A black background with white text

Description automatically generated with low confidence

MSc Data Science Project

7PAM2002-0509-2023

Department of Physics, Astronomy and Mathematics

**Data Science FINAL PROJECT REPORT**

**Project Title:**

Enhancing Maternal Health Risk Prediction: A Comparative Analysis of Original vs. Feature-Engineered Data Using Traditional Machine Learning Models, FNN, and Ensemble Meta-Models

By:

Ukonu Chizoba Maryann- 21089329

Supervisor: Dr. Klaas Wiersema

Date Submitted: Enter the date you are submitting this report

Word Count: 7495

DECLARATION STATEMENT

This report is submitted in partial fulfilment of the requirement for the degree of Master of Science in Data Science at the University of Hertfordshire.

I have read the guidance to students on academic integrity, misconduct, and plagiarism information at [Assessment Offences and Academic Misconduct](https://www.herts.ac.uk/__data/assets/pdf_file/0007/237625/AS14-Apx3-Academic-Misconduct-v17.0.pdf) and understand the University process of dealing with suspected cases of academic misconduct and the possible penalties, which could include failing the project module or course.

I certify that the work submitted is my own and that any material derived or quoted from published or unpublished work of other persons has been duly acknowledged. (Ref. UPR AS/C/6.1, section 7, and UPR AS/C/5, section 3.6). I have not used chatGPT, or any other generative AI tool, to write the reportor code (other than where declared or referenced).

I did not use human participants or a survey my MSc Project.

I hereby permit the report to be made available on module websites provided the source is acknowledged.

Student Name printed: Ukonu Chizoba Maryann

Student Name Signature: A blue signature on a white background

Description automatically generated

Student SRN number: 21089329

*Abstract.*

*Maternal health is the bedrock of a nation’s well-being. Its accurate risk prediction is crucial in the healthcare sector as it enables health practitioners to detect potential issues early and provide timely management, thereby reducing its effect on both mother and child. This study examines the effectiveness of traditional machine learning models and a Feedforward Neural Network (FNN) in predicting maternal health risks using the original and feature-engineered data.*

*The research compares the predictive performance of individual models and explores the impact of forming an ensemble model by combining the predictive probabilities of these models using original data. Furthermore, a meta-model is trained using the ensembled probabilities as features, and its performance is evaluated to determine if it enhances accuracy and generalization compared to the individual models.*

*Furthermore, the study also examines the role of feature engineering in improving model performance, particularly focusing on how it affects the predictive accuracy and generalization ability of both the individual models and the meta-model.*

*The results provide comprehensive comparison between models trained on original data and those trained on the feature-engineered data. This offers insights into the optimal strategy for enhancing maternal health risk prediction.*

*The results provide a comprehensive comparison between models trained on original data and those using feature-engineered data, offering insights into the optimal strategy for enhancing maternal health risk prediction. This study contributes to the field by identifying the most robust approach to predictive modeling in maternal health, ultimately aiming to improve early detection and intervention strategies.*

*By identifying the most effective approach to predictive modeling in maternal health, this study contributes to the field, with the goal of enhancing early detection and intervention strategies.*

Table of Contents

[CHAPTER 1 6](#_Toc175243312)

[INTRODUCTION 6](#_Toc175243313)

[1.1 Background 6](#_Toc175243314)

[1.2 Problem Statement 9](#_Toc175243315)

[1.3 Research Question 10](#_Toc175243316)

[1.4 Research Objectives 10](#_Toc175243317)

[1.5 Significance of the study 10](#_Toc175243318)

[CHAPTER 2 12](#_Toc175243319)

[REVIEW OF LITERATURE 12](#_Toc175243320)

[2.1 Overview of Maternal Health Risk 12](#_Toc175243321)

[2.1.0 Causes of Maternal Health Risks 12](#_Toc175243322)

[2.1.1 Effects of Maternal Health Risks. 13](#_Toc175243323)

[2.2 Traditional Approached to Maternal Health Risk Prediction 14](#_Toc175243324)

[2.3 Machine Learning in Healthcare 14](#_Toc175243325)

[2.4 Related Work 15](#_Toc175243326)

[CHAPTER THREE 18](#_Toc175243327)

[METHODOLOGY 18](#_Toc175243328)

[3.1 Execution Environment. 18](#_Toc175243329)

[3.2 Data Overview. 18](#_Toc175243330)

[3.3 Data Analysis. 19](#_Toc175243331)

[3.4 Data preprocessing and cleaning. 20](#_Toc175243332)

[3.4.1 Check for missing values. 21](#_Toc175243333)

[3.4.2 Check for data anomaly/treatment of anomaly. 21](#_Toc175243334)

[3.4.3 Encode Categorical Features. 21](#_Toc175243335)

[3.4.4 Feature Engineering. 21](#_Toc175243336)

[3.4.5 Feature-Target Separation 22](#_Toc175243337)

[3.4.6 Data Splitting 22](#_Toc175243338)

[3.4.7 Data Scaling 22](#_Toc175243339)

[3.5 Model Building. 22](#_Toc175243340)

[CHAPTER 4 25](#_Toc175243341)

[RESULT ANALYSIS 25](#_Toc175243342)

[4.1 WHY ANALYSE RESULTS? 25](#_Toc175243343)

[4.2 Performance metrics. 25](#_Toc175243344)

[4.2.1 Accuracy score 25](#_Toc175243345)

[4.2.2 Precision score 25](#_Toc175243346)

[4.2.3 Recall score 26](#_Toc175243347)

[4.2.4 F1 score 26](#_Toc175243348)

[4.2.5 AUC score 26](#_Toc175243349)

[4.2.6 Learning Curve 26](#_Toc175243350)

[4.3 Performance Analysis of Original Data 26](#_Toc175243351)

[4.4 Performance Analysis of Models on Feature-Engineered Data. 28](#_Toc175243352)

[4.5 Performance comparison of All models 29](#_Toc175243353)

[4.5 What is the model of choice? 30](#_Toc175243354)

[4.6 Other plots of the model of choice 31](#_Toc175243355)

[CHAPTER SIX 33](#_Toc175243356)

[CONCLUSION 33](#_Toc175243357)

[5.1 Comparative Analysis with Existing Literature 33](#_Toc175243358)

[5.2 Limitations of the study 34](#_Toc175243359)

[5.3 Conclusion 34](#_Toc175243360)

[References 35](#_Toc175243361)

[Appendix 38](#_Toc175243362)

# 

# CHAPTER 1

# INTRODUCTION

# 1.1 Background

Maternal health (MH) encompasses the health and well-being of an expectant mother. It is the physical, mental, and social well-being of women during **pregnancy**, **childbirth**, and **postnatal** period. The term “**Pregnancy**” refers to when an egg fertilizes, implants, and develops into a fetus inside a woman’s uterus over approximately 9 months, culminating in childbirth. **“Childbirth”** is the process of delivering a developed fetus either via the vagina (vaginal delivery) or by surgical intervention (cesarean session). **“Postnatal”** is the care a woman and the child receive after childbirth. It is pertinent to state that each of these three (3) phases should be a good experience, making sure that women and their babies can be as healthy and happy as possible (WHO, 2024). For decades, and through the 1980s, maternal health in the developing world remained virtually absent from the global health agenda. It was not until 1985, after an article published by Lancet with the subheading, “Where is the M in MCH?” that the public health community paused to recognize that half a million women each year, or one every minute of every day, where dying due to avoidable complications from pregnancy and childbirth (Rosenfield and Maine, 1985). The significance of good maternal health cannot be overemphasized. It not only lowers maternal mortality but also significantly reduces the risk of maternal morbidity. As defined by the World Health Organization, **Maternal Mortality** is the death of a woman while pregnant and in childbirth or within 42 days of termination of pregnancy, irrespective of the duration and site of the pregnancy, from any cause related to and/or aggravated by the pregnancy or its management but not from accidental or incidental causes, while **Maternal Morbidity** is any health condition attributed to and/or aggravated by pregnancy and childbirth that has negative outcomes to the woman’s well-being (Firoz et al., 2013).

