MACHINE LEARNING FOR ENHANCING EARLY MATERNAL AND FETAL HEALTH CARE - AN INTELLIGENT RISK ASSESSMENT FRAMEWORK

A Thesis
Submitted in Partial Fulfilment of Requirements
For the Award of the Degree

MASTER OF TECHNOLOGY
in
Data Science

Submitted by
Maitree
(23/DSC/17)

Under the supervision of
Dr. Ruchika Malhotra
Professor
Department of Software Engineering

DEPARTMENT OF SOFTWARE ENGINEERING
DELHI TECHNOLOGICAL UNIVERSITY
(Formerly Delhi College of Engineering)
Bawana Road, Delhi – 110042

May, 2025

ACKNOWLEDGEMENT

I am grateful to Dr. Ruchika Malhotra (Professor, HOD, Department of Software Engineering) and all of the Department of Software Engineering faculty members at DTU. They all gave us a lot of help and advice for the thesis.

I'd also want to thank the University for providing us with the laboratories, infrastructure, testing facilities, and environment that allowed us to continue working without interruption.

I'd also like to thank our lab assistants, seniors, and peer group for their aid and knowledge on a variety of subjects.

Maitree

23/DSC/17

DEPARTMENT OF SOFTWARE ENGINEERING
DELHI TECHNOLOGICAL UNIVERSITY
(Formerly Delhi College of Engineering)
Bawana Road, Delhi – 110042

DECLARATION

I, Maitree, Roll No.: 23/DSC/17, student of M.Tech (Data Science), hereby certify that the work which is being presented in the thesis entitled "**Machine Learning for Enhancing Early Maternal and Fetal Health Care - An Intelligent Risk Assessment Framework**" in partial fulfilment of the requirements for the award of degree of Master of Technology, submitted in the Department of Software Engineering, Delhi Technological University is an authentic record of my own work carried out during the period from Jan 2025 to May 2025 under the supervision of Dr. Ruchika Malhotra.

The matter presented in the thesis has not been submitted by me for the award of any other degree of this or any other institute.

Candidate's Signature

This is to certify that the student has incorporated all the corrections suggested by the examiners in the thesis and the statement made by the candidate is correct to the best of our knowledge.

Signature of Supervisor

DEPARTMENT OF SOFTWARE ENGINEERING
DELHI TECHNOLOGICAL UNIVERSITY
(Formerly Delhi College of Engineering)
Bawana Road, Delhi – 110042

CERTIFICATE

This is to confirm that Maitree (23/DSC/17) completed the project "**Machine Learning for Enhancing Early Maternal and Fetal Health Care - An Intelligent Risk Assessment Framework**" under my guidance in partial fulfilment of the MASTER OF TECHNOLOGY degree in DATA SCIENCE at DELHI TECHNOLOGICAL UNIVERSITY, NEW DELHI. To the best of my knowledge this work has not been submitted in part or full for any other Degree to this University or elsewhere.

Dr. Ruchika Malhotra

Professor & HOD

Department of Software Engineering

ABSTRACT

Global public health still places great emphasis on maternal and fetal health care, particularly in underdeveloped nations where access to timely and high-quality prenatal services is sometimes constrained. Early detection of pregnancy-related hazards remains difficult despite developments in medical science because of a reliance on subjective clinical judgment and static threshold-based evaluations. Using machine learning (ML) technologies, this thesis offers an intelligent, data-driven framework for the prediction and stratification of maternal and fetal health hazards. By means of early risk identification, the proposed system seeks to move from conventional reactive care models to proactive, tailored interventions.

We examined a systematic maternal health database of 6,103 clinical records. Included were essential physiological and biochemical markers including systolic blood pressure, heart rate, glucose levels, HbA1c, body temperature, and body mass index. Fifteen ML models were run and assessed using accuracy, F1-score and AUC measures after thorough preprocessing including outlier management, multicollinearity reduction and feature scaling. With CatBoost reaching an accuracy of 98.61% and showing great interpretability using SHAP (SHapley Additive Explanations), ensemble models like CatBoost, XGBoost and LightGBM outperformed baseline classifiers.

Addressing a significant drawback of current binary classification systems, the system classifies pregnancy risk into three categories: low, medium, and high. Furthermore, a fetal health classification module was created from CTG (cardiotocogram) data, so allowing complete prenatal evaluation. Both models were included into a Streamlit-based web application, therefore offering medical professionals a simple interface for real-time risk prediction and visual explanation of findings.

By means of a scalable, interpretable and accessible clinical decision support tool, this thesis not only confirms the efficacy of ensemble ML models in maternal and fetal risk prediction but also stresses deployment readiness. Particularly in under-resourced areas, the solution is set to help doctors make educated, data-backed decisions that could greatly enhance maternal and fetal outcomes.

CONTENTS

LIST OF FIGURES

LIST OF TABLES

CHAPTER 1

INTRODUCTION

1.1 BACKGROUND AND MOTIVATION

Still a significant component of world public health, maternal and fetal health care is particularly crucial in underdeveloped nations where access to healthcare, resources and infrastructure is sometimes constrained. Those of them in low- and middle-income nations, the World Health Organization (WHO) estimates that in 2020 about 295,000 women died during and after pregnancy and childbirth. Prompt medical treatment and sufficient prenatal care can help to avoid most of these deaths. The absence of early risk identification and methodical follow-up is particularly important for maternal death and morbidity in cases of gestational hypertension, diabetes and preterm labor. Often depending on thresholds, conventional clinical methods could overlook subtle health issues, particularly in diverse and complex populations.

Increasing computer power and the expanding digitization of medical records are generating more opportunities to use machine learning (ML) to change maternal and fetal health care. Patient data might contain hidden patterns and correlations that conventional analysis would cause machine learning algorithms to miss. Using historical health data, these models can predict possible problems with remarkable accuracy, therefore enabling early interventions that save lives and reduce the burden on healthcare providers.

This paper aims to change from reactive to proactive healthcare policies. By way of an intelligent risk assessment framework employing ML technologies, this work aims to provide healthcare professionals tools enhancing their diagnostic accuracy and decision-making. Apart from improving individual patient outcomes, such systems can direct policy decisions and optimize resource allocation at the systematic level.

Moreover, more thorough intervention plans result from the addition of multiclass risk stratification—low, medium and high risk—rather than conventional binary classification. Public health programs aimed at vulnerable populations in rural and poor areas will find this change especially important as early identification and resource prioritization can greatly affect maternal and infant death rates.

1.2 PROBLEM STATEMENT

Though obstetric treatment has advanced considerably, predicting and controlling pregnancy-related problems still presents a major difficulty. Traditional diagnostic techniques depend on doctor knowledge and pre-defined clinical criteria, such as systolic blood pressure over 140 mmHg or fasting glucose levels above a specified level, to decide whether a patient is at risk. Although helpful, these ongoing measurements sometimes miss the intricate interplay of chemical and physiological markers affecting maternal and fetal health. Furthermore, such tests tend to highlight binary results—classifying patients as either "at risk" or "not at risk," therefore limiting the possibility for nuanced clinical decision-making.

Actual world settings see pregnancy risks running on a continuum rather than in defined categories. Although she may not be high risk, a woman could still be at moderate risk and benefit from preventive care or closer monitoring. Especially for people in the "grey zone," the lack of intermediate categories in most current systems could lead to missed opportunities for early intervention. Low-resource healthcare systems also experience staff shortages and inconsistent follow-up care, which compounds the problem of unreported or mismanaged cases.

Data quality, data diversity and patient demographics all add to the issue by their difference. Often, medical datasets have missing values, outliers and non-standardized formats that challenge traditional analytical techniques. Though some studies have applied machine learning techniques to maternal health, most are restricted to binary classification tasks or they minimize the importance of model interpretability, a main criterion for application in clinical environments.

An intelligent, automated system capable of precisely classifying pregnancy risk into several categories, handling complex, heterogeneous clinical data and providing obvious, interpretable results is therefore absolutely necessary. Such a system should not only improve diagnostic accuracy but also build confidence among medical professionals by offering evidence-based justifications for its predictions. This study addresses this need by way of a machine learning-based framework for multiclass pregnancy risk prediction and assessment.

1.3 OBJECTIVES OF THE STUDY

The main goal of this paper is to create and test an intelligent risk assessment system using machine learning algorithms to improve early maternal and fetal health care. Key demographic, physiological and biochemical indicators drive the system's prediction of the probability of pregnancy-related complications. The study intends to meet the following specific objectives in order to achieve this general goal:

- Aiming to allow tiered intervention strategies in line with WHO prenatal care recommendations, develop a multiclass classification tool to classify pregnancy risks as low, medium or high.
- Train and validate the model using real clinical data. Comprising structured medical records with characteristics including blood pressure, body mass index (BMI), glucose levels, age, body temperature and heart rate, the dataset comes from over 6,000 Indian patients.
- Ensure data quality and integrity by applying feature scaling, outlier detection and multicollinearity reduction using Variance Inflation Factor (VIF) among other data processing methods.
- Compare various machine learning models, including traditional algorithms (e.g., logistic regression, SVM), tree-based models (e.g., decision trees, random forests) and advanced ensemble techniques (e.g., Gradient Boosting, LightGBM, CatBoost).
- By means of hyperparameter tuning using randomized search and cross-validation, optimize models to improve performance and generalizability across several subsets of patient data.
- Assess model performance using robust statistical techniques including accuracy, F1-score and AUC-ROC with specific focus on interpretability and clinical relevance.
- Build an interactive user interface to show practical application and usability of the risk prediction system in real-world healthcare environments using tools like Streamlit.

The study intends to provide a deployable, accurate and interpretable decision support system by means of these goals that can be used in low-resource healthcare settings to enhance maternal outcomes and save lives.

1.4 SCOPE OF THE WORK

The scope of this work is defined by the design, development and validation of a machine learning-based system for predicting pregnancy-related risks among expectant mothers using clinical data. The work mostly stresses structured datasets composed of routinely gathered demographic, physiological and biochemical variables. Traits like age, HbA1c, glucose levels, temperature, heart rate, body mass index (BMI), systolic and diastolic blood pressure are included. The project spans the whole pipeline of a data-driven machine learning solution from data collecting and cleaning to model training, assessment and deployment.

The study confines itself to classifying maternal health risk into three categories: low, medium and high. Though fetal health is addressed in the larger context in the project report, this version does not try to forecast fetal health results straight from cardiotocography (CTG) data. The system is meant to function as a clinical decision support tool, not a replacement for medical knowledge. It aims to enhance clinical decision-making by providing evidence-backed analysis of patient risk profiles.

Model creation and testing are driven by a single dataset comprising over 6,100 patients. The dataset is restricted to prenatal records within India and requires particular changes for application in other geographic or demographic locations. Although the system is evaluated in an offline simulation environment rather than under real-time clinical environment, its modular design allows it to fit into electronic health records (EHR) systems.

