CHAPTER 1

INTRODUCTION

**Research Question**

**"How do tree-based ensemble models, feedforward neural networks, and their combined predictions compare in terms of performance for predicting maternal health risk, and which approach provides the most accurate and reliable risk assessments?"**

INTRODUCTION

* 1. Background.

Maternal health 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 World Health Organization 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 World Health Organization 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. Educated women are more likely to access antenatal care, skilled birth attendance, and postnatal care services (Karlsen et al, 2011). It also enables women to actively participate in making informed decisions regarding their reproductive health.

The launch of the Safe Motherhood Initiative made maternal health record significant improvement in recent years, but substantial challenges still lie ahead. The number of births attended by skilled health personnel has risen 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%. This is 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).

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 issue of maternal health is multifaceted and presents a complex challenge in the healthcare sector. The use of machine learning (ML) in recent years in the healthcare sector has grown exponentially. 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). The methodology of machine learning focuses on developing algorithms and statistical models that permit computers to execute assignments without explicit instructions, relying instead on patterns and inference from data. The use of machine learning in pregnancy diseases and complications is relatively recent, with the most reviewed articles published in the last five years (Carvajal et al., 2023). It has been on the rise in the background of maternal and fetal health, offering promising solutions for early diagnosis, screening, and risk determination of pregnancy-related complications (Carvajal et al., 2023). Machine learning has proven to be a powerful branch of artificial intelligence with robust technology that can uncover complicated patterns, correlations, and subtle risk factors that traditional analytical approaches may not be able to discern, potentially leading to timely interventions and improved results. The large quantity of data generated during pregnancy, childbirth, and the postpartum period, combined with the complex nature of timely interventions, makes maternal health ideal for machine learning applications (Paydar et al., 2017).

* 1. 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. While this issue might seem light, suffice to state that maternal health comprises a complex interplay of physiological, psychological, and social factors that significantly impact both the mother and the developing fetus throughout pregnancy, childbirth, and the postpartum period (WHO, 2019). The complex nature of maternal health comprises various aspects, including:

* Prenatal Care and Nutrition
* Management of pre-existing medical conditions
* Prevention and early detection of pregnancy-related complications
* Mental health and emotional well-being
* Access to quality healthcare services
* Socioeconomic factors influencing health outcomes
* Postpartum care and support

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 Internet of Things (IoT) is a ground technology that is changing the way maternal health risks are monitored. IoT devices enable continuous, real-time data gathering and analysis, which enables healthcare providers to monitor health indicators in pregnant women closely. The IoT devices can monitor vital signs such as heart rate, blood pressure, fetal movements, and temperature, providing a comprehensive and up-to-date picture of a pregnant woman’s health status (Ahmed et al., 2020). This real-time monitoring allows for the early detection of potential complications, enabling healthcare providers to intervene proactively and provide personalized care (Mutlu et al., 2023).

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. Research Question

This study intends to answer the below question:

“Can an ensemble model, which combines prediction probabilities from traditional machine learning models and deep learning models, enhance the accuracy of maternal health risk prediction when used to train a meta-model compared to using individual models alone?”

* 1. Research Objectives

Maternal health is the foundation of a nation’s development as when neglected impacts both the mother and the fetus, thereby influencing the broader society. Hence the importance of its timely intervention cannot be over-emphasized.

The objectives of this study are to.

1. Understand Features

* Use predictions from the traditional model and deep learning model to identify and analyze features that are most important in classifying risk in pregnancy

1. Form ensemble Model

* Combine predictions from the models in (a) to form an ensemble model to understand if their combined strength gives an improved predictive performance.

1. Train a meta-model

* Use the prediction probabilities from (a) to form a new data frame and train a meta-model. This is aimed at further enhancing the accuracy of maternal health risk prediction.

1. Compare predictive outcomes

* Compare results obtained from the traditional model, deep learning model, ensemble model, and meta-model to determine which approach provides the most accurate and reliable predictions for maternal health risks
  1. 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 detection of high-risk pregnancy will imply timely management of severe complications that might lead to mortality or morbidity. Hence interventions will be employed that will prevent the conditions from escalating to loss of life or life-threatening issues.
* Increase Maternal Education and commitment: Timely detection of maternal risk allows 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. Health mothers are more likely to raise healthy children, participate in the workforce and contribute to economic growth (WHO, 2019)

Chapter 2

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 Categories of Maternal Health Risk.