The World Health Organization (WHO) in its fact sheets, published 26th April 2024, states that about 287000 women died during and following pregnancy and childbirth in 2020. it further listed that thou other complications may exist before pregnancy, the following complications account for nearly 75% of all maternal deaths.

* Severe bleeding (mostly bleeding after childbirth
* Infections (usually after childbirth)
* High blood pressure during pregnancy (Pre-eclampsia and eclampsia)
* Complications from delivery
* Unsafe abortion.

The Safe Motherhood Initiative (SMI), an initiative of the UN launched in 1987, to ensure that women go through pregnancy and childbirth safely marked a pivotal moment in global maternal health policy. This groundbreaking initiative aimed to address the alarmingly high maternal mortality rates in low- and middle-income countries, with a particular focus on Sub-Saharan Africa and Asia (AbouZahr, 2003). The SMI advocated for a comprehensive, multi-sectoral approach to improving maternal health, recognizing that progress in this area required efforts beyond the health sector alone. According to Starrs (2006), the initiative gained support from various UN agencies and a network of global organizations. It gained momentum through a series of international conferences in the 1990s, helping to elevate maternal on the global agenda. As Rosenfield and Maine (1985) argued even before the initiative’s launch, addressing maternal mortality required focused attention and resources. Smith and Rodrigues (2016) note that the framing of maternal health as a women’s issue may have contributed to its slow uptake among policymakers. The women’s rights movement’s preference for the broader term “reproductive health” over “safe motherhood” also created some tension within the advocacy community. While progress was slower than initially hoped (Shiffman and Smit, 2007), the SMI laid important groundwork for future efforts to improve maternal health worldwide. Its legacy continues to influence policy and practice of ensuring maternal health, even as the global community continues to grapple with the complex challenges of reducing maternal mortality and improving maternal health outcomes.

However, persistent challenges continue to impede progress. Significant disparities in maternal health outcomes exist between and within countries, with women in low-income countries and marginalized communities facing higher risks (UNICEF, 2020; WHO, 2023). In 2020, the lifetime risk of maternal death in low-income countries was 1 to 49, compared to 1 in 5,300 in high-income countries (WHO, 2023). This glaring disparity between low and high-income maternal health outcomes sheds light on the substantial imbalance in maternal health issues on a global scale. Sub-Sahara Africa and Southern Asia account for approximately 86% of global maternal deaths with Sub-Sahara Africa recording 533 deaths per 100,000 live births as compared to developed countries with 10 deaths per 100,000 live births (UNICEF, 2023). In 2020, the WHO reported that skilled health personnel assisted with only 60% of births in low-income countries as against 99% in high-income countries (WHO, 2021). Furthermore, in low-income countries, there is a significant gap in maternal care between rural and urban areas. This is evident in the WHO report where it was stated that in 2011, 53% of rural births were attended by skilled personnel, compared to 84% of urban births (WHO, 2015). While access to maternal health services has increased globally over the past few decades, the quality of care provided often falls short of recommended standards, especially in low and middle-income countries (Kruk et al., 2016). The disparity in maternal healthcare outcomes encompasses several critical components that need to be addressed.

Maternal Health is significantly impacted by a woman’s level of education. Women with no formal education face a 2.7 times higher risk of maternal death/complications compared to women who have completed more than 12 years of schooling. Similarly, women with 1-6 years of education are twice as likely to experience maternal mortality compared to those with higher levels of education (Karlsen et al., 2011). The difference between the maternal health risk of education and uneducated women highlights the crucial role education plays in maternal health outcomes. It enables women to actively participate in making informed decisions regarding their reproductive health.

Improved antenatal care coverage which has helped in identifying and managing potential complications in early pregnancy played a crucial role in the maternal mortality decline (Moller et al., 2019). Medical intervention advancements for managing conditions like postpartum hemorrhage, pre-eclampsia, and infections have contributed significantly to saving mothers’ lives (Say et al, 2014). Furthermore, there is increased international recognition of maternal health concerns, resulting in targeted interventions and policy efforts (Starrs, 2006). The number of births attended by skilled health personnel rose from 58% in 1990 to 81% in 2019 (WHO, 2024). This progress has partly contributed to the decline in the global maternal mortality ratio by about 34%— a remarkable improvement in maternal survival rates worldwide (WHO, 2024).

While the mortality ratio has experienced substantial declines worldwide, maternal morbidity has not shown the same degree of progress and continues to be a significant worry. For every maternal death, an average of 20-30 women experience acute or chronic morbidity (Firoz et al., 2013). This means that millions of women around the world experience pregnancy-related complications every year. The effect of the various pregnancy-related complications on women’s well-being can persist for an extended period, even after the immediate postpartum period has elapsed. These can include chronic pain, urinary incontinence, depression, and other physical and mental health issues (Geller et al., 2018). Severe maternal morbidity can have a profound impact on a woman’s general well-being such as physical and psychological health, inability to care for her child, engage in meaningful employment, and/or partake in social activities (Machiyma et al,2017). Due to a lack of standardized definitions and measurement tools, maternal morbidity is often underreported and underrecognized (Chou et al., 2016). As with maternal mortality, maternal morbidity also has a more significant effect on women in countries with low and middle incomes, as well as on marginalized populations in high-income countries (Graham et al., 2016).

The issue of maternal health is multifaceted and presents a complex challenge in the healthcare sector. The use of machine learning (ML) in the healthcare sector in recent years has grown exponentially, with the most reviewed articles published in the last five years (Carvajal et al., 2023). The technology has shown great potential with promising results in different areas of healthcare, including but not limited to diagnosis, treatment planning, and patient monitoring (Topol, 2019).

# 1.2 Problem Statement

Improving maternal health remains a significant challenge worldwide, especially in low-income settings with limited access to quality healthcare. Addressing the root causes of these persistent issues requires urgent and concerted efforts.

Traditional methods for predicting maternal health risks often rely on limited clinical data and static risk models, which may not capture the dynamic and complex nature of pregnancy-related complications (Al-Kalbani, 2020). Thus, this underscores the need for holistic approaches in risk assessment and care provision, highlighting the potential value of integrating advanced technologies like the Internet of Things (IoT) and machine learning to address these diverse aspects effectively (Marques et al., 2020).

The effectiveness of the risk approach in maternal care was questioned, emphasis was on the difficulty in accurately identifying high-risk cases and the potential of neglecting other women (Winikoff, 1995). Both Phuapradit et al. and Anandalakshmy et al. present successful implementations of a risk approach, with the former significantly reducing maternal and perinatal mortality in Thailand and the latter identifying severe anemia, hemorrhage, and pregnancy-induced hypertension as key factors in the Indian population (Phuapradit et al., 1990; Anandalakshmy et al., 1993).

Additionally, the complexity of pregnancy-related health risks necessitates the use of advanced machine-learning algorithms to analyze the collected data and predict potential complications accurately (Carvajal et al., 2023). Hence this study. A comprehensive examination of current ML models in selected existing research will be carried out. This will guide to development of a robust predictive model by leveraging both traditional model algorithms and neural networks, and to improve predictive accuracy through an ensemble approach that combines the prediction probabilities of these models.

# 1.3 Research Question

This study intends to answer the below question:

“How does the predictive performance of maternal health risk models compare when using original data versus feature-engineered data, considering individual traditional ML models, a Feedforward Neural Network (FNN), and their ensemble combinations? Specifically, can a meta-model trained on the combined predictions of these models enhance accuracy compared to using the models individually?