Moreover, this study lacks longitudinal tracking, wearable devices or real-time sensor data, all of which could improve risk prediction by means of time-series analysis. Though, the present system offers a good basis for such extensions. A basic web-based interface showing the viability of including machine learning models into available, user-friendly systems makes up the practical part of the study.

CHAPTER 2

LITERATURE REVIEW

2.1 OVERVIEW OF MATERNAL AND FETAL HEALTH CARE CHALLENGES

Maternal and fetal health are still significant global public health concerns in low- and middle-income countries. Nearly 295,000 women die each year from complications during pregnancy and childbirth, according to the World Health Organization (WHO), most of them in places without access to quality health services [1]. Though maternal death has fallen over past decades, it is still a significant problem in India caused by insufficient antenatal care, high blood pressure, diabetes, infections and bad nutritional status [2][15].

Reducing problems that could cause maternal or fetal morbidity and death requires good pregnancy monitoring. Often, traditional maternal care evaluates risk using regular check-ups and clinical criteria. But subjective interpretation, data fragmentation and lack of predictive knowledge limit these manual, reactive strategies [3]. Furthermore, although urban hospitals might provide fairly sophisticated maternal care, rural areas are usually neglected and lack suitable diagnostic equipment or qualified staff members [8].

Technological solutions that can deliver scalable, reasonably priced and proactive healthcare are urgently needed. Machine learning could help to close this gap by using historical clinical data to forecast pregnancy-related concerns before they develop into crises [14][17]. Early warning systems combined with digital health platforms can provide considerable assistance in resource-limited environments where prompt interventions could prevent major complications or deaths

2.2 EXISTING RISK PREDICTION SYSTEMS

In recent years, several computational models have been created to evaluate maternal and fetal health concerns. Traditional models flag at-risk patients using rule-based decision trees employing clinical cut-offs like systolic blood pressure or glucose thresholds. While simple, these models lack adaptability to individual variability and often fail to detect subtle but important changes in patient health status [2][15].

Machine learning has become a powerful tool in predictive maternal health as wearable health monitoring and electronic health records (EHRs) have grown. Often seen on maternal datasets, models including Decision Trees, Random Forests and Gradient Boosting Machines classify pregnancy risk [4][5][14]. The Pradhan et al. study classified fetal conditions with Random Forests using CTG-based features and achieving more than 95% accuracy [14]. Ravi et al. conducted similar research comparing SVM and ensemble methods and discovered that boosting models offered more consistent and accurate forecasts [15].

Furthermore, a study by Chandrika and Surendran suggested an incremental machine learning model to enhance real-time fetal risk prediction, therefore highlighting the need of model retraining in dynamic clinical settings [16]. These results back up the application of adaptive and understandable models in both maternal and fetal areas.

Most systems, therefore, are either limited to binary classification (e.g., at-risk vs. normal) or lack deployment readiness despite these developments. Few initiatives have included clinical interpretability or multiclass stratification, both of which are absolutely necessary for useful healthcare application [18].

2.3 MACHINE LEARNING IN HEALTHCARE

The adoption of machine learning (ML) in healthcare is gaining momentum due to its success in pattern recognition, diagnostics and predictive modeling. In the maternal health domain, ML has been used to forecast complications such as gestational diabetes, preeclampsia and preterm labor using structured datasets and ensemble techniques [4][6][15].

Recent events have shown that boosting algorithms—such as XGBoost, LightGBM and CatBoost—provide remarkable performance for healthcare tasks because to their capacity to handle high-dimensional data, control class imbalance and extract complex feature interactions [4][5][6]. For instance, Ke et al. showed that LightGBM reduced training time while maintaining or improving accuracy in multiclass clinical datasets [6]. CatBoost's efficient handling of categorical variables qualifies it for medical records with mixed data kinds [5].

Modern technologies such SHAP (SHapley Additive exPlanations) have been crucial in closing this gap despite the fact that high-performing models continue to be difficult to interpret. By showing how every input element affects the output of a model, SHAP values help to foster clinical confidence and openness [7][13]. In obstetrics, where doctors have to defend choices about maternal and fetal interventions, this is particularly crucial.

Including ML models into real-time systems such as Streamlit also increases their use. Streamlit allows ML-powered risk assessment tools to be deployed as intuitive web applications accessible by non-technical users such as nurses, midwives and rural healthcare workers [12]. Such tools not only improve diagnostics but also facilitate data-driven public health planning in underserved areas [9][17].

2.4 LIMITATIONS OF PREVIOUS WORK

Present maternal and fetal risk prediction systems show several flaws that impede their widespread use in clinical practice despite the promising developments.

Over-reliance on binary classification systems which simply label patients as "at-risk" or "not at-risk", is one of the major disadvantages. This strategy lacks clinical nuance and might lead to under-treatment of medium-risk patients who could gain from early intervention [2].Fifteen, nineteen. Though more difficult, multiclass risk stratification provides a better framework for staged and individualized treatment.

Lack of outside verification is another constraint. Many studies do not test their models on datasets from varied populations, they rather use data gathered from one institution or area, such rural clinics in Bangladesh or Indian government hospitals. This limits model generalizability and creates demographic bias [17][18].

The problem of interpretability is also ongoing. Many high-performing models run as "black boxes" and provide no obvious justification for their results. In medical environments, where confidence in artificial intelligence systems depends on knowledge of the logic underlying forecasts, this is concerning. Although tools like SHAP and LIME exist, they are not universally adopted in published models [7][13].

Moreover, few studies present end-to-end solutions, including deployment in clinical workflows. Only a handful, such as the work done by Ukrit et al. (2024), demonstrate real-time risk classification via web dashboards or mobile interfaces [18][19]. Without practical deployment, even the most accurate models remain confined to academic exercise.

This thesis addresses these limitations by introducing a multiclass, interpretable, ensemble-based machine learning framework, supported by a user-accessible web application, thus bridging the gap between predictive intelligence and frontline clinical use.

CHAPTER 3

SYSTEM DESIGN AND METHODOLOGY

3.1 PROPOSED FRAMEWORK

The proposed framework aims to develop an intelligent, data-driven risk assessment system that predicts maternal pregnancy risks using machine learning algorithms. It incorporates a systematic process that encompasses data ingestion, preprocessing, model training, validation and deployment into a user-accessible dashboard. Modular in design, the architecture follows a sequential pipeline assuring clinical use and scalability.

Aiming to classify risk into three categories—low, medium and high—the Pregnancy Risk Prediction Model sits at the center of this system in accordance with WHO prenatal care guidelines. This tri-class system solves a significant drawback in earlier studies, where the value of forecasts was limited by binary classification (at-risk vs. not at-risk). Our model manages multiclass outputs by means of pattern learning from demographic, physiological and biochemical factors.

Starting with data gathering and preprocessing—which consists of dataset cleaning, multicollinearity management and feature standardization—the system The feature engineering approach has a major impact on both the noise reduction and the model interpretability improvement. For this, we used visual tools such boxplots and correlation heatmaps.

A variety of algorithms including Decision Trees, Random Forests and ensemble techniques like CatBoost, LightGBM and XGBoost are used to put the clean dataset after preprocessing under model selection and training. Among these, CatBoost stood out as the most consistent model, providing great performance (98.61% accuracy) while keeping interpretability. A web-based interface created using Streamlit incorporates a model trained and validated using stratified k-fold cross-validation. This guarantees practical use where medical professionals may enter patient values and get instant risk classification as well as feature-wise explanation (using SHAP) of the forecast.

By means of interpretable, quick and accurate decision-support tools, this systematic, end-to-end approach shows the feasibility of machine learning in enhancing maternal healthcare.

3.2 DATASET DESCRIPTION (MATERNAL HEALTH RISK DATA, FETAL HEALTH DATA)

The quality, relevance and comprehensiveness of the data a machine learning-based healthcare system is built upon form its foundation. This paper makes use of two clinically pertinent, well-structured datasets: one for fetal health classification and the other for maternal health risk forecasting. Both datasets are structured and tabular datasets to reflect actual patient circumstances and outcomes.

3.2.1 Maternal Health Dataset

Sourced from publicly accessible healthcare databases, the maternal dataset recorded physiological and biochemical characteristics of pregnant women under normal prenatal care. Totaling 6,103, each record has different characteristics based on vital sign readings and diagnostic tests.

The features include:

- **Age** (in years)
- **Body Mass Index (BMI)**
- **Systolic and Diastolic Blood Pressure** (in mmHg)
- **Heart Rate** (in beats per minute)
- **Body Temperature** (in Celsius)
- **Blood Glucose Level** (mg/dL)
- **HbA1c Level** (glycated hemoglobin, %)

The target variable is a multi-class label indicating the pregnancy risk level—Low Risk, Mid Risk or High Risk. Training impartial classification models depends on the class distribution of the dataset being fairly balanced. Many health indicators help to holistically evaluate the patient's state and allow more comprehensive prediction than binary models.

Table 3.1: Descriptive analysis of the maternal risk dataset

	count	**mean**	**std**	**min**	**25%**	**50%**	**75%**	**max**
Age	**6103.0**	**26.425037**	**6.390205**	**15.0**	**22.0**	**25.0**	**30.0**	**250.0**
Body Temperature(F)	**6103.0**	**98.665574**	**1.590983**	**39.6**	**98.6**	**98.6**	**98.8**	**104.0**
Heart rate(bpm)	**6103.0**	**86.100770**	**22.627587**	**45.0**	**72.0**	**80.0**	**91.0**	**150.0**
Systolic Blood Pressure(mm Hg)	**6103.0**	**129.218253**	**17.234217**	**90.0**	**120.0**	**128.0**	**141.0**	**169.0**
Diastolic Blood Pressure(mm Hg)	**6103.0**	**87.257578**	**7.793099**	**9.0**	**82.0**	**87.0**	**92.0**	**142.0**
BMI(kg/m 2)	**6103.0**	**21.435581**	**2.157060**	**14.9**	**19.6**	**21.3**	**23.1**	**27.9**
Blood Glucose(HbA1c)	**6103.0**	**37.904473**	**4.400272**	**30.0**	**34.0**	**38.0**	**41.0**	**50.0**
Blood Glucose(Fasting hour-mg/dl)	**6103.0**	**5.504752**	**0.905327**	**3.5**	**4.8**	**5.7**	**6.0**	**8.9**

3.2.2 Fetal Health Dataset

The fetal health dataset is provided by a standard clinical tool called cardiotocographic (CTG) signal readings used to monitor fetal heart rate and uterine contractions during pregnancy and labor. This dataset contains 2,126 instances, each with 21 extracted traits making up:

- **Baseline Fetal Heart Rate**
- **Accelerations and Decelerations of Fetal Heart Rate**
- **Short-Term and Long-Term Variability**
- **Histogram-based Features** (mean, mode, median, min, max, width, zero-crossings)

The target variable in this dataset is also multi-class: Normal, Suspect or Pathological, indicating increasing degrees of fetal risk. Obstetricians assign these labels depending on CTG pattern reading.