Maternal health Risks can be categorized as follows:

1. Pregnancy-Related Conditions: Some circumstances spring up during pregnancy. Such circumstances could be gestational diabetes, preeclampsia, and eclampsia and they inadvertently pose a significant risk to both the mother and the fetus

* Gestational Diabetes: This can lead to macrosomia (a newborn baby who is much larger than average baby weight), birth injuries, and neonatal hypoglycemia (a plasma glucose level of less than 30mg/dL (1.65 mmol/L) in the first 24 hours of life and less than 45 mg/dL (2.5 mmol/L) thereafter (Mutlu et al., 2023).
* Preeclampsia: This is characterized by high blood pressure and organ damage and leads to eclampsia if not timely managed (Anandalakshmy et al., 1993)
* Eclampsia: A severe complication of preeclampsia which is characterized by seizure (WHO, 2019)

1. Pre-existing conditions: Women who have the below-existing conditions before pregnancy will potentially be at risk.

* Diabetes: This can lead to macrosomia, birth defects, and increased risk of cesarean delivery (Ahmed et al, 2020).
* Hypertension: Hypertensive women are at risk of developing preeclampsia, placenta abruption, and fetal growth restriction (Phuapradit et al., 1990).
* Heart and Kidney Disease: Existing heart disease before pregnancy can worsen during pregnancy, leading to complications for both mother and fetus (Carvajal et al., 2023). Likewise, women associated with kidney disease are at risk of preeclampsia and preterm birth (Raza et al., 2022).

1. Obstetric Risks: This could be hemorrhage before or after delivery, preterm labor, or placenta abnormalities. Placenta abnormalities such as placenta previa (placenta covering the cervix) or abruption(premature separation of the placenta from the uterus) can cause bleeding before or after delivery (Phuapradit et al., 1990). Excessive bleeding after delivery is one of the leading causes of maternal mortality (Anandalakshmy et al., 1993). Preterm Labor can lead to premature birth, and this gives rise to various complications for the newborn including respiratory distress syndrome, intraventricular hemorrhage, and long-term developmental issues (Carvajal et al., 2023).
2. Infectious Risks: Infections such as HIV, malaria, urinary tract infections, and sexually transmitted infections can adversely affect pregnancy outcomes (WHO, 2019). Malaria is associated with maternal anemia, stillbirth, and low birth weight (Marques et al., 2020). HIV if not carefully managed during pregnancy and delivery, can be transmitted to the fetus (WHO, 2019).
3. Psychological Risk: Depression and Anxiety also constitute a risk in pregnancy. Anxiety is associated with an increased risk of preterm birth and low birth weight (Marques et al., 2020) while depression can affect maternal-fetal bonding and increase the risk of postpartum depression (Ghasemi et al., 2020).

2.1.2 Causes of Maternal Health Risks

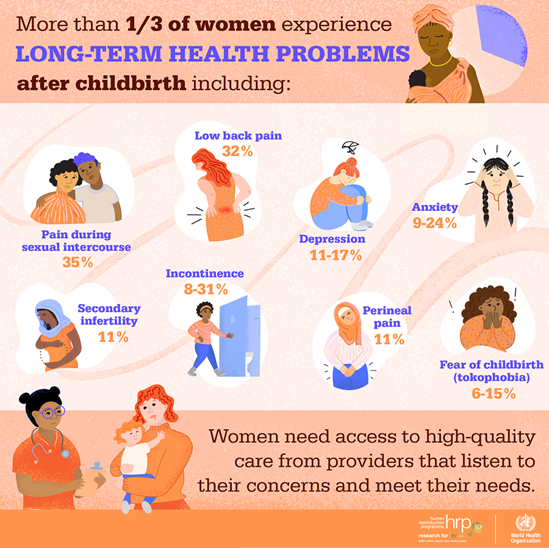
Maternal health risks can arise from genetics, physiology, lifestyle, or a combination. Failures in the healthcare system and exposure to environmental dangers can also lead to maternal health risks. Awareness of these possible causes is essential in developing effective strategies to address and prevent 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. The presence of skilled healthcare providers can significantly reduce maternal and neonatal mortality rates (WHO, 2019) as they have trained to recognize and address potential complications such as bleeding, obstructed labor, and infections.

(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).

2.1.4 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. 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. In the Healthcare sector, maternal health risk will mean an increased need for specialized care, prolonged hospital stays, and potential long-term health issues for both mother and child, thereby leading to higher healthcare costs (WHO, 2019). Maternal health risks also have a psychological impact as high-risk pregnancies and their complications can lead to increased anxiety, stress, and depression for both the mother and family (Ghassemi et al., 2020). Maternal health risks also have intergenerational implications. Recent research has shown that maternal health status during pregnancy can influence gene expression in offspring through epigenetic changes, potentially affecting long-term health outcomes (Ahmed et al. 2020). Socially, maternal health risks can lead to reduced women's participation in education and the workforce leading to reduced family income, especially in communities where maternal health outcomes are poor (Winikoff, 1995).