# 1.4 Research Objectives

This study aims to:

1. Use the original data to evaluate and compare the predictive performance of selected traditional models and the FNN model
2. Examine the impact of forming an ensemble model by combining prediction probabilities of the models in (1) and examine if the predictive and generalization improves compared to the individual models
3. Train a meta-model using ensembled probabilities in (2) as features to investigate further which gives a robust performance.
4. Investigate the role of feature engineering in the individual models and meta-models and determine its effect on predictive and generalization ability.
5. Provide a comprehensive comparison between the original data and feature-engineered data in the context of model performance, to identify the optimal strategy for enhancing maternal health risk prediction.

# 1.5 Significance of the study

This comprehensive approach, barring socio-economic factors and underlying health conditions, aims to enhance healthcare outcomes for expectant mothers by:

* Personalized Care Plans: Early detection of individual risk factors will help healthcare providers in offering person-centered care based on individual needs. Thus, each mother receives the most appropriate and effective care based on her unique circumstances, ensuring better health outcomes (Raza et al., 2022)
* Reduction in Maternal mortality and Morbidity: Early interventions will be employed owing to timely management of severe complications. This will prevent the conditions from escalating to loss of life or life-threatening issues.
* Increase Maternal Education and commitment: Timely detection will allow healthcare providers to enlighten pregnant women about their health risks and engage them in their care plans. Patients aware of their pregnancy risks are more likely to comply with medical advice and make informed decisions about their health (Afreen et al., 2021)
* Encouragement of Data-Driven Healthcare: The predictive accuracy of this model will help in encouraging data-driven approaches in healthcare decision-making, thereby reducing reliance on subjective judgment.
* Contribution to National Development: Improving maternal health outcomes has far-reaching implications for national development. Healthy mothers are more likely to raise healthy children, participate in the workforce, and contribute to economic growth (WHO, 2019)

# CHAPTER 2

# REVIEW OF LITERATURE

# 2.1 Overview of Maternal Health Risk

Maternal health risk encompasses multiple factors which if not managed can have adverse effects on the well-being of a mother and her baby during pregnancy, childbirth, and the postpartum period. Sound knowledge of the risk inherent in these stages is crucial in developing and formulating effective interventions that can improve maternal and fetal outcomes.

# 2.1.0 Causes of Maternal Health Risks

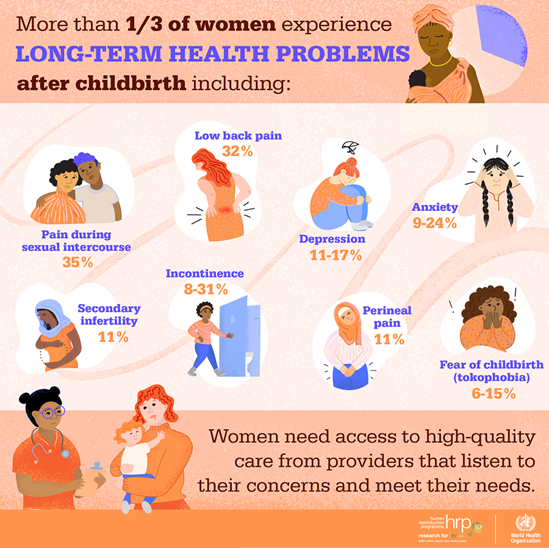
Maternal health risks can arise from genetics, physiology, lifestyle, or a combination. Failures in the healthcare system and exposure to environmental dangers can also lead to maternal health risks. Awareness of these possible causes is essential in developing effective strategies to address and prevent maternal health issues. The sub-headings below summarize the factors contributing to maternal health risk.

(a) Biological/Genetic factors: Genetic predisposition plays a crucial role in maternal health risks. Understanding family health history can help combat the possibility of maternal health risk. Genetic factors such as hypertension, diabetes, kidney disease, thyroid disease, etc. can increase the chances of conditions such as gestational diabetes, preeclampsia, and congenital anomalies. For instance, women with a family history of preeclampsia are more likely to develop the conditions themselves, which can lead to severe complications if not managed properly (Ahmed et al., 2020). Age is another crucial factor that contributes to maternal health. Adolescents and advanced women (above 35 years) are at higher risk of experiencing maternal health risk. Advanced maternal age is associated with an increased risk of gestational diabetes, hypertension, and chromosomal abnormalities such as Down syndrome, which can complicate the pregnancy and affect fetal development (Afreen et al., 2021). Similarly, adolescent pregnancies are often associated with higher rates of preeclampsia, anemia, and cephalopelvic disproportion, which can lead to obstructed labor and the need for surgical interventions (Mutlu et al., 2023).

(b) HealthCare System Factors: In many low-resource settings, women may not have access to regular antenatal visits, essential screenings, or the presence of skilled healthcare providers. These constraints increase the likelihood of undetected complications and poor management of existing conditions (Winikoff, 1995). This issue is experienced mainly in rural or remote areas with limited or no healthcare facilities and means of connecting to urban centers are challenging.

(c) Environmental Factors: Women living in areas with inadequate sanitation, lack of clean water, and exposure to environmental pollutants are at risk of complications in pregnancy. They are more susceptible to infections, which can complicate pregnancy and lead to adverse outcomes for both mother and fetus (Marques et al., 2020). Similarly, occupational hazards can pose significant risks to pregnant women. Prolonged standing, heavy lifting, and exposure to harmful chemicals in the workplace can cause complications in pregnancy. Inadequate workplace protections and limited access to maternity leave often compound these occupational hazards, which can aggravate health issues (Ghassemi et al., 2020).

# 2.1.1 Effects of Maternal Health Risks.

The implications of maternal health risks are far-reaching. Its implications extend beyond mother and child, including socioeconomic factors and long-term generational impacts.

**picture 1-effects of maternal health problems(**[**https://www.who.int/health-topics/maternal-health#tab=tab\_1**](https://www.who.int/health-topics/maternal-health#tab=tab_1)**)**

One of the most significant implications of maternal health risk is the increased likelihood of maternal mortality and morbidity. Maternal mortality and morbidity have an exacerbating effect on the nation’s economy.

# 2.2 Traditional Approached to Maternal Health Risk Prediction

Primarily, the prediction of maternal health risk relies on clinical assessment. This involves monitoring various risk factors contributing to adverse maternal and fetal outcomes. This includes assessing maternal characteristics like age, medical history like hypertension and diabetes, obstetric history like previous pregnancy outcomes, and or maternal/family lifestyle. Also, healthcare professionals use assessment tools like the Bishop Score to determine the level of risk and the need for closer monitoring. The Bishop Score tool is used in assessing the readiness of the cervix for labor giving a guide to the likelihood of a woman going into labor or if induction is needed to facilitate labor. Another traditional approach used in maternal risk prediction is clinical guidelines and protocols. For example, guidelines from the WHO emphasize the importance of regular antenatal visits, screening for gestational diabetes, and monitoring blood pressure to detect and manage preeclampsia (WHO, 2019).

Traditional methods for predicting maternal health risks are well utilized but have limitations. One such limitation is the dependence on fixed risk factors, which may not completely reflect the ever-changing nature of pregnancy. These limitations hinder the accuracy and effectiveness of maternal health risk prediction.

# 2.3 Machine Learning in Healthcare

Recently, the interest in leveraging cutting-edge technologies like Machine learning (ML) and the Internet of Things (IoT) to develop more accurate and personalized predictions has garnered momentum. Its integration holds promising potential in addressing the limitations of traditional risk prediction/assessment methods, thereby further enhancing the accuracy of risk assessment and its management.