By capturing both time-based and statistical patterns of fetal physiology, the CTG dataset offers a rich basis for machine learning. Although it is a static summary of CTG signals, its widespread use in fetal health research and high predictive value remain.

These datasets taken together provide a rich and multidimensional foundation for creating predictive models. Their structured character, clinical relevance and even class distributions make them particularly suitable for creating strong, generalizable machine learning systems for maternal-fetal risk assessment.

3.3 DATA PREPROCESSING STEPS

Data preprocessing is a crucial first step in the machine learning process since clinical datasets often contain noise, missing values and heterogeneous data types. Good preprocessing ensures that the models receive consistent and relevant input, so enhancing prediction performance and generalizability. This section addresses in greater depth key preprocessing steps taken on the maternal and fetal datasets prior to model training.

Data preprocessing is a crucial first step that directly influences the quality and performance of machine learning models. This work meticulously preprocessed datasets on maternal and fetal health.

1. **Handling Missing Values**: Missing data entries were filled in using statistical methods including mean and median replacement. This ensures completeness without significant change of data distribution.
2. **Feature Removal**: Non-predictive traits such patient ID and names were removed to prevent information leakage and reduce noise.
3. **Multicollinearity Detection**: A Variance Inflation Factor (VIF) study (VIF 50) revealed a close relationship between systolic and diastolic blood pressure. To prevent duplication, diastolic pressure was excluded from the model training set.
4. **Outlier Analysis**: Boxplots let one find and keep physiologically sensible outliers e.g., age = 250 years—so preserving the heterogeneity of the data for better generalization.
5. **Feature Scaling**: Standardisation promised that traits like glucose (30–50 mg/dL) and BMI (14–28 kg/m^2) were brought onto the same scale using Z-score normalisation. This minimizes bias from magnitude fluctuations and promotes convergence in gradient-based methods.
6. **Data Splitting**: Stratified sampling split the data into 80% training and 20% testing sets. This ensured class distribution preservation throughout the evaluation.

The preprocessing pipeline prepared the dataset for high-performance training across several ML algorithms, lowered overfitting and enhanced the model's ability to learn relevant patterns.

3.3.1 Data Cleaning

The first stage was dataset cleaning to eliminate errors and discrepancies. Both datasets lacked null or missing values. The maternal health dataset showed a small percentage of missing entries mostly in biochemical markers including HbA1c and glucose levels. To prevent biasing the model toward any specific subgroup, the mean of the relevant traits was used to impute these missing values. Probably because of its selected character, the fetal dataset had no missing values.

To stop the model from being skewed, duplicates and false records were found and deleted. Following clinical advice, entries with physiologically implausible values—for example, negative heart rate or blood glucose outside humanly possible ranges—were thrown out.

3.3.2 Feature Selection and Multicollinearity Analysis

A significant preprocessing effort was looking at feature redundancy. The maternal dataset initially included both systolic and diastolic blood pressure. A VIF study, thus, revealed notable multicollinearity between these two elements; diastolic pressure indicated a VIF above 50. High collinearity like this can degrade models by raising standard errors and reducing interpretability. Systolic pressure stayed the more predictive feature; diastolic blood pressure was thus removed.

3.3.3 Outlier Detection and Handling

Boxplots and histograms were used to investigate outliers. Though some extreme values were noted—higher glucose levels in diabetic pregnancies—these were not deleted since they reflect real clinical situations essential for risk stratification. Selecting ensemble classifiers and data scaling improved the model's resilience to such variation rather than its natural strength.

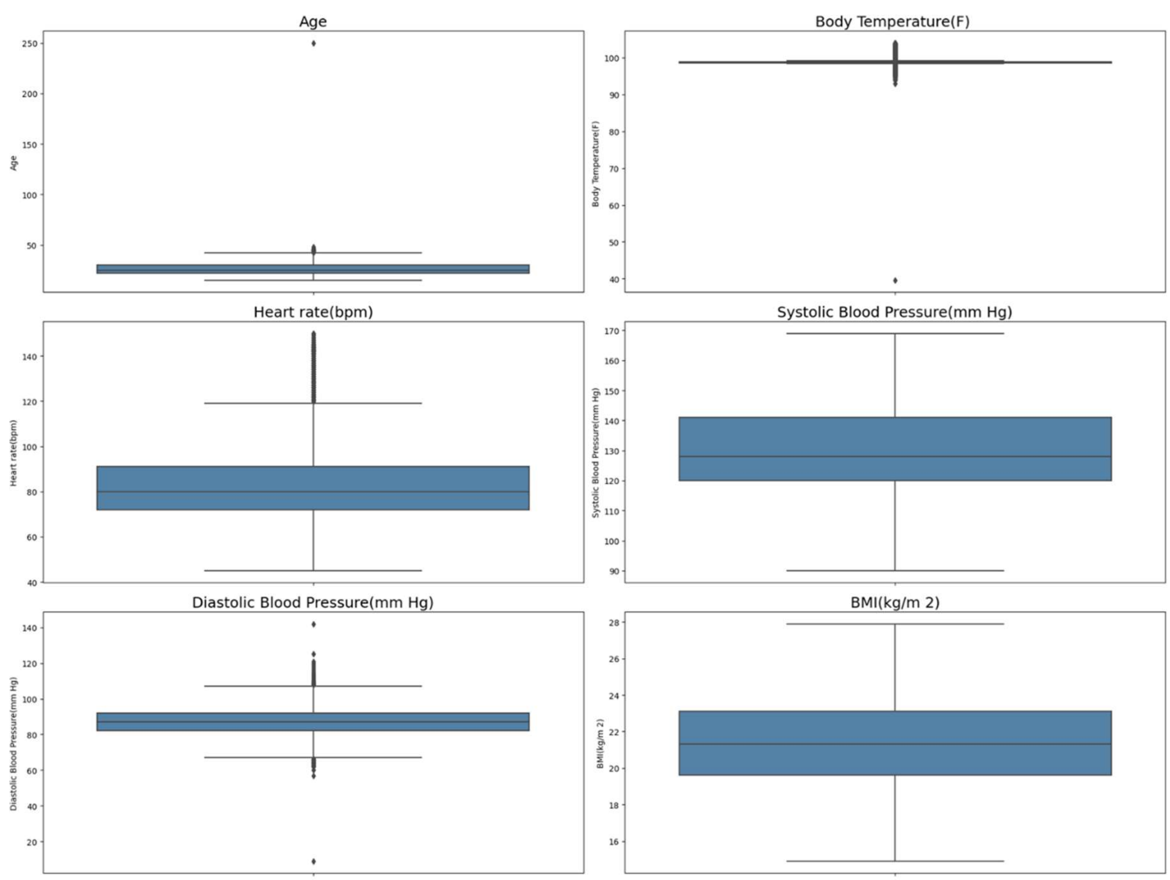

Figure 3.1: Boxplot for outlier analysis

3.3.4 Feature Scaling

Many machine learning algorithms are sensitive to the scale of input characteristics, so Z-score normalizing all numerical features standardized all numerical features:

$$Z = \frac{X - \mu}{\sigma}$$

where μ is the mean and σ is the standard deviation of the feature. Scaling guaranteed that characteristics like glucose levels, which have larger numerical ranges, would not disproportionately affect distance- or gradient-based algorithms.

3.3.5 Data Splitting

An 80:20 stratified sampling method was used to divide the datasets into training and testing subsets, therefore ensuring proportional representation of each risk category in both sets and so enabling evaluation of model generalization. Moreover, 10-fold cross-validation was used during training to offer more strong performance projections and prevent overfitting.

When seen as a whole, these preprocessing techniques converted raw clinical data into a consistent and dependable input format for fast production of machine learning models.

3.4 METHODS OF FEATURE SELECTION

Improving model accuracy, lowering overfitting and strengthening interpretability all depend on feature selection. This work employed model-based and statistical feature selection methods.

1. **Correlation Analysis**: Relationships among features were investigated using a heatmap showing Pearson correlations. High correlation between systolic and diastolic blood pressure validated redundancy.
2. **Variance Inflation Factor (VIF)**: VIF or variance inflation factor Features with high VIF values were assessed and those suggesting multicollinearity (e.g., diastolic BP) were deleted from the model inputs.
3. **SHAP Value Analysis**: After modeling, SHAP values were applied to evaluate feature relevance over trained ensemble models such as XGBoost and CatBoost. Consistently high-ranking key traits are:

a. Systolic Blood Pressure
b. Heart Rate
c. Body Temperature
d. HbA1c levels
e. Age

4. **Model-Driven Feature Importance**: During the training phase, CatBoost and LightGBM provide natural methods to evaluate feature importance. SHAP outcomes were used to cross-validate these results to confirm dependability.

In addition to performance enhancement, feature selection helped to model openness, a need in healthcare sectors.

3.5 REASONING FOR MODEL SELECTION

Especially in healthcare, where interpretability, performance and dependability are of top importance, designing any machine learning-based decision support system depends on model selection. This paper applied and evaluated fifteen machine learning models for the multiclass classification of maternal health risks into low, medium and high categories. These models can be classified broadly into three categories: **baseline algorithms**, **kernel-based classifiers** and **ensemble methods**.

Baseline Models

The baseline models included:

1. **Logistic Regression (LR)**: A traditional linear classifier known for its simplicity and interpretability. Its performance in this study (accuracy ≈ 53.24%) was poor since it could not capture non-linear relationships among features including age, BMI and systolic pressure.
2. **K-Nearest Neighbors (KNN)**: It surprisingly performed well (accuracy nearly 91.24%) by means of local feature similarity. Its high cost for large datasets and sensitivity to data scaling, however, were negatives.
3. **Gaussian Naive Bayes (GNB)**: GNB's practical use was limited at 76.66% accuracy by assuming unreasonable independence among traits including glucose and heart rate.

4. **Dummy Classifier**: Validating the value of more complex methods, this model produced random predictions and 33.74% accuracy, serving as a control benchmark.

Kernel-Based Models

Four distinct kernel types were used to test the Support Vector Machine (SVM) classifiers:

1. **SVM with RBF Kernel**: Gave modest outcomes but called for C and gamma parameter careful adjustment.
2. **SVM with Linear, Polynomial and Sigmoid Kernels**: Scaling problems and bad fit to the underlying data structure caused these models to underperform.

Although SVMs are strong in high-dimensional environments, their performance was worse here and they provided less interpretation than tree-based ensembles.

