Health* **24**(1): 286.

Maheswari, B. U., Dixit, A. and Karn, A. K. (2024). Machine learning algorithm for maternal health risk classification with smote and explainable ai, *2024 IEEE 9th International Conference for Convergence in Technology (I2CT)*, pp. 1–6.

Mihali, S.-I. and Niță, -L. (2024). Credit card fraud detection based on random forest model, *2024 International Conference on Development and Application Systems (DAS)*, IEEE, pp. 111–114.

Min, J., Lee, W., Oh, J., Kwag, Y., Kim, E., Kim, J. M., Lee, K. A. and Ha, E. (2024). Disparities in the association between ambient temperature and preterm birth according to individual and regional characteristics: a nationwide time-stratified case-crossover study, *Environmental Health* **23**(1): 23.

Mndala, L., Kondoni, C., Gadama, L., Bamuya, C., Kuyere, A., Maseko, B., Kachale, F., Gondwe, M. J., Lissauer, D. and Nyondo-Mipando, A. L. (2024). Developing comprehensive woman hand-held case notes to improve quality of antenatal care in low-income settings: participatory approach with maternal health stakeholders in malawi, *BMC Health Services Research* **24**(1): 628.

Niazkar, M., Menapace, A., Brentan, B., Piraei, R., Jimenez, D., Dhawan, P. and Righetti, M. (2024). Applications of xgboost in water resources engineering: A systematic literature review (dec 2018–may 2023), *Environmental Modelling & Software* p. 105971.

Price, J., Yamazaki, T., Fujihara, K. and Sone, H. (2022). Xgboost: Interpretable machine learning approach in medicine, *2022 5th World Symposium on Communication Engineering (WSCE)*, pp. 109–113.

Reza, S. E., Islam, S. A. and Al Mamun, K. A. (2024). A feasibility study on the rooftop solar net metering system for union level health care facilities in bangladesh, *2024 7th International Conference on Development in Renewable Energy Technology (ICDRET)*, IEEE, pp. 1–6.

Rosales, M., Huacacolque, E., Castillo-Sequera, J. L. and Wong, L. (2024). Framework for monitoring peruvian patients with hypertension using a smartwatch and gpt, *2024 35th Conference of Open Innovations Association (FRUCT)*, pp. 588–595.

Rumbeli, N. M., August, F., Silvestri, V. and Sirili, N. (2024). Factors influencing maternal death surveillance and review implementation in dodoma city, tanzania. a qualitative case study, *Learning Health Systems* **8**(2): e10390.

Semeia, L., Bauer, I., Sippel, K., Hartkopf, J., Schaal, N. K. and Preissl, H. (2023). Impact of maternal emotional state during pregnancy on fetal heart rate variability, *Comprehensive Psychoneuroendocrinology* **14**: 100181.

Setiawan, S. D., Zanetta, M., Chandra Kesuma, K. A., Edbert, I. S. and Suhartono, D. (2023). Predicting diabetes risk in pregnant women: Leveraging adaboost as a boosting algorithm, *2023 International Conference on Advanced Mechatronics, Intelligent Manufacture and Industrial Automation (ICAMIMIA)*, pp. 1–6.

Shaheed, K., Abbas, Q., Hussain, A. and Qureshi, I. (2023). Optimized xception learning model and xgboost classifier for detection of multiclass chest disease from x-ray images, *Diagnostics* **13**(15): 2583.

Siti, H., Abou Elassad, Z. E., El Meslouhi, O., Abou Elassad, D. E., Lamjadli, S. and Lakhel, H. (2024). Improving cardiovascular disease prediction: An optimized tree-based strategy using smote, *2024 4th International Conference on Innovative Research in Applied Science, Engineering and Technology (IRASET)*, IEEE, pp. 1–6.

Strahm, A. M., Hilmert, C. J., Campos, B., Dvorak, R. and Schenker, M. (2024). Maternal blood pressure and birth weight associations in us-born and foreign-born latinas., *Cultural Diversity and Ethnic Minority Psychology* .

Wan, W., Zhu, Y., Tian, J., Cheng, Y., Zeng, L. and Zhu, Z. (2024). Associations of parental age at pregnancy with adolescent cognitive development and emotional and behavioural problems: a birth cohort in rural western china, *BMC Public Health* **24**(1): 775.

Waseem, M. and Abidin, S. (2023). Issues and challenges of kdd model for distributed data mining techniques and architecture, *2023 10th International Conference on Computing for Sustainable Global Development (INDIACom)*, pp. 1612–1616.

Wierzchowska-Opoka, M., Grunwald, A., Rekowska, A. K., Łomża, A., Mekler, J., Santiago, M., Kabała, Z., Kimber-Trojnar, Ż. and Leszczyńska-Gorzelak, B. (2023). Impact of obesity and diabetes in pregnant women on their immunity and vaccination, *Vaccines* **11**(7): 1247.

Wu, Y., Yuan, J., Yuan, Y., Kong, C., Jing, W., Liu, J., Ye, H. and Liu, M. (2023). Effects of ambient temperature and relative humidity on preterm birth during early pregnancy and before parturition in china from 2010 to 2018: a population-based large-sample cohort study, *Frontiers in Public Health* **11**: 1101283.