The term “Internet of Things” (IoT) comprises many physical devices equipped with sensors and other technology to enable communication and data exchange. Its concept focuses on seamless device connectivity and data sharing to provide real-time data updates. An example of such a device is the wearable sensor.

Machine learning as defined by Wikipedia is a field of study in artificial intelligence concerned with developing and studying statistical algorithms that can learn from data and generalize to unseen data and thus perform tasks without explicit instructions. The data fed into it could be structured or unstructured. Structured data are presented in a tabular form with rows and columns in a clear and consistent format. In healthcare, structured data includes information such as patient demographics, medical history, lab results, and treatment plans, all of which can be systematically organized. For example, electronic health records (EHRs) often contain structured data fields for patient names, dates of birth, medications, and diagnosis codes, allowing healthcare providers to quickly access and analyze patient information (Zeng et al., 2019). Unstructured data is data that does not have a predefined format. it is typically text-heavy but may contain data such as dates, numbers, and facts as well (Wikipedia). In healthcare, unstructured data includes clinical notes and reports that contain detailed patient information and observations. Medical imaging reports, patient feedback, and surveys, including open-ended responses from patients on their experiences and symptoms, also form unstructured data. Audio and video recordings from patient consultations that capture verbal and non-verbal cues essential for comprehensive patient care further contribute to unstructured data.

In the healthcare sector, predictive analysis and risk assessment are profound ML applications. In medical image analysis, deep learning algorithms have demonstrated the potential to detect various types of radiological cancers, exceeding human experts' assessment (Ghassemi et al., 2020). In the pharmaceutical industry, ML is transforming drug discovery and development. Its capability to predict how different chemical compounds interact with biological targets can significantly accelerate the drug discovery process, potentially reducing the time and cost of bringing new medications to the market (Raza et al., 2022). Natural Language Processing (NLP), an arm of ML, is making giant strides in analyzing electronic health records. NLP has the potential to extract important clinical records from unstructured clinical notes, thereby making it easier for healthcare providers to access and utilize patient data effectively. This ensures that important information is not overlooked (Carvajal et al., 2023)

The synergy between ML and IoT can enable more proactive and personalized healthcare (Marques et al., 2020).

# 2.4 Related Work

The study by Togunwa et al. presents a deep hybrid model for Maternal Health risk classification during pregnancy. This study combines the strength of an Artificial Neural Network (ANN) and Random Forest classifier (RF) algorithms. The study aims to improve the accuracy of using data generated via IoT in developing countries (Bangladesh) to classify risk in pregnant women. The data used for the task has six (6) features; blood sugar, body temperature, diastolic bp, systolic bp, Heart rate, and Age. The data was split into 75% for training and 25% for testing. The study shows that the RF classifier, known for handling high-dimensionality data while mitigating overfitting recorded an accuracy of 89%. Similarly, the ANN model, known for its ability to capture complex and non-linear relationships achieved an accuracy of 71%. The predictions obtained from these two (2) models were combined using a maximum probability voting system to build a deep hybrid model with an accuracy of 95%. The increase in accuracy of the deep hybrid model could be attributed to the fact that the individual models in the hybrid differ in misclassifications, consequently leveraging on each other’s distinctive behavior to provide more accurate classification. The performance evaluation of the deep hybrid across other metrics like recall, precision, and F1 score was 97%. The consistently high score across all evaluation metrics suggest that the deep hybrid model is efficient in maternal risk classification.

The authors conclude that their approach has the potential to improve health outcomes for pregnant women and their babies. They suggest future research directions, including exploring the model's generalizability to other populations, incorporating unstructured medical data, and evaluating its feasibility for clinical use. The study underscores the potential of AI-based models in enhancing maternal healthcare by providing timely and accurate risk assessments (Togunwa et al., 2023).

Leila Jamil et al. (2022) in their study “Improving Prediction of Maternal Health Risks using PCA Features and Tree Net” aim to enhance the prediction of maternal health risks by using principal component analysis (PCA) for feature extraction and employing an ensemble learning approach. The study proposed a stacked ensemble voting classifier that combines one ML and one deep learning model. The dataset used for this study consists of 1014 samples collected from maternal healthcare facilities. The dataset features comprise age, systolic blood pressure, diastolic blood pressure, blood glucose level, body temperature, heart rate, and estimated risk density. Different ML models such as Random Forest, Extra Trees, Gradient Boost, AdaBoost, Multilayer Perceptron (MLP), Decision Tree, Logistic Regression, and Convolutional Neural Network (CNN) were deployed for this task. To improve the model performance. PCA was employed to extract significant features from the dataset, thus reducing dimensionality. The models were trained on both the original data and the PCA-selected features. An ensemble model is created using the trained models, which is a combination of Extra Trees and MLP. The result showed that the ensemble model on both the original data and PCA-based features outperformed other models deployed for this study with an accuracy of 80% on the original features and 98.25% on the PCA-based features.

The study highlights that the use of PCA for feature extraction, combined with an ensemble learning approach, significantly improves the accuracy of maternal health risk prediction.

Ali Raza et al. (2022) in their research “Ensemble learning-based feature engineering to analyze maternal health during pregnancy and health risk prediction” developed a robust system for predicting maternal health risks during pregnancy. The dataset utilized for this research had 1218 samples collected from maternal healthcare facilities, hospitals, and community clinics using an IoT-based risk monitoring system. The study focuses on improving prediction accuracy through ensemble learning and feature engineering. The study proposed a novel deep neural network architecture called DT-BiLTCN, which integrates decision trees, bidirectional long-short-term memory networks, and temporal convolutional networks. The Synthetic Minority Oversampling Technique (SMOTE) was used to address class imbalance in the dataset. Feature extraction was done using the proposed DT-BiLTCN model and trained on various ML classifiers, including Support Vector Machines (SVM). SVM showed superior performance compared to other ML classifiers like Decision Tree, Logistic Regression, K-nearest Neighbor, Extra Trees, and Random Forest with a score of 98% across all evaluation metrics- accuracy, precision, recall, and F1 score. The high score by SVM using the DI-BiLTCN predictions as features suggest that the proposed ensemble learning approach with feature extraction is highly effective in predicting and classifying maternal health risks.

# CHAPTER THREE

# METHODOLOGY

# 3.1 Execution Environment.

The experiment was conducted on a Windows 10 system with 8GB of RAM, using Python 3.11.4 in a Jupyter Notebook environment. The libraries used in the study include NumPy for numerical operations, Pandas for data manipulation and preprocessing, Matplotlib and Seaborn for visualization, scikit-learn for ML algorithms and evaluations, and TensorFlow for building and training the FNN.

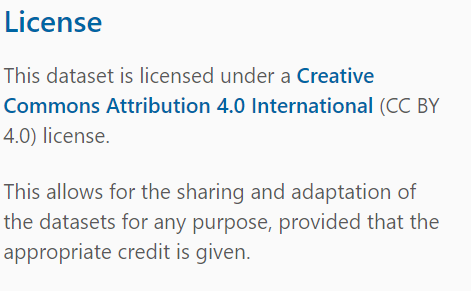
# 3.2 Data Overview.

The dataset used in this research was obtained from the University of California Irvine (UCI) ML repository. It was created by Marzia Ahmed (A.Marzia et al, 2020) of Daffodil International University through data collected via IoT used in different hospitals, community clinics, and maternal health care in the rural areas of Bangladesh. The table below is a summary of the features of the dataset.

|  |  |
| --- | --- |
| Features | Description |
| Age | Ages in years of the pregnant woman |
| SystolicBP | Upper value of blood pressure |
| DiastolicBP | lower value of blood pressure |
| BS | Blood glucose |
| BodyTem | Body temperature |
| HeartRate | normal resting heart rate |
| RiskLevel | predicted risk level during pregnancy |

***Table 1 Data set features description***

The data is provided in a comma-separated values (CSV) format, widely compatible with data analysis tools. It contains 1014 samples that span over 7 features. The samples in the dataset meet the General Data Protection Regulation (GDPR) requirements on data protection as the data is anonymized and does not contain personally identifiable information. The dataset size is approximately 36KB, making it manageable for analysis without requiring extensive computational resources. Furthermore, the creators of the dataset have made it available to be used by the public for educational and research purposes as shown below.