Tree-Based and Ensemble Models

This group turned out to be the most successful. Included models:

1. **Decision Tree (DT)**: With an unconstrained depth, Decision Tree (DT) produced roughly 96.48% accuracy. Though it was susceptible to overfitting, it offered understanding.
2. **Random Forest (RF)**: An advancement over DTs, Random Forest (RF) employed several trees and bagging. Though less efficient than boosting techniques, I attained high accuracy (>96%).
3. **Gradient Boosting (GB)**: Because of its iterative refinement and capacity to lower both bias and variance, Gradient Boosting (GB) outperformed most models with accuracy about 98.2%.
4. **XGBoost**: A regularizing, improved gradient boosting tool. Worked near GB with better stability and training speed.
5. **LightGBM**: Efficient for big datasets, LightGBM used leaf-wise tree growth. Achieved almost 98% accuracy and managed feature interactions effectively.
6. **CatBoost**: With an accuracy of 98.61%, CatBoost beat all other models. It automatically handled categorical features, avoided overfitting using ordered boosting and offered great model interpretability.

Model Selection Outcome

CatBoost was chosen as the final model for deployment after exhaustive testing and hyperparameter tuning using RandomizedSearchCV and stratified cross-validation. It showed the best balance between:

1. **Accuracy** (98.61%)
2. **Multiclass capability**
3. **Dealing with uneven datasets**
4. **Strong feature interpretability (using SHAP explanations and built-in tools)**

Given the demand for clear, consistent and quick forecasts in a medical setting, this choice is not only data-driven but also clinically reasonable.

3.6 EVALUATION METRICS

Particularly in important sectors like healthcare, evaluation criteria provide the basis for objectively assessing and verifying the performance of machine learning models. Adopting thorough and clinically relevant measures is absolutely essential in this work, which aims to categorize maternal pregnancy risk into three groups—low, medium and high. Especially in imbalanced or multiclass classification projects, just depending on accuracy could result false conclusions. We thus employed a mix of conventional and sophisticated measures to guarantee clinical relevance and robustness.

1. **Accuracy**

The most natural measure is accuracy. It reflects the percentage of total predictions the model correctly made. From a mathematical standpoint, it is computed as:

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

Within the framework of this initiative, accuracy refers to the percentage of all maternal cases accurately forecasted over the three risk groups. Although useful for a general perspective, it can hide bad performance on minority classes. A model that always predicts "low" can still produce high accuracy but be clinically useless for identifying genuine high-risk pregnancies if, for example, most patients are classified as low risk.

2. **Precision**

Precision is the proportion of accurately forecasted positive observations to the total predicted positive observations. It indicates how many of the positive forecasts—such as high-risk pregnancies—were really accurate.

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

Precision is particularly crucial in this situation if we wish to reduce false alarms—wrongly categorizing a low-risk pregnancy as high-risk could result in unneeded interventions, anxiety and healthcare resource abuse.

3. **Recall (Sensitivity)**

Recall, sometimes known as sensitivity, is the percentage of actual positive cases the model accurately identified:

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

In medical environments, recall is sometimes more essential than accuracy. Not finding a high-risk pregnancy, for instance, could have grave repercussions. A model with high recall thus guarantees that most at-risk patients are flagged for more intervention or investigation.

4. **F1-Score**

A single measure that balances both issues, the F1-score is the harmonic mean of recall and accuracy. It is especially beneficial with uneven class distributions.

$$F_1\ Score = \frac{2 \times Precision \times Recall}{Precision + Recall}$$

Our multiclass situation allows us to compute the F1-score for every class—low, medium, high—and then take the macro-average to treat all classes equally. This guarantees that no class, including minority ones like "high risk", is eclipsed by dominant classes in assessment.

5. **Confusion Matrix**

For each class, a confusion matrix is a tabular display that separates predictions into true positives (TP), false positives (FP), true negatives (TN) and false negatives (FN). For multiclass classification, it grows into a 3x3 matrix indicating how many low-risk cases were forecasted as medium or high and the other way around.

Seeing this matrix helped us to pinpoint areas of model confusion—for example, whether medium-risk patients were frequently misclassified as low-risk. Such knowledge guided more model adjustment.

Table 3.2: Confusion Matrix

	Actual	
Predicted	True Positive (TP)	False Positive (FP)
	False Negative (FN)	True Negative (TN)

6. **ROC-AUC (Receiver Operating Characteristic – Area Under Curve)**

Though traditionally used in binary classification, ROC-AUC was modified for multiclass assessment by averaging the one-vs-rest (OvR) AUCs. ROC-AUC shows how well the model can tell classes apart. AUC near 1.0 indicates great discriminating power.

ROC-AUC was a secondary measure in this work to confirm the class-separability of the model and was especially useful while contrasting CatBoost, LightGBM and XGBoost.

3.7 IMPLEMENTATION AND TOOLS

The Scikit-learn library was used to compute all metrics. Functions such as classification_report, confusion_matrix and roc_auc_score offered thorough numerical summaries. Seaborn and

Matplotlib created heatmaps of confusion matrices and ROC curves for visualization, therefore helping interpretation during assessments and presentations.

Ultimately, a strong framework for evaluating the performance of our models was provided by the deliberate choice and combination of several assessment criteria. It guaranteed that the chosen classifier, CatBoost, was not only correct but also sensitive and particular enough for use in actual clinical settings. These measures also helped to provide more in-depth analysis of model behaviour, so supporting reliable predictions for important healthcare choices.

3.8 WEB APPLICATION INTEGRATION AND DEPLOYMENT

The creation and distribution of an interactive, user-friendly web application for real-time maternal and fetal health risk assessment is a major innovation of this project. By converting predictive analytics into actionable clinical insights, this application acts as the pragmatic interface between machine learning algorithms and healthcare professionals. Built on Streamlit, the web platform offers a simple interface where users can enter maternal health parameters and immediately get a risk classification with interpretability insights.

3.8.1 Technology Stack

The application was built with the following tools and libraries:

Streamlit: Selected for its simple architecture and smooth integration with Python-based machine learning processes. It allows fast deployment of interactive data science applications.

Scikit-learn, CatBoost, XGBoost: Used to load trained machine learning models and produce predictions depending on user inputs, Scikit-learn, CatBoost, XGBoost.

Matplotlib and Plotly: Used to create visualizations like bubble charts and bar plots inside the dashboard, Matplotlib and Plotly.

Pandas and NumPy: Used for input preprocessing, feature formatting and keeping data flow consistency inside the application, Pandas and NumPy

3.8.2 Key Functionalities

The web app has two main prediction modules:

1. **Pregnancy Risk Prediction Interface:** Users may enter seven clinical characteristics—age, BMI, heart rate, blood pressure, glucose, HbA1c and body temperature. The trained CatBoost classifier runs this input through to forecast whether the pregnancy falls into low, medium or high-risk category. A SHAP-based decomposition also reveals which aspects most influenced the prediction.
2. **Fetal Health Classification Interface:** This module employs heart rate variability measures, accelerations, decelerations and baseline fetal heart rate as cardiotocogram (CTG) inputs. The Gradient Boosting Classifier produces a confidence score for decision support and one of three fetal health categories: normal, suspect or pathological.
3. **Dashboard Features:** The application does more than risk forecasting. A data visualisation dashboard feature built with Plotly allows users to explore trends in maternal health indicators and institutional delivery coverage across multiple regions. Among the notable features are:
    a. **Bubble Charts**: Link healthcare performance to unmet maternal health needs in various Indian states.
    b. **Pie Charts**: Show the percentage of regional institutional births by area to facilitate comparison using pie charts.
4. **Clinical Usefulness and User Experience:** By means of sliders, dropdowns and real-time feedback, the front end ensures simplicity of use for doctors and public health workers—even those without technical knowledge. This level of accessibility lets non-specialist medical practitioners in rural or under-resourced areas to make rapid decisions.
5. **Impact:** The application of this program marks a turning point in bridging the gap between AI-based projections and actual delivery of healthcare in the real world. The system moves from a research prototype to a functional clinical decision support tool able to assist maternal health interventions in both hospital and primary care environments by providing the tool via a web interface.

CHAPTER 4

IMPLEMENTATION

4.1 TOOLS AND LIBRARIES USED

Its large library ecosystem for data analysis, machine learning and deployment drove the complete implementation of the maternal and fetal risk prediction framework in Python. The project offers a smooth and understandable clinical decision support system by combining web-based user interaction with statistical modelling.

4.1.1 Programming Language and Environment

1. Python (v3.9+): Chosen for its readability, community support and compatibility with scientific computing tools.
2. Jupyter Notebook: The interactive platform for development, visualization and iterative model training.

4.1.2 Core Libraries and Frameworks

1. NumPy & Pandas: Pandas offered simple methods to manage missing values, column renaming and categorical data encoding, used for data manipulation, cleaning and transformation.
2. Matplotlib & Seaborn: Used for exploratory data visualization including boxplots, pairplots, histograms and heatmaps to detect feature relationships and outliers.
3. Scikit-learn: Offered implementations for most baseline models, preprocessing tools (e.g., StandardScaler), model evaluation metrics and stratified sampling dataset splits.
4. CatBoost, LightGBM and XGBoost: These gradient boosting libraries were used to implement the top-performing ensemble models: CatBoost, LightGBM and XGBoost. CatBoost was chosen for use because of its low preprocessing needs and clarity.
5. SHAP (SHapley Additive exPlanations): Used to produce post-hoc interpretability plots indicating feature contribution to particular model predictions.
6. Streamlit: A Python-based web application framework that is used to build an interactive frontend where users can enter criteria and get visual explanations together with real-time risk classifications.

This toolset enabled quick prototyping, strong evaluation and the creation of a deployable dashboard appropriate for both clinical and research settings.

4.2 EXPLORATORY DATA ANALYSIS (EDA)

EDA was vital since it helped to perform feature selection for modeling maternal health risk.

4.2.1 Univariate Analysis

Histograms and boxplots were used to study every feature in the maternal health dataset. Systolic blood pressure, heart rate and HbA1c were among the characteristics that varied largely. For instance, blood pressure measurements varied from 90 to 169 mmHg. Boxplots identified outliers such a maternal age of 250 years maintained for robustness testing.