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 lifestyle/family lifestyle. For instance, advanced maternal age (over 35 years) and adolescent pregnancies are both associated with higher risks of complications such as preeclampsia, gestational diabetes, and preterm birth (Mutlu et al., 2023). 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 World Health Organization (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 and the Internet of Things (IoT) to develop more accurate and personalized predictions has garnered momentum. These technologies when integrated can revolutionize the healthcare system by providing more accurate and personalized risk assessment using real-time data generated by IoT devices, thereby leading to improved healthcare outcomes. IoT are wearable sensors capable of collecting real-time data. 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 a wide range of physical devices equipped with sensors and other technology to enable communication and data exchange over the Internet. 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 data presented in a tabular form having rows and columns with 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, patients feedback and surveys which include open-ended responses from patients on their experiences and symptoms also form unstructured data. Audio and video recordings from patient consultations which captures verbal and non-verbal cues essential for comprehensive patient care further contributes to unstructured data.

In the healthcare sector, one of the profound applications of machine learning is predictive analysis and risk assessment. ML algorithms can analyze large datasets and extract information with remarkable accuracy, thereby facilitating early intervention and personalized care plans by healthcare providers. In medical image analysis, deep learning algorithms have demonstrated the potential to detect various types of cancers from radiological, exceeding assessment by human experts (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 Random Forest 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 TreeNet” 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 RandomForest, ExtraTrees, GradientBoost, AdaBoost, Multilayer Perceptron (MLP), DecisionTree, LogisticRegression, 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 ExtraTrees 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 Neighbour, 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 suggests that the proposed ensemble learning approach with feature extraction is highly effective in predicting and classifying maternal health risks.

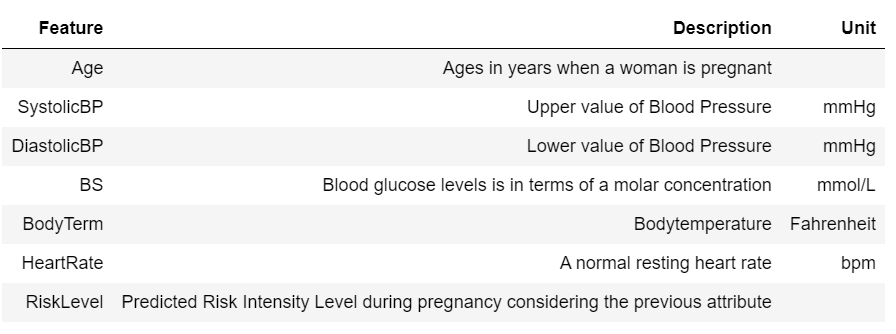
CHAPTER THREE

Methodology.

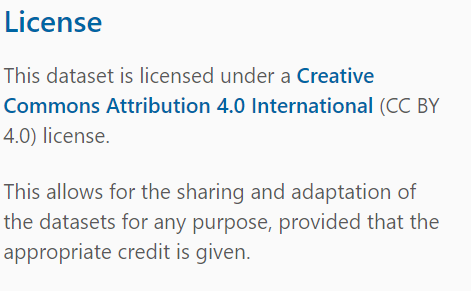
3.1. Data Overview.

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.

**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 1- Data usage License

3.2. 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.

A diagram of a graph

Description automatically generated

Figure 2 Figure 3

Figure 2 and Figure 3 show that risk levels are spread across different ages both younger and older.

3.2 Data preprocessing and cleaning.

Data preprocessing is an intricate part of preparing the data for ML algorithm usage. It involves cleaning, transforming, and preparing raw data for analysis, to ensure that ML algorithms can understand and learn effectively to provide accurate predictions. Data preprocessing, as part of data preparation for ML use, ensures that the quality and integrity of the data are preserved (Hossain et al., 2023). 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.

* Check for missing values
* Check for data anomaly/treatment of anomaly
* Encode categorical features
* Feature engineering
* Feature-Target Separation
* Data Splitting
* Data Scaling

3.2.1. 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.

3.2.2. Check for data anomaly/treatment of anomaly.

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.**

A table with numbers and a few letters

Description automatically generated with medium confidence

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.

3.2.3. 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). In the dataset, the target feature ‘RiskLevel’ is in category form. 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. In this form, the ML algorithm can understand, interpret, and perform calculations yielding robust performance.

3.2.4 Feature Engineering.

Feature engineering 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

3.2.5. Feature-Target Separation

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

3.2.6. Data Splitting.

This involves dividing the dataset into subsets, typically including the training, validation, 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).

3.2.7. 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 that the dataset is generally acceptable for any ML to train on as some ML are sensitive to the scale of input features. The mathematical formula for Z-score for a 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.*

Important.

* 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.

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

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

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

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

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

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), e013903.

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

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].

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.

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.

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. <https://doi.org/10.1007/978-3-030-45340-4_3>

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

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].

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].

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.

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.

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, 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].

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].

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

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.

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].

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.

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.

Marques, G., Bhatt, C., Bhatia, S. and Prieto-González, L., 2020. Internet of Things and Machine Learning Applications for Smart Healthcare Systems. In: Cognitive Internet of Medical Things for Smart Healthcare. Springer, Cham.

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), p.1-12. certified

Jamel, L., Umer, M., Saidani, O., Alabduallah, 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. certified

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. certified

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], certified

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

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].

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]. certified