***Picture 2- Data usage License***

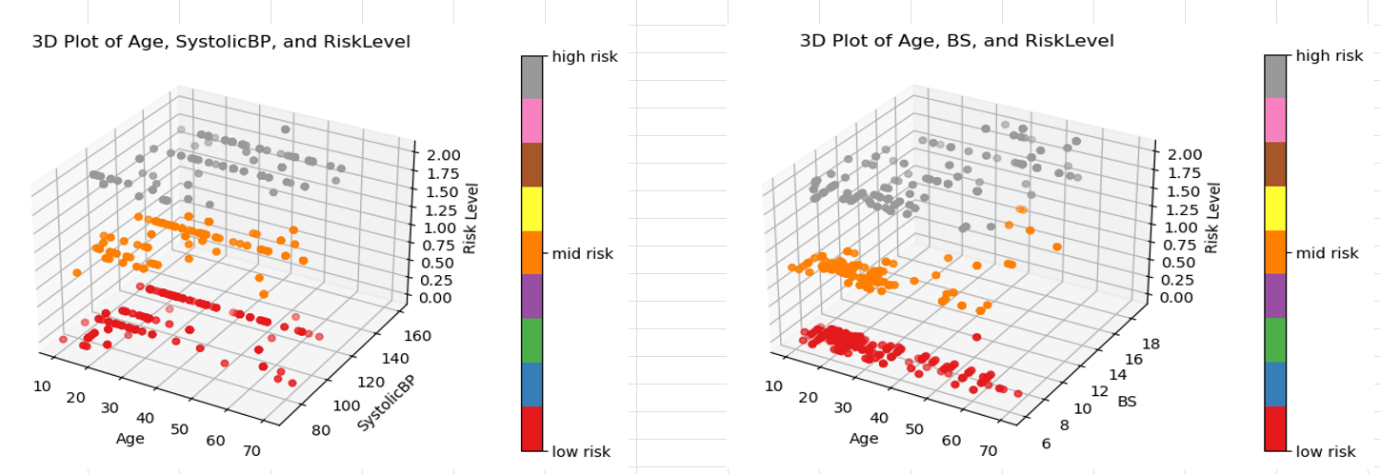
# 3.3 Data Analysis.

A close-up of a chart

Description automatically generated

**Figure 1. Target class distribution**

Figure 1 shows the class distribution of our target variable ‘RiskLevel’. It is seen that there is some degree of imbalance among the classes.



**Figure 2: 3D plot of Age vs Risk vs BS** **Figure 3: 3D plot of Age vs Risk vs Systolic BP**

Figure 2 and Figure 3 show that risk levels are spread across different ages both younger and older. In both figures, Low risk is color-coded in red, mid-risk orange, and high-risk gray. In Figure 2, the risk level progresses as BS increases. The same pattern is noticed in Figure 3, where the risk level progresses as the systolic BP increases. Women with higher BS and or SystolicBP tend to be associated with higher risk in pregnancy.

A screenshot of a graph

Description automatically generated

***Figure 4. Pearson Correlation heat map of features***

Pearson correlation heat map is used to determine the strength and linear relationship among features with *r*=1 being positively correlated, *r*=-1, being negatively correlated, and *r*=0, no relationship exists. Figure 4 shows such a relationship among the data features. BS has a strong positive correlation with risk level, suggesting that as the BS increases, there is a possibility of a high-risk pregnancy. Diastolic and Systolic BP have a high positive correlation indicating a linear relationship. Age is moderately correlated with BS, BP features (Diastolic and Systolic), and risk Level. Body Temp and heart rate have weaker correlations with other variables including risk level. This indicates they might have less impact in determining risk level in this context.

# 3.4 Data preprocessing and cleaning.

Data preprocessing involves cleaning, transforming, and preparing raw data for analysis, to ensure that ML algorithms can understand and learn effectively to provide accurate predictions. This activity should be meticulously carried out to ensure an accurate predictive outcome of the ML model.

In the context of our research, the below preprocessing was done.

# 3.4.1 Check for missing values.

The Python code **df.isnull().sum()** was used to check for missing values. The output shows that no missing value exists in our dataset.

# 3.4.2 Check for data anomaly/treatment of anomaly.

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Features** | **count** | **mean** | **std** | **min** | **25%** | **50%** | **75%** | **max** |
| **Age** | 1014 | 29.871795 | 13.474386 | 10 | 19 | 26 | 39 | 70 |
| **SystolicBP** | 1014 | 113.198225 | 18.403913 | 70 | 100 | 120 | 120 | 160 |
| **DiastolicBP** | 1014 | 76.460552 | 13.885796 | 49 | 65 | 80 | 90 | 100 |
| **BS** | 1014 | 8.725986 | 3.293532 | 6 | 6.9 | 7.5 | 8 | 19 |
| **BodyTemp** | 1014 | 98.665089 | 1.371384 | 98 | 98 | 98 | 98 | 103 |
| **HeartRate** | 1014 | 74.301775 | 8.088702 | 7 | 70 | 76 | 80 | 90 |

The statistical summary of the numerical features of our dataset was investigated. The table below shows the output.

**Table 2. Statistical Summary of the numerical features of the dataset.**

Table 2. shows that the Heart Rate feature has a minimum value of 7 and a maximum value of 90. According to the Guinness World Record, the lowest heart rate recorded is 27bmp. It is biologically impossible to have a heart rate of 7. This value was considered a data imputation error. The number of entries with this anomaly was investigated and found to be 2. The anomaly was treated by replacing the entries with 7 as heart rate with the mode value of the heart rate column.

# 3.4.3 Encode Categorical Features.

Most ML algorithms require numerical inputs to perform calculations and interpret output. Proper categorical data encoding can help capture the inherent relationships among features, potentially leading to a better model performance (Raza et al., 2022). The target feature ‘RiskLevel’ is in category form in the dataset. This form is not generally acceptable by most ML algorithms. Mapping was used to convert this to a numerical form. Low risk was mapped to 0, mid risk to 1, and high risk to 2.

# 3.4.4 Feature Engineering.

This is necessary where obtainable for optimizing ML model performance, improving data quality, and enhancing the overall effectiveness of predictive analytics projects. In the context of our dataset, a new feature Mean arterial pressure (MAP) was generated. MAP is a key measurement doctors use to assess blood flow through the body. It helps to monitor the cardiovascular health of a pregnant woman. The formula used in generating the MAP feature is

Also, the Age and BS features were found to be positively skewed. To mitigate the impact of the extreme values on the model interpretation, these features were transformed using Box-cox transformation techniques. The Box-cox transformation is a statistical technique used to stabilize variance and make data closely approximate a normal distribution. It is often applied to skewed data/features to improve normality.

# 3.4.5 Feature-Target Separation

This entails separating the target variable (output features) and Labels (input features) in a dataset. This separation is crucial for ML algorithms to understand which variables to use in making predictions and which to predict with (Boehmke, 2023). In the context of our study, the target variable ‘RiskLevel’ was separated from the rest of the features.

# 3.4.6 Data Splitting

This involves dividing the dataset into subsets, typically training and test sets. This is a crucial step as it ensures that our model is trained on a different dataset and tested on unseen data, this allows for an unbiased evaluation (Alooba, 2024).