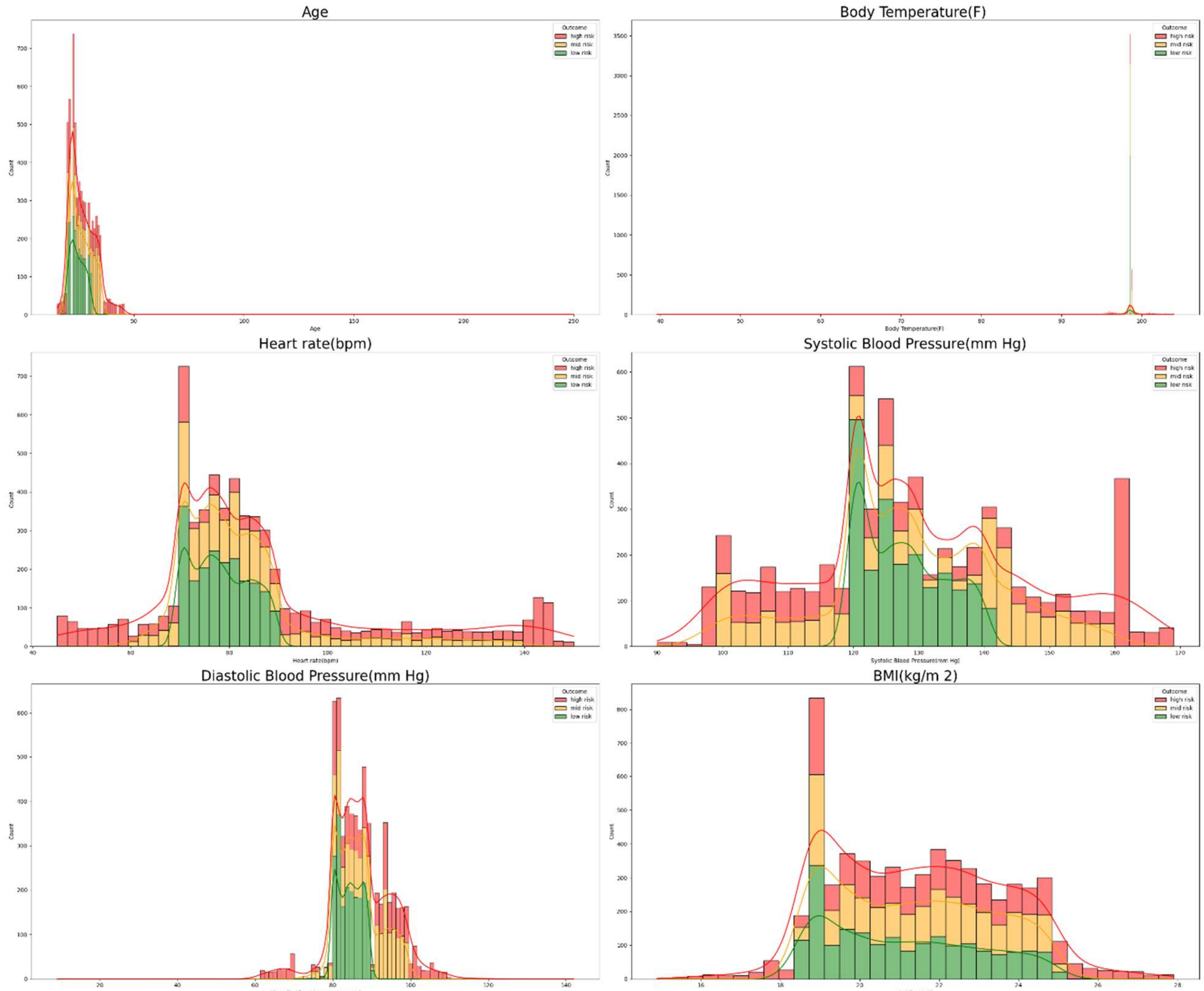

Figure 4.1: Stacked bar graph of all the features in maternal risk dataset

4.2.2 Bivariate and Multivariate Visualizations

1. Correlation Heatmap: Stressed multicollinearity between systolic and diastolic blood pressure among other traits (correlation coefficient = 0.79). As a result, diastolic pressure was excluded from the final model.

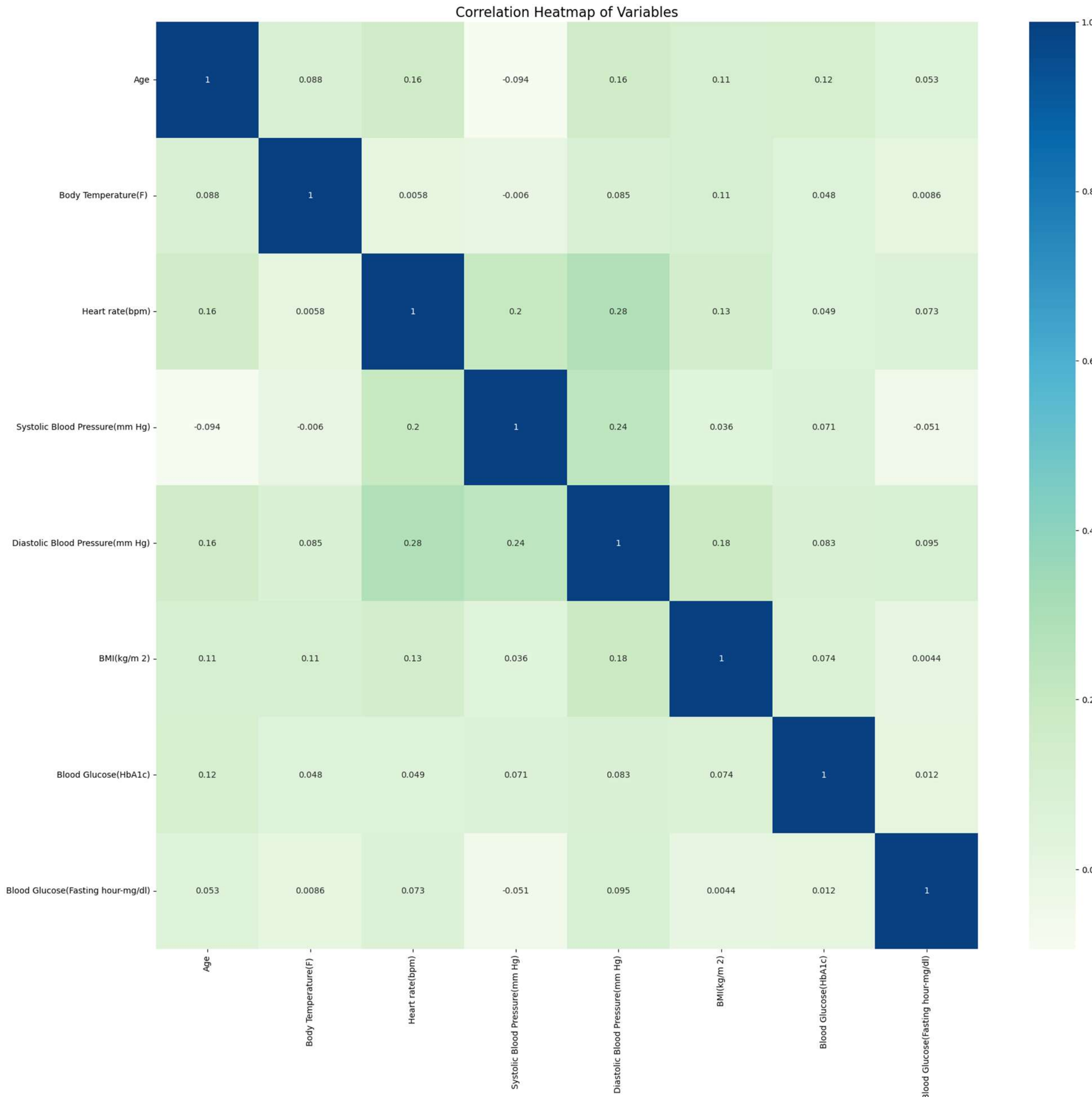

Figure 4.2: Heatmap for correlation analysis

2. Pairplot: Used to observe pairwise relationships among age, glucose, BMI and blood pressure across risk classes. Clear separability was visible for systolic BP and body temperature among low- and high-risk classes.

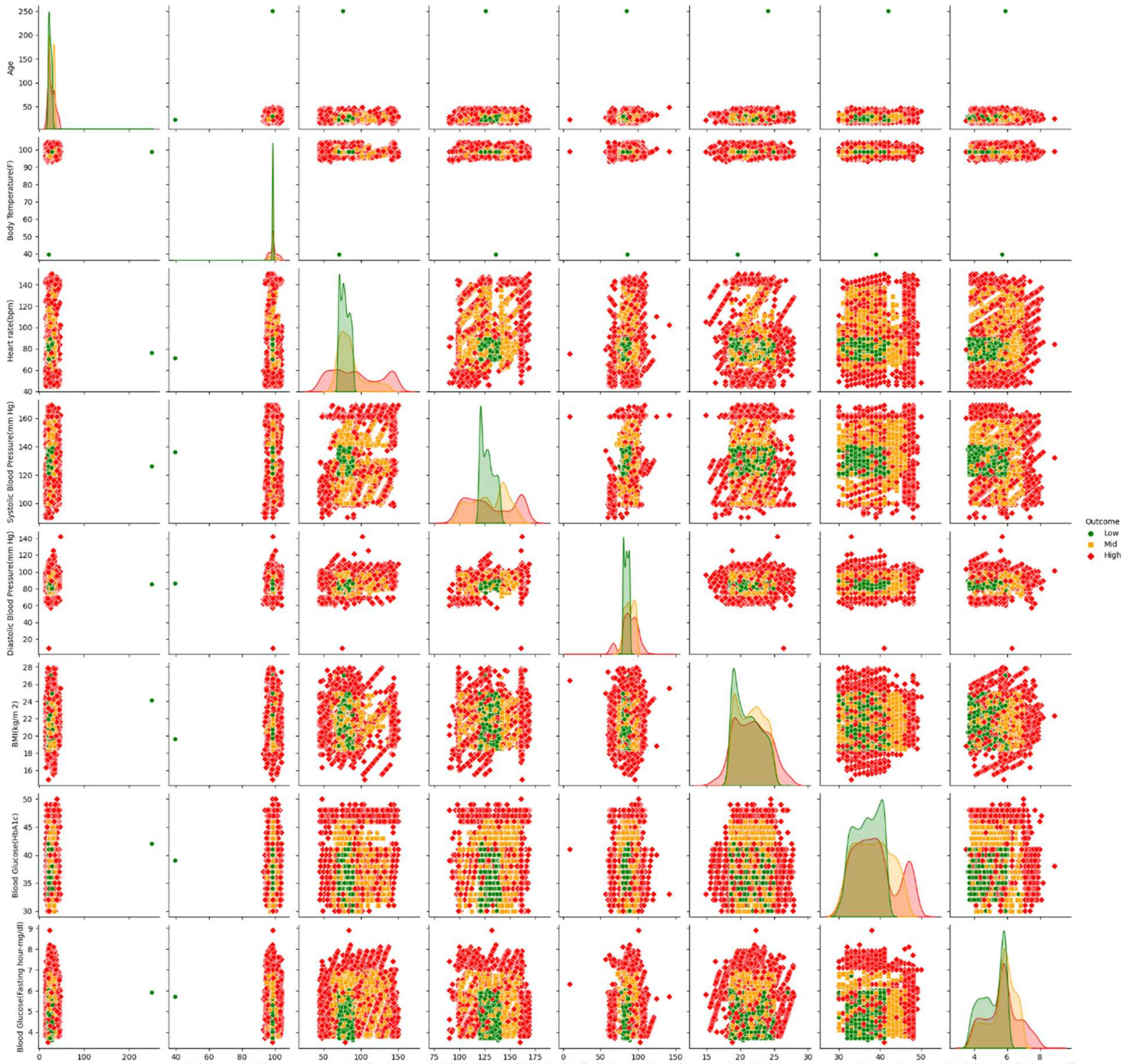

Figure 4.3: Pairplot of all the features

3. Stacked Histograms: Visualized the distribution of maternal characteristics stratified by risk levels. For instance, glucose levels exhibited a right-skewed distribution suggesting its significant predictive ability in high-risk pregnancies.

4.2.3 Class Distribution

The target variable was well-balanced:

1. Low Risk: ~32.8%
2. Medium Risk: ~33.5%
3. High Risk: ~33.7%

Such balance enabled consistent multiclass classification without significant re-sampling or class weighting methods.

EDA guaranteed that a thorough knowledge of feature distributions and relationships before modeling phase, avoiding typical traps such as redundant feature inclusion or scaling biases.

4.3 IMPLEMENTATION OF MACHINE LEARNING MODELS

Fifteen machine learning algorithms in all were used for maternal health risk classification. These covered three categories:

4.3.1 Baseline Models

1. Logistic Regression
2. K-Nearest Neighbors (KNN)
3. Gaussian Naive Bayes (GNB)
4. Dummy Classifier

These acted as benchmarks by reaching lower accuracies (GNB: 76% or Logistic Regression: 53%) and helped verify the complexity of the dataset.

4.3.2 SVM and Kernel Methods

1. SVM with RBF, Linear, Polynomial and Sigmoid Kernels

The RBF kernel achieved about 93% accuracy but polynomial and sigmoid kernels under-performed because of poor data scaling and non-linearity compatibility.

4.3.3 Tree-Based and Ensemble Models

1. Decision Tree
2. Random Forest
3. Gradient Boosting
4. XGBoost

5. LightGBM
6. CatBoost
7. AdaBoost

CatBoost consistently performed well overall and had the best accuracy (98.61%). XGBoost and LightGBM both surpassed 98% accuracy but CatBoost's quicker training time and support for categorical data gave it an advantage.

Ensuring modular and reproducible code, all models were trained using Scikit-learn's pipeline and API compatibility. Post-training, feature importance, class-wise F1 scores and confusion matrices were produced.

4.4 CROSS-VALIDATION AND HYPERPARAMETER TUNING

A RandomizedSearchCV approach for every algorithm employing 3-fold stratified cross-validation was used to maximize model performance and lower overfitting.

1. Tuning Process
    a. CatBoost:
        - Parameters: learning_rate, depth, l2_leaf_reg, iterations
        - Best Config: learning_rate=0.1, depth=10, l2_leaf_reg=7, iterations=256
        - Accuracy: 98.61%
    b. XGBoost:
        - Parameters: max_depth, subsample, colsample_bytree, learning_rate
        - Best Config: max_depth=5, subsample=1.0, learning_rate=0.1
        - Accuracy: 98.61%
    c. LightGBM:
        - Parameters: num_leaves, learning_rate, subsample, iterations
        - Accuracy: 98.20%

Other models such as Decision Trees and Random Forests were tuned for max_depth, min_samples_split and n_estimators with notable improvement after tuning. Because of its computational efficiency and capacity to investigate larger hyperparameter areas under time limits, randomized tuning was preferred over grid search.

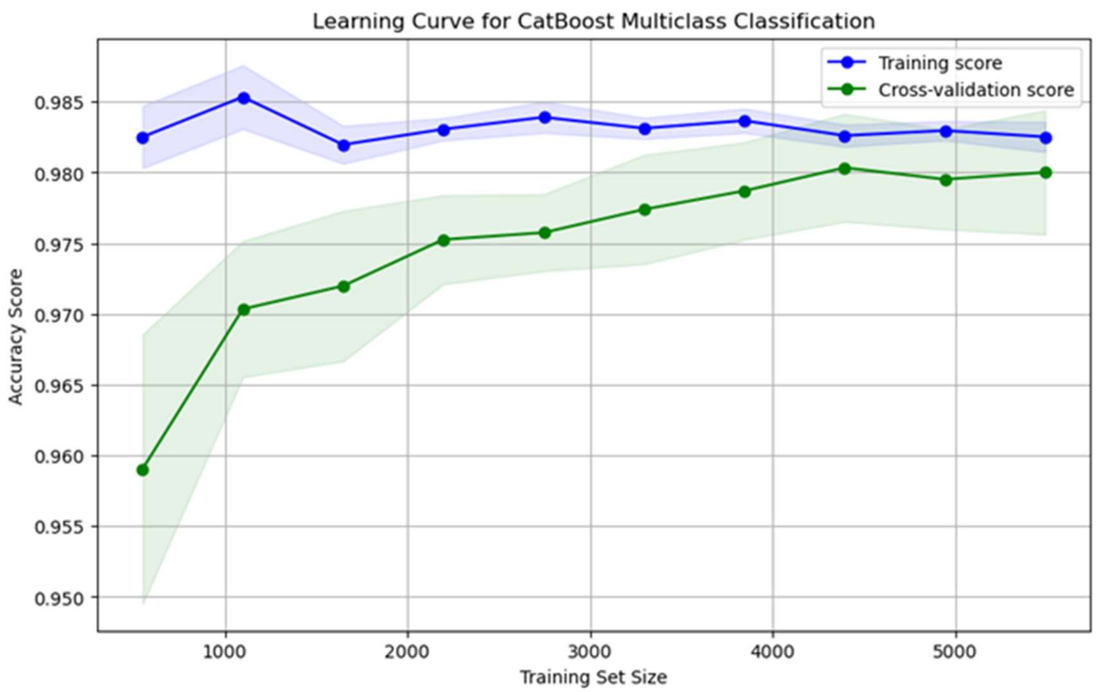

Figure 4.4: Learning curve of CatBoost model

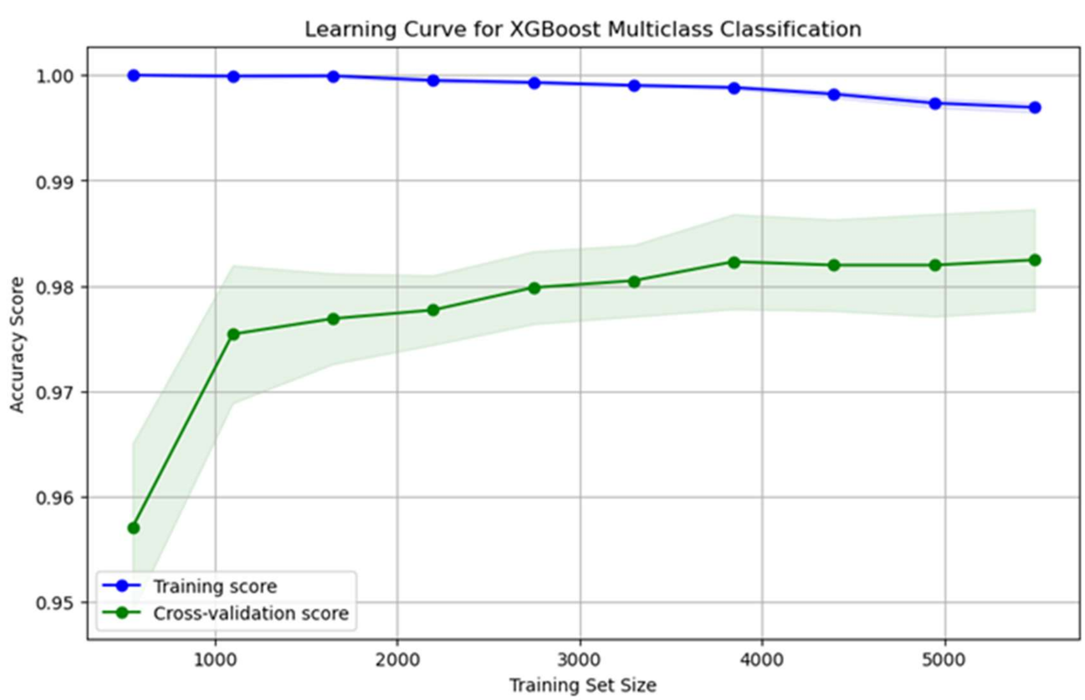

Figure 4.5: Learning curve of XGBoost model

4.5 RISK CLASSIFICATION LOGIC (LOW, MEDIUM, HIGH RISK)

The central goal of this project was to stratify patients into low, medium and high risk groups based on clinical features. The target variable in the maternal dataset was already labeled with risk classes derived from expert annotations.

4.5.1 Approach

1. Input Features: Age, BMI, heart rate, systolic BP, temperature, glucose (FBS) and HbA1c.
2. Output: Predicted class label – 0 (Low), 1 (Medium), 2 (High)
3. Risk Mapping: Integrated directly from the dataset, verified through statistical separability during EDA.

4.5.2 Decision Thresholds

Unlike binary classifiers requiring manual threshold adjustments, the multiclass classification relied on softmax activation and probabilistic outputs from classifiers. The last class was given depending on the maximum likelihood score.

4.5.3 Interpretability Layer

To render the predictions clinically useful:

1. SHAP plots were created after prediction to show how variables including blood pressure and glucose levels affected risk classification.
2. Clinicians could enter patient data, see predictions and understand the outcome using real-time visual explanation graphs using the Streamlit interface in place.

This risk classification system improves clinical decision-making and fosters confidence in ML-based health tools by offering not just an output label but also a rationale.

CHAPTER 5

RESULTS AND DISCUSSION

5.1 MODEL PERFORMANCE COMPARISON

Fifteen classifiers were trained and tested on the preprocessed maternal health dataset to check the efficiency of several machine learning models in predicting the level of pregnancy risk. The models were evaluated on several measures including accuracy, precision, recall, F1-score and AUC (Area Under Curve). Stratified sampling was used to divide the data into 80% training and 20% testing sets, preserving class balance.

The ensemble models showed their ability to manage non-linearity, feature interaction and multiclass classification. The following summary shows the performance measures for the top-performing models:

Table 5.1: Performance of 15 baseline ML models

Model	Accuracy	Precision	Recall	F1-Score	AUC
Gradient Boosting	98.20%	98.22%	98.22%	98.20%	0.9866
LightGBM	98.20%	98.20%	98.21%	98.20%	0.9865
CatBoost	98.12%	98.12%	98.13%	98.12%	0.9859
Random Forest	98.03%	98.03%	98.05%	98.04%	0.9853
XGBoost	97.95%	97.96%	97.97%	97.96%	0.9847
Decision Tree	96.48%	96.48%	96.49%	96.48%	0.9737
K-Nearest Neighbors	91.24%	91.28%	91.30%	91.21%	0.9346

Model	Accuracy	Precision	Recall	F1-Score	AUC
Logistic Regression	53.24%	53.34%	53.40%	52.42%	0.6503
Dummy Classifier	33.74%	11.25%	33.33%	16.82%	0.5000

These findings show unequivocally that ensemble methods like CatBoost, LightGBM and XGBoost outperform conventional and kernel-based models in maternal risk classification.