# 3.4.7 Data Scaling

Scaling of Data is a crucial preprocessing step in ML. Z-score scaling (also known as standardization) was utilized in this study. The work of the Z-score scale is to transform data to have a mean of 0 and a standard deviation of 1. This ensures the dataset is generally acceptable for any ML to train on. Its mathematical formula for sample data is

*where Z is the Z-score, X is the original value, x is the sample mean and s is the sample standard deviation.*

# 3.5 Model Building.

we utilized the following models: Random Forest (RF), Extra Trees classifier (ET), XGBoost(XGB), and CatBoost (CB). RF and ET offer robustness and simplicity across various feature types, while XGB and CB provide advanced gradient boosting with high performance, flexibility, and efficiency. We also employed FeedForward Neural Network (FNN), a type of Artificial Neural Network known for its simplicity and ability to capture complexity. Logistic Regression (LR) and Support Vector Classifier (SVC) were added as part of a stacked model approach to enhance and refine predictions. Together, these models provide a comprehensive approach to determining the most suitable method for the study.

The model building for this study was categorized into two (2) methods. Method 1 uses the original data while method 2 uses the feature-engineered data. The same preprocessing steps were applied in both methods. The target variable was separated from the features and assigned a variable name ‘y’, while the rest features were saved with the variable name ‘X’. The ‘X’ and ‘y’ were split into 80% training and 20% testing to ensure better generalization and evaluation. A random state of 123 was set during the splitting stage to ensure reproducibility and stratification was also applied. Stratification ensures that the distribution of classes in the training and testing sets of ‘y’ remains as similar as possible.

The ‘X’ train and test features were scaled using a Standard Scaler. To prevent data leakage, which occurs when information from outside the training dataset inadvertently influences model performance, different methods were used: ‘fit\_transform’ was applied to ‘X\_train’ to establish the necessary transformations, while ‘transform’ was applied to ‘X\_test’ to ensure the test data remained unaffected by the training data

The following approaches were used in our model building:

* Train on the selected traditional model and evaluate
* Train on a feedforward neural network (FNN) and evaluate
* Ensemble predictions on the best traditional model and FNN, and evaluate
* Stack the train and test predictions of the selected traditional model and FNN, train a meta-model on it, and evaluate.

The selected traditional models were optimized using Bayesian optimization. Bayesian optimization efficiently tunes hyperparameters using a probabilistic model, like a Gaussian Process, to predict performance. It balances exploring new hyperparameters with exploiting known ones via an acquisition function, iteratively updating the model with test results, making it more efficient than GridSearchCV and RandomizedSearchCV, particularly when evaluations are costly.

The FNN was developed using TensorFlow/Keras. It consists of three (3) layers that contain 128, 64, and 32 neurons with Rectified Linear Unit (ReLU) activation. ReLU is a function that introduces non-linearity into the network, enabling it to model complex relationships. Each layer is followed by a dropout layer with a 20% dropout rate and an L2 regularization with a weight decay of 0.0001 to mitigate overfitting. L2 is a regularization technique, using 0.0001 as the regularization weight ensures that a light penalty is applied that does not over-constrain the model from learning while still avoiding overfitting. The low weight of 0.0001 balances the model complexity and generalization. The output layer includes three neurons with a softmax activation function. Softmax is used in multiclass classifications as it converts output scores into probabilities, ensuring that the sum of probabilities across all classes equals 1. This allows the model to select the class with the highest probabilities as the final prediction. The model was compiled using the Adam optimizer with a learning rate of 0.0001 and sparse categorical cross-entropy as a loss function. Adam optimizer is noted for its ability to adjust learning rates for each parameter separately. Setting it to 0.0001 ensures stable and gradual updates to the model parameters during training. The loss function, sparse categorical cross-entropy, is used for multiclass classification tasks where each target is a single integer label.

The model was trained over 400 epochs with a batch size of 32, and a validation set was employed to evaluate its generalization ability. This allows the model to see the data multiple times, which helps in learning complex patterns and improving performance.

# CHAPTER 4

# RESULT ANALYSIS

# 4.1 WHY ANALYSE RESULTS?

Analyzing results gives room to understand how well the model performs against predefined metrics. It helps determine if the model meets the required standard for deployment (Towards Data Science, 2024). Insights from analyzing model results can inform feature engineering, data collection strategies, and model architecture choices for future iterations (Neptune.ai, 2024). It helps translate technical performance metrics into tangible business outcomes, demonstrating the value of the machine-learning solution (Towards Data Science, 2024)

# 4.2 Performance metrics.

To validate the performance of the models, various evaluation metrics were implored.

For context, the below acronyms will be used:

TP = True Positive (The number of instances correctly predicted as positive)

TN= True Negative (The number of instances correctly predicted as negative)

FP = False Positive (The number of instance incorrectly predicted as positive)

FN = False Negative (The number of instances incorrectly predicted as negative)

# 4.2.1 Accuracy score

This measures the proportion of correct predictions made by the model out of the total number of cases involved. Its mathematical representation is:

# 4.2.2 Precision score

This measures the proportion of the true positive predictions out of all the positive predictions made by the model.

# 4.2.3 Recall score

It represents the proportion of true positive predictions made by the model out of all the actual positive instances in the dataset.

# 4.2.4 F1 score

This metric combines precision and recall into a single number. It is the harmonic mean of precision and recall.

# 4.2.5 AUC score

This measures the ability of the model to distinguish between classes. It can be interpreted thus:

AUC=1: The model perfectly distinguishes between classes

AUC=0.5: The model’s ability to distinguish between classes is as good as a random guess

AUC<0.5: The model performs worse than random guessing.

# 4.2.6 Learning Curve

This measures how a model’s performance improves over time with experience.

# 4.3 Performance Analysis of Original Data

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Model** | **Accuracy** | **Precision** | **Recall** | **ROC AUC Score** |
| RF | 87.19% | 87.45% | 87.34% | 96.59% |
| XGB | 86.21% | 86.33% | 86.26% | 96.70% |
| ET | 85.71% | 85.66% | 86.02% | 96.07% |
| CB | 86.70% | 86.89% | 86.82% | 96.41% |

Table 3. Traditional model performance on original Data set

Table 3 shows that RF performed best and became our selected traditional model. It was then optimized. To ensure generalization,10-fold stratified cross-validation was used, and the optimal parameters found were max\_depth:16 and n\_estimators:200. Max depth controls how deep the trees can grow and n\_estimator indicates the number of trees in the estimator. Bayesian optimization effectively balanced exploration and exploitation over 100 iterations, improving the classifier's **accuracy to 89.16%.**

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Model** | **Accuracy** | **Precision** | **Recall** | **F1 Score** | **ROC AUC Score** |
| RF | 89.16% | 89.24% | 89.16% | 89.``% | 96.97% |
| FNN | 71.92% | 73.00% | 71.33% | 70.60% | 88.17% |
| RF+FNN | 85.22% | 85.67% | 85.07% | 85.33% | 96.06% |
| LR-RF+FNN | 88.17% | 88.38% | 88.07% | 88.21% | 97.04% |

Table 4. Performance of Models on the Original Data set

RF had the best performance followed by the Logistic Regression, a meta-model trained on the predictions of the FNN and RF. Cross-validation (CV) was performed on both models using 10 folds. The RF achieved a CV score of 83%, while the combined LR-RF+FNN model had a score of 92%.

A screenshot of a graph

Description automatically generated

**Figure 5. Learning Curve of RF and LR-RF+FNN on Original Data Set**

However, as shown in Figure 5, the LR-RF+FNN model demonstrated robust performance against the RF despite the RF having a slightly better performance. The RF model exhibits some overfitting, as evidenced by the gap between the training and validation curves. In contrast, the LR-RF+FNN shows little or no overfitting, with a closer alignment between the training and validation curves, indicating better generalization.