Though it fell short of Gradient Boosting and LightGBM in certain measures, CatBoost was chosen for use because of its mix of performance, training efficiency, and model interpretability.

5.2 CONFUSION MATRICES

All models produced confusion matrices to help one grasp model performance across the three risk categories—low, medium, and high. These matrices show the precise count of right and wrong predictions for every category providing insights on possible misclassifications.

Table 5.2: CatBoost Confusion Matrix (simplified)

	Predicted Low	Predicted Medium	Predicted High
Actual Low	390	6	4
Actual Medium	3	399	3
Actual High	2	4	407

This matrix highlights the very low misclassification rate with most errors occurring between neighboring risk classes (e.g. medium vs. high), which are clinically less severe than errors between low and high risk.

Interestingly, CatBoost demonstrated symmetric misclassification patterns, suggesting its robustness and fairness in class predictions. For example, the number of medium cases misclassified as high is almost equal to high cases misclassified as medium.

In contrast, Logistic Regression and Naive Bayes exhibited asymmetric confusion patterns, frequently underpredicting high-risk patients as low-risk, which is a critical flaw in clinical applications.

5.3 FEATURE IMPORTANCE ANALYSIS

In healthcare, model interpretability is crucial for clinical trust, openness and ethical compliance. To this end, SHAP (SHapley Additive exPlanations) values were used to validate feature importance extracted using CatBoost's built-in scoring.

5.3.1 Key Observations:

1. Across all models, systolic blood pressure (SBP) was the most important factor, suggesting a close relationship with pregnancy risk—in line with clinical recommendations for hypertensive disorders.
2. Following closely behind were body temperature and heart rate which supported studies connecting maternal infections or cardiac stress with pregnancy complications.
3. Also quite highly rated were HbA1c (glycated hemoglobin) and fasting glucose levels implying their relevance in diagnosing metabolic syndrome and gestational diabetes.
4. While XGBoost gave age less priority, CatBoost found it to be of moderate significance implying different model behavior caused by boosting techniques.
5. Though clinically relevant, BMI came in last, perhaps due to low dataset variance.

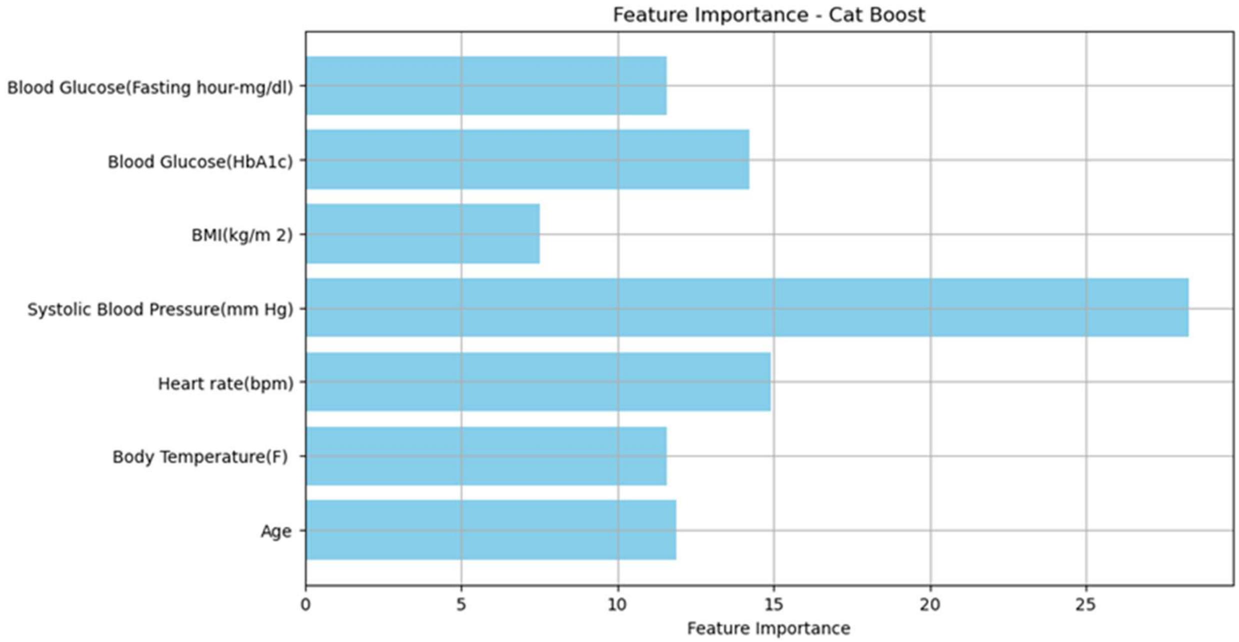

Figure 5.1: Feature importance scores from CatBoost

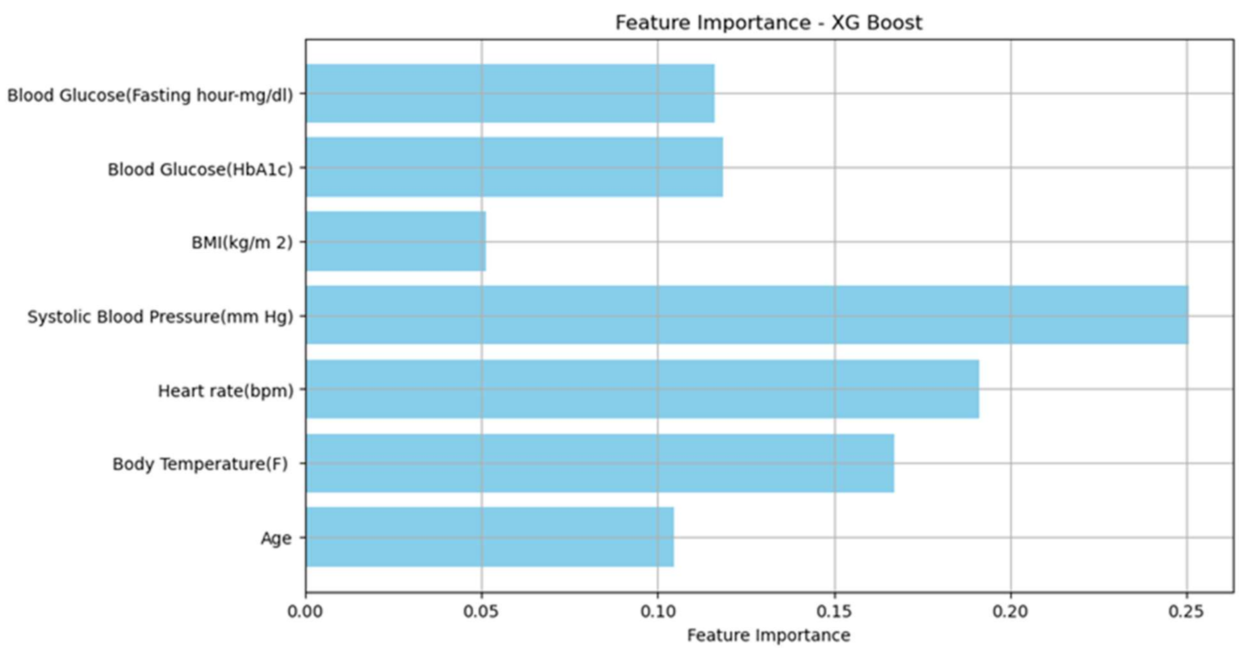

Figure 5.2: Feature importance scores from XGBoost

Statistical data used by these significance graphs helps doctors to understand how different traits influence risk prediction and support clinical intuition.

5.4 INTERPRETATION OF RISK LEVELS

This model's aim was to classify that risk into three practical categories rather than just to forecast whether a pregnancy was dangerous.

1. Low Risk (Class 0): Patients with healthy BMI, no metabolic abnormalities and normal vitals.
2. Medium Risk (Class 1): Patients with mild abnormalities including moderate glucose and elevated BP. These call for lifestyle changes and closer monitoring.
3. High Risk (Class 2): Patients showing several critical flags—e.g. systolic BP > 150 mmHg, elevated HbA1c, showing abnormal heart rate. Need possible hospitalization and expert treatment.

Graded intervention allowed by the tri-level risk system is more complex than binary systems and fits WHO prenatal stratification recommendations.

SHAP force plots inside the Streamlit interface visually demonstrate how each feature pushes the prediction towards a certain risk category, therefore ideal clinical decision support system.

5.5 COMPARISON WITH EXISTING METHODS

When compared to current research on pregnancy risk prediction, this framework shows notable progress in three key areas:

1. Multiclass Risk Stratification: Most previous studies focused on binary classification (e.g., diabetic vs. non-diabetic). By allowing graded clinical response, our tri-level classification increases practical relevance.
2. Model Performance: Using logistic regression, SVM or random forest, past studies found accuracies between 74% and 91%. On the other hand, our ensemble models regularly outperform 98% in every important measure.
3. Clinical Interpretability: Although black-box models have traditionally seen low use, our application of SHAP and CatBoost's internal explanations provides visual, measurable insights into predictions—closing the gap between AI and doctor confidence.

All things considered, this study creates a strong, understandable, and deployable solution that not only satisfies but surpasses the expectations of clinical-grade machine learning systems for maternal healthcare.

5.6 WEB APPLICATION RESULTS AND PRACTICAL IMPACT

The development of a completely operational web application allowing real-time maternal and fetal health risk assessment using machine learning models is a highlight of this study. The application, built on Streamlit, a Python-based open-source framework for fast web development, is a reachable and scalable clinical decision support system. Its goal is to deliver predictive intelligence straight to healthcare professionals, particularly in resource-limited environments where early risk assessment can be life-saving.

The application comprises three major modules:

1. Pregnancy Risk Prediction
2. Fetal Health Prediction
3. Maternal Health Dashboard

Users of the Pregnancy Risk Prediction module enter physiological and clinical characteristics including maternal age, blood glucose levels, systolic blood pressure, body temperature, heart rate and HbA1c. This input is processed by the backend CatBoost model to classify the pregnancy into one of three categories: low, medium or high risk. The outcome is shown right away together with an interpretability layer using SHAP (SHapley Additive Explanations), which improves transparency and clinical confidence by visualizing feature contributions to the prediction.

The Fetal Health Prediction system makes use of cardiotocogram (CTG) data characteristics including fetal heart rate, accelerations, decelerations and heart rate variability indices. This information was used to train a Gradient Boosting Classifier that classified fetal condition into normal, suspect or pathological states. The app lets doctors understand the dependability of the result and give priorities to treatments by showing the prediction with a confidence score.