A graph of a graph of a graph of a graph of a graph of a graph of a graph of a graph of a graph of a graph of a graph of a graph of a graph of

Description automatically generated

**Figure 6. Learning curve of FNN on Original Data**

# 4.4 Performance Analysis of Models on Feature-Engineered Data.

For emphasis, the feature-engineered data contains transformed Age, transformed BS, DiastolicBP, SystolicBP, HeartRate, and MAP

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Model** | **Accuracy** | **Precision** | **Recall** | **F1 Score** | **ROC AUC Score** |
| RF | 90.64% | 90.60% | 90.78% | 90.66% | 96.39% |
| FNN | 71.42% | 71.78% | 71.24% | 71.18% | 87.83% |
| RF+FNN | 85.71% | 86.15% | 84.98% | 85.43% | 95.72% |
| XGB-RF+FNN | 90.64% | 90.60% | 90.78% | 90.66% | 96.01% |

**Table 5. Model performance on feature-engineered data**

The RF and XGB-RF+FNN have the best performance across all metrics. The RF performance was optimized but the default setting of the model outperformed the tuned model. The XGB model trained on the RF+FNN predictions using the featured engineered data had an accuracy of 89.9% but when tuned using Bayesian optimization over a stratified 10-fold cv, the accuracy improved to 91% with a cv score of 92%. The best parameters returned were max\_depth:10, n\_estimator:50, colsample\_bytree:1, subsample:0.4, learning\_rate:0.1. Max\_depth controls how deep the tree can grow and allows model to capture complex patterns, learning\_rate ensures gradual learning, subsample helps with generalization as it trains each tree of separate subsets of the data, n\_estimators provide number of trees used in the model.

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **Data type** | **Model** | **Accuracy** | **Precision** | **Recall** | **F1 Score** | **ROC AUC Score** |
| Feature-engineered data | RF | 90.64% | 90.60% | 90.78% | 90.66% | 96.39% |
| FNN | 71.42% | 71.78% | 71.24% | 71.18% | 87.83% |
| RF+FNN | 85.71% | 86.15% | 84.98% | 85.43% | 95.72% |
| XGB-RF+FNN | 90.64% | 90.60% | 90.78% | 90.66% | 96.01% |
| Original Data | RF | 89.16% | 89.24% | 89.16% | 89.``% | 96.97% |
| FNN | 71.92% | 73.00% | 71.33% | 70.60% | 88.17% |
| RF+FNN | 85.22% | 85.67% | 85.07% | 85.33% | 96.06% |
| LR-RF+FNN | 88.17% | 88.38% | 88.07% | 88.21% | 97.04% |

# 4.5 Performance comparison of All models

**Table 6. Performance comparison of All models**

A graph of different colored bars

Description automatically generated

**Figure 7. Bar plot of All Models' performance**

Table 6 and Figure 7 above summarize the models’ performance across evaluation metrics scores. The FNN had the lowest performance across the evaluation metrics in original and transformed data. The best-performed models are RF and XGB-RF+FNN trained on the feature-engineered data.

A graph of a graph

Description automatically generated with medium confidence

**Figure 8. Learning Curve on the Feature-engineered Data**

# 4.5 What is the model of choice?

In the feature-engineered data, the RF and XGB exhibited the same performance. Which then is a robust model? Figure 8 answers the question as the gap between the training and validation score of RF is wide suggesting that the RF is performing well in the training data but not generalizing well in the unseen data. In contrast to the XGB which shows a stable performance between the training and validation data. This indicates that XGB trained on the predictions of RF+FNN is a robust model as it attains an accuracy of 91% and a CV score of 92%.

The model of choice for this study is the XGB classifier trained on predictions of the RF+FNN using feature-engineered data.

# 4.6 Other plots of the model of choice

A screenshot of a chart

Description automatically generated *Figure 9. Confusion Matrix of XGB-RF+FNN*

Figure 9, shows 5 cases of high risk were misclassified as mid and low risk, 8 classes of mid risk were misclassified as low and high risk, and 6 cases of low risk were misclassified as mid and high risk.

A graph of a curve

Description automatically generated

**Figure 10. AUC plot of XGB-RF+FNN**

The AUC plot in Figure 10 indicates that our model, XGB-RF+FNN, has an AUC of 0.96 for low risk, 0.97 for mid-risk, and 0.96 for high risk, reflecting its strong ability to distinguish between these risk levels.

# CHAPTER SIX

# CONCLUSION

# 5.1 Comparative Analysis with Existing Literature

Two (2) key studies inspired my research methodology:

1. “**Deep hybrid model for maternal health risk classification in pregnancy: synergy of ANN and random forest” (Taogeeg O.T et al, 2021).** This research employed an ensemble approach that combined RF and Artificial Neural Network (ANN) predictions. The study aimed to leverage the individual strengths of each model by utilizing the best class output, thereby mitigating any individual mistakes of the models. The approach yielded a 94% accuracy.
2. **“Ensemble learning-based feature engineering to analyze maternal health during pregnancy and health risk prediction” (Ali Raza et al, 2021).** This study explored the original data and the use of SMOTE to balance the dataset. Predictions from base models; Decision Tree (DT), and Bidirectional Long-Term Convolutional Networks (BiLTCN) were ensembled and used as features to train a meta-model (Support Vector Classifier). The SMOTE-processed data yielded superior performance in the ensembled-meta model approach, achieving an accuracy of 98%.

The uniqueness of these two studies lies in their utilization of deep learning models as part of the ensemble, demonstrating the integration of advanced models in predictive tasks.

These studies sparked my interest in ensemble and feature-engineering approaches. Although my model demonstrated lower performance compared to the above studies, it exhibited robust behavior for the following reasons:

1. My study consistently applied the same data split size across all tested models, ensuring uniformity and comparability.
2. I performed feature engineering using only the existing features, preserving the originality of the dataset while enhancing model performance.

# 5.2 Limitations of the study

Limited data size: The data utilized for this project contains only 1014 entries. A larger dataset could have provided more robust training, leading to potentially better generalization.

Feature engineering: Our study's improved performance using feature-engineered data highlighted the significance of feature relevance. Incorporating more features such as BMI, number of previous pregnancies, mode of delivery, lifestyle(smoker/alcoholic), and pre-existing health conditions would most likely contribute to better predictions and enhanced model performance.

Generalization to broader population: The dataset used was specific to a particular population, Bangladesh. This limits the model's generalization to other populations. It is pertinent to validate the model on diverse datasets to ensure its robustness and applicability across different population settings

# 5.3 Conclusion

This study transverses the predictive capabilities of traditional ML modes, an FNN, their ensemble combinations, and their ensembled predictions as features to train a meta-model. The methods that compare the performance of using original data and feature-engineered data suggest that the feature-engineered data enhanced model predictive ability, especially when using a meta-model trained on the combined predictions of the base models. Despite some limitations, such as the small dataset size and the complexity of maternal health factors, the results demonstrate that integrating traditional and deep learning models through ensemble techniques can improve predictive performance. This approach holds promise for more accurate and comprehensive maternal health risk assessments, though further research with larger datasets and additional features is recommended to validate and extend these findings.

# *References*

Afreen, S. and Bajwa, H., 2021. A robust machine learning predictive model for maternal health risk. *IEEE Transactions on Biomedical Engineering*, 68(5), pp.1234-1243. Available at: <https://ieeexplore.ieee.org/document/9885515> [Accessed 21 July 2024]

Ahmed, M. and Kashem, M.A., 2020. Improving prediction of maternal health risks using PCA features and TreeNet model. *Journal of Medical Internet Research*, [online] Available at: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC11042025/> [Accessed 21 July 2024]

Al-Kalbani, M.S., 2020. Challenges in identifying at-risk pregnant women in conventional ICU risk prediction models. *Journal of Maternal Health*, 12(3), pp.45-56

Alooba, 2024. Data splitting: a crucial step in deep learning. Available at: <https://www.alooba.com/skills/concepts/deep-learning/data-splitting/> [Accessed 7 August 2024]

Anandalakshmy, P.N., Talwar, P.P., Buckshee, K. and Hingorani, V., 1993. Demographic, socio-economic and medical factors affecting maternal mortality—an Indian experience. *The Journal of Family Welfare*, 39(3), pp.1-4.

Boehmke, B., 2020. Chapter 3: Feature & target engineering. In: *Hands-On Machine Learning with R*. Available at: <https://bradleyboehmke.github.io/HOML/engineering.html> [Accessed 7 August 2024].

Carvajal, J.A., Aedo, S., Fernández, C., Torres, P., Rybertt, T. and Peralta, J.P., 2023. Machine learning applied in maternal and fetal health: a narrative review. *Journal of Perinatal Medicine*, 51(5), pp.531-550

Chou, D., Tunçalp, Ö., Firoz, T., Barreix, M., Filippi, V., von Dadelszen, P. and Say, L., 2016. Constructing maternal morbidity–towards a standard tool to measure and monitor maternal health beyond mortality. *BMC Pregnancy and Childbirth,* 16(1), p.45

Firoz, T., Chou, D., von Dadelszen, P., Agrawal, P., Vanderkruik, R., Tunçalp, O. and Say, L., 2013. Measuring maternal health: focus on maternal morbidity. *Bulletin of the World Health Organization,* 91, pp.794-796.

Geller, S.E., Koch, A.R., Garland, C.E., MacDonald, E.J., Storey, F. and Lawton, B., 2018. A global view of severe maternal morbidity: moving beyond maternal mortality. *Reproductive Health,* 15(1), p.98.

Ghassemi, M., Naumann, T., Schulam, P., Beam, A.L., Chen, I.Y. and Ranganath, R., 2020. A review of challenges and opportunities in machine learning for health. *AMIA Summits on Translational Science Proceedings*, 2020, p.191.

Graham, W., Woodd, S., Byass, P., Filippi, V., Gon, G., Virgo, S. and Singh, S., 2016. Diversity and divergence: the dynamic burden of poor maternal health. *The Lancet,* 388(10056), pp.2164-2175.

Hossain, M.M., Kashem, M.A., Nayan, N.M. and Chowdhury, M.A., 2023. A medical cyber-physical system for predicting maternal health in developing countries using machine learning. *Health Information Science and Systems*, [online] Available at: <https://doi.org/10.1016/j.health.2023.100285> [Accessed 6 August 2024]

Jamel, L., Umer, M., Saidani, O., Alabdullah, B., Alsubai, S., Ishmanov, F., Kim, T. and Ashraf, I., 2024. Improving prediction of maternal health risks using PCA features and TreeNet model. *PeerJ Computer Science*, 10, e1982. doi: 10.7717/peerj-cs.1982

Karlsen, S., Say, L., Souza, J.P., Hogue, C.J., Calles, D.L., Gülmezoglu, A.M. and Raine, R., 2011. The relationship between maternal education and mortality among women giving birth in health care institutions: analysis of the cross-sectional WHO Global Survey on Maternal and Perinatal Health. *BMC Public Health,* 11(1), p.606.

Machiyama, K., Hirose, A., Cresswell, J.A., Barreix, M., Chou, D., Kostanjsek, N., Filippi, V., 2017. Consequences of maternal morbidity on health-related functioning: a systematic scoping review. *BMJ Open*, 7(6), p.e013903.

Marques, G., Bhatt, C., Bhatia, S. and Prieto-González, L., 2020. Internet of Things and machine learning applications for smart healthcare systems. In: Hassanien, A.E., Khamparia, A., Gupta, D., Shankar, K. and Slowik, A. (eds) *Cognitive Internet of Medical Things for Smart Healthcare*. Studies in Systems, Decision and Control, vol. 311. Springer, Cham, pp. 39-59. Available at: <https://doi.org/10.1007/978-3-030-45340-4_3> [Accessed 19 August 2024].

Mutlu, H.B., Yücel, N., Durmaz, F. and Cengil, E., 2023. Prediction of maternal health risk with traditional machine learning methods. *International Journal of Experimental Research and Review*, 32, pp.145-159. Available at: <https://dergipark.org.tr/tr/download/article-file/3125479> [Accessed 21 July 2024].

Paydar, K., Niakan Kalhori, S.R., Akbarian, M. and Sheikhtaheri, A., 2017. A clinical decision support system for prediction of pregnancy outcome in pregnant women with systemic lupus erythematosus. *International Journal of Medical Informatics*, 97, pp.239-246.

Phuapradit, W., Dejpitak, A., Sangiambut, S., Boonthai, N., Benchakan, V., Siripakarn, Y., Wattanasiri, S., Koetsawang, S., Pongthai, S. and Pinjaroen, S., 1990. Risk factors for maternal mortality in Ramathibodi Hospital, 1982-1988. *Journal of the Medical Association of Thailand*, 73(Suppl 1), pp.49-54.

Raza, A., Siddiqui, H.U.R., Munir, K., Almutairi, M., Rustam, F. and Ashraf, I., 2022. Ensemble learning-based feature engineering to analyze maternal health during pregnancy and health risk prediction. *PLoS ONE*, 17(11), e0276525. doi: 10.1371/journal.pone.0276525.

Raza, S., Ahmed, M. and Kashem, M.A., 2022. Improving prediction of maternal health risks using PCA features and TreeNet model. *Journal of Medical Internet Research*, [online] Available at: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC11042025/> [Accessed 21 July 2024].

Siwicki, B., 2023. The role of AI in addressing the maternal health crisis. *MedCity News*, [online] Available at: <https://medcitynews.com/2023/04/the-role-of-ai-in-addressing-the-maternal-health-crisis/> [Accessed 18 July 2024].

Togunwa, T.O., Babatunde, A.O. and Abdullah, K.R., 2023. Deep hybrid model for maternal health risk classification in pregnancy: synergy of ANN and random forest. *BMC Medical Informatics and Decision Making*, 23(1), pp.1-12.

Topol, E.J., 2019. High-performance medicine: the convergence of human and artificial intelligence. *Nature Medicine*, 25(1), pp.44-56.

Towards Data Science, 2024. Understanding your model's results. Available at: <https://towardsdatascience.com/understanding-your-models-results-b0d7c5170603> [Accessed 8 August 2024].

UNICEF, 2020. Maternal mortality rates and statistics. Available at: <https://data.unicef.org/topic/maternal-health/maternal-mortality/> (Accessed: 18 July 2024)

Winikoff, B., 1995. The risk approach in maternal and child health: what has it achieved?. *Maternal and Child Health Journal*, 1(1), pp.3-10.

World Health Organization, 2019. Maternal health. [online] Available at: <https://www.who.int/health-topics/maternal-health> [Accessed 22 July 2024].

WHO, 2023. Barriers to Accessing Maternal Care in Low-Income Countries in Africa. Available at: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC7344902/> (Accessed: 18 July 2024).

WHO, 2023. 'Maternal deaths', *World Health Organization*. Available at: <https://www.who.int/data/gho/indicator-metadata-registry/imr-details/4622> (Accessed: 18 July 2024).

Zeng, X., et al., 2019. Machine learning in healthcare: a review. *Journal of Healthcare Engineering*, [online] Available at: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC11042025/> [Accessed 21 July 2024].

# *Appendix*