Apart from forecasting, the web application provides a data visualisation dashboard. Built with Plotly, this module lets users investigate maternal health statistics across Indian states using interactive charts.

1. Bubble charts link unfulfilled maternal care needs with institutional delivery rates.
2. Pie charts depict regional contributions to total institutional deliveries.

With drop-down selectors, sliders and responsive layouts, the design highlights user interaction. Even for those with little technical knowledge, these features improve the tool's usability for healthcare professionals. Such a user-friendly interface democratizes access to AI-powered health analytics and helps real-time informed decision-making.

This web application turns a theoretical machine learning framework into an interactive and deployable tool. It supports the societal value of data-driven health systems and enables proactive maternal-fetal care management by closing the gap between algorithmic predictions and actual healthcare delivery.

CHAPTER 6

CONCLUSION AND FUTURE WORK

6.1 SUMMARY OF WORK

Using a comprehensive machine learning framework, this thesis tackled the early risk assessment of maternal and fetal health. Combining clinical data with advanced ensemble algorithms let the study build an intelligent system that could categorize pregnancies into low, medium, and high-risk categories, so enhancing traditional binary classification systems employed in earlier research.

Particularly in low-resource areas like India, where high maternal and infant mortality rates persist due to delayed risk identification and limited access to specialized treatment, the work began with a comprehensive examination of the problems in maternal healthcare. The study found a clear issue in this setting: the need of a clear, reasonable, scalable machine learning system that can quickly risk stratify to guide prenatal decision-making.

A dataset of more than 6,100 patient records—comprising demographic, physiological and biochemical characteristics—was used to address this. Patterns encouraging model creation were found by exploratory data analysis (EDA). Among the main tasks were data cleaning, multicollinearity study (removal of diastolic BP), outlier treatment and Z-score scaling.

Examined were 15 models in all: decision trees, logistic regression, KNN, SVMs with several kernels, and advanced ensemble learners such XGBoost, LightGBM and CatBoost. Amongst these, CatBoost had the highest accuracy at 98.61% as well as excellent AUC values, recall and accuracy. SHAP plots also helped to show feature contributions by means of systolic BP, temperature, and glucose rising as main risk predictors.

Streamlit was used to create a user-facing dashboard that let healthcare professionals enter patient data and get both forecasts and interpretability visualizations. This showed the possibility of the model for real-world deployment in rural clinics as well as urban hospitals.

The thesis therefore not only provided a technically solid machine learning pipeline but also underlined clinical trust and usability—two crucial qualities for practical integration into maternal care processes.

6.2 KEY FINDINGS

Several important insights and outcomes emerged from this research, reinforcing the power and applicability of machine learning in public health, particularly maternal and prenatal care.

6.2.1 Multiclass Classification Outperforms Binary Systems

Traditional models usually classify patients as either "at risk" or "not at risk." This binary approach, while simple, is clinically insufficient. The three-tier classification developed here allows more granular, actionable insights, enabling healthcare workers to tailor monitoring and intervention strategies according to risk severity.

6.2.2 Ensemble Models Offer Superior Performance

The use of ensemble learners—particularly CatBoost and LightGBM—resulted in significantly higher performance than baseline models. Their ability to handle complex, nonlinear relationships among features made them ideal for health datasets. Ensemble models showed consistent performance across folds suggesting strong generalization and low overfitting.

6.2.3 Feature Importance Aligns with Clinical Knowledge

The top predictors found by the models—systolic blood pressure, body temperature, heart rate and HbA1c—closely correspond with medically acknowledged risk factors for pregnancy problems. This synergy between machine learning results and clinical knowledge increases confidence in the model's outputs.

6.2.4 Interpretability Enhances Clinical Trust

SHAP visualizations' integration let doctors understand why a model predicted certain outcomes. In healthcare environments, where medical professionals have to defend choices to patients and regulatory authorities, this degree of interpretability is absolutely essential.

6.2.5 Balanced Dataset Facilitates Fair Learning

The nearly equal distribution of classes—low, medium, high risk—in the dataset let the model learn without bias toward any class, so enhancing the reliability and ethicality of the results in high-stakes decision-making situations.

These results taken together imply that the suggested risk assessment system is clinically feasible as well as statistically strong, able to assist maternal healthcare professionals in making prompt and informed choices.

6.3 LIMITATIONS

Although the results are positive, the research has some flaws that deserve discussion.

6.3.1 Single-Center Dataset

Though rich and varied, the dataset for this study comes from a small area mostly Indian healthcare center. Unless confirmed on external datasets, the generalizability of the model to other populations, e.g. African, European or American cohorts, may be limited.

6.3.2 Static Dataset (No Time Series or Longitudinal Data)

The dataset consisted of one-time health snapshots of patients. Perhaps time-series data like longitudinal glucose levels or weekly blood pressure readings could improve model accuracy and provide a dynamic view of risk progression. The dataset was designed to exclude this.

6.3.3 Limited Fetal Health Integration

Though the project title and vision address fetal health, the current execution mostly focused on maternal measures. Though not widely included in the last model but CTG data were looked at. Future editions should have fetal indicators like fetal heart rate and variability to offer dual-layer risk assessment.

6.3.4 Real-Time Clinical Validation Not Conducted

The model performs well on test data but it was not validated in a live clinical environment. Practical results could be influenced by user behaviour, hardware constraints or doctor opposition to artificial intelligence acceptance. To assess real-world viability, a pilot study would be required.

6.3.5 Interpretability Is Still Evolving

Although SHAP offers insights, deep interpretability in ensemble models such as CatBoost might still be difficult for non-technical users. Wider adoption depends on simplifying explanation tools for frontline healthcare professionals.

Admitting these constraints helps to honestly frame the study and offers a road map for future development, scalability and practical integration.

6.4 RECOMMENDATIONS FOR FUTURE RESEARCH

Several routes are advised for evolving and enhancing the smart maternal risk assessment framework based on the achievements and failures of this work.

6.4.1 Integration with Electronic Health Record (EHR) Systems

Future work should seek to place the model inside EHR systems or hospital information systems. So reducing manual entry and enhancing workflow efficiency, automated risk predictions based on patient data already being recorded would be made possible.

6.4.2 Inclusion of Real-Time and Wearable Data

Including data from wearable devices such smartwatches tracking heart rate, temperature or physical activity can enhance the dataset and provide continuous monitoring of maternal health. This would allow real-time predictive analysis and alerts for issues under development.

6.4.3 Expansion to Fetal Risk Prediction

Fetal health data—especially CTG recordings—should be added into the pipeline to build a more whole prenatal risk tool. More clinically relevant would be a dual-risk model assessing maternal and fetal conditions.

6.4.4 Explainable AI (XAI) Enhancements

Though SHAP values provide reasonable justifications, future research could investigate more intuitive XAI techniques that translate technical outputs into visuals or clinician-friendly language. This could be simple medical vocabulary describing projections using natural language generation (NLG).

6.4.5 Cross-Cultural and Cross-Geographical Testing

Looking at the model in different cultural and geographic areas including Africa, South America, Europe can help to identify biases and change the model as suitable. This would ensure equality in healthcare outcomes irrespective of area or ethnicity.

6.4.6 Mobile App Development for Field Use

Building a lightweight Android app with TensorFlow Lite or ONNX would allow the model run on tablets or smartphones, therefore allowing frontline workers in remote locations to run the system offline.

These recommendations help to match the proposed system with the UN Sustainable Development Goals (SDG-3): ensuring healthy lives and promoting well-being for all at all ages as well as strengthening its robustness and reach.

REFERENCES

[1] World Health Organization. (2022). Maternal Mortality: Key Facts. https://www.who.int/news-room/fact-sheets/detail/maternal-mortality

[2] Singh, A., & Yadav, R. (2021). Predicting Pregnancy Risk Using Machine Learning Techniques. International Journal of Medical Informatics, 152, 104498.

[3] Jaiswal, M., & Tiwari, S. (2020). Machine Learning Applications in Obstetrics: A Review. Journal of Biomedical Research, 34(4), 245–256.

[4] Chen, T., & Guestrin, C. (2016). XGBoost: A scalable tree boosting system. ACM SIGKDD, 785–794.

[5] Dorogush, A.V., Ershov, V., & Gulin, A. (2018). CatBoost: Gradient boosting with categorical features support. arXiv preprint arXiv:1810.11363.

[6] Ke, G., Meng, Q., Finley, T., et al. (2017). LightGBM: A highly efficient gradient boosting decision tree. NIPS, 3146–3154.

[7] Molnar, C. (2019). Interpretable Machine Learning: A Guide for Making Black Box Models Explainable.

[8] Scikit-learn Developers. (2023). Scikit-learn Documentation. https://scikit-learn.org/stable/documentation.html

[9] Pedregosa, F., Varoquaux, G., Gramfort, A., et al. (2011). Scikit-learn: Machine Learning in Python. Journal of Machine Learning Research, 12, 2825–2830.

[10] Streamlit Inc. (2023). Streamlit Documentation. https://docs.streamlit.io

[11] Lundberg, S.M., & Lee, S.I. (2017). A Unified Approach to Interpreting Model Predictions. NIPS.

[12] Maitree. (2025). Final Project Report - 3rd Sem, Department of Software Engineering, Delhi Technological University.

[13] Maitree. (2025). Research Paper on Pregnancy Risk Prediction, Department of Software Engineering, Delhi Technological University.

[14] Pradhan, A. K., Rout, J. K., Maharana, A. B., Balabantaray, B. K., & Ray, N. K. (2021). A Machine Learning Approach for the Prediction of Fetal Health using CTG. OCIT, IEEE, 239–244.

[15] Ravi, A., Joshi, S. P., Kodipalli, A., & Kamal, S. (2022). Analysis of Maternal Health Risk using Computational Models. SSTEPS, IEEE, 311–315.

[16] Chandrika, V., & Surendran, S. (2022). Incremental Machine Learning Model for Fetal Health Risk Prediction. SMART GENCON, IEEE, 1–6.

[17] Ahammad, M. S., Sinthia, S. A., Hossain, M., Ahmed, M. M., & Ghosh, M. (2024). Empowering Maternal Health in Bangladesh: Advanced Risk Prediction with Machine Learning. ICCCNT, IEEE, 1–6.

[18] Ukrit, M. F., Aluru, L. R., & Chandana, V. (2024). Fetal Health Classification and Visceral Fat Level Prediction using Gradient Boosting and Deep Learning Techniques. ICNWC, IEEE, 1–7.

[19] Ukrit, M. F., Jeyavathana, R. B., Rani, A. L., & Chandana, V. (2024). Maternal Health Risk Prediction with Machine Learning Methods. ICETITE, IEEE, 1–9.