#### ASSESMENT OF MATERNAL HEALTH CARE STATUS AMONG AFGHAN REFUGEES USING MACHINE LEARNING BASED IN DISTRICT HARIPUR

**By**

# Aleesha Azam

# Hira Bibi

**CS-21407 and CS-21413**



**Department of Computer Science**

**Govt Akhter Nawaz Khan (S) Degree Collage**

**KTS Haripur**

**Khyber Pakhtunkhwa, Pakistan**

**June,2025**

#### ASSESMENT OF MATERNAL HEALTH CARE STATUS AMONG AFGHAN REFUGEES USING MACHINE LEARNING BASED IN DISTRICT HARIPUR

**By**

**Aleesha Azam**

**Hira Bibi**

**CS-21407 and CS-21413**



**Supervised By**

**Department of Computer Science**

**Govt Akhter Nawaz Khan (S) Degree Collage**

**KTS Haripur**

**Khyber Pakhtunkhwa, Pakistan**

# June,2025

#### ASSESMENT OF MATERNAL HEALTH CARE STATUS AMONG AFGHAN REFUGEES USING MACHINE LEARNING BASED IN DISTRICT HARIPUR

**By**

**Aleesha Azam**

**Hira Bibi**

**CS-21407 and CS-21413**



A thesis submitted in partial fulfilment of the requirements for the degree of Computer Science BSCS

**Department of Computer Science**

**Govt Akhter Nawaz Khan (S) Degree Collage**

**KTS Haripur**

**Khyber Pakhtunkhwa, Pakistan**

# June,2025

#### ASSESMENT OF MATERNAL HEALTH CARE STATUS AMONG AFGHAN REFUGEES USING MACHINE LEARNING BASED IN DISTRICT HARIPUR

**By**

**Aleesha Azam**

**Hira Bibi**

**CS-21407 and CS-21413**



## COMMITTEE

Mr. Shahid Gillani Mr. Shahid Gillani FYP Supervisor FYP Coordinator

Dr. Dr. Abid Ali External Examiner Head of Department

**Department of Computer Science**

**Govt Akhter Nawaz Khan (S) Degree Collage**

**KTS, Haripur**

**Khyber Pakhtunkhwa, Pakistan**

DECLARATION

I hereby declare that the thesis titled “Assessment of Maternal Healthcare Status Among Afghan Refugees Using Machine Learning Based in District Haripur” submitted to the University of Haripur, in partial fulfillment of the requirements for the degree of BSCS, is my original work and has not been submitted previously by me or any other person for any degree or diploma at any other university or institution.

I further declare that all sources of information and data have been acknowledged, and that this research has been conducted in accordance with the ethical standards and guidelines of the institution.

This work is the result of my own investigation and efforts, and all contributions from other individuals or sources have been duly cited and acknowledged.

**Department of Computer Science**

**Govt Akhter Nawaz Khan (S) Degree Collage**

**KTS, Haripur**

**Khyber Pakhtunkhwa, Pakistan**

# DEDICATION

In the name of Allah, the most mercifal, the most gracious, we humbly dedicate this thesis to express our deepest gratitude and appreciation. Without the guidance, blessing and support of Almighty, this endeavor would not have been possible.

First and foremost, we extend our heartfelt gratitude to our beloved parents and esteemed supervisors, who have been unwavering pillars of strength throughout this journey. This unwavering love, encouragement, and sacrifices have been the driving force behind our pursuit of knowledge and personal growth.

To our parents we owe an immeasurable debt of gratitude for nurturing and shaping us into the individuals we are today. Their unwavering faith in our abilities and their constant support have been instrumental in our academic and personal achievements. May Allah shower his blessing upon them for their boundless love and care.

Additionally, we express our profound gratitude to the faculty and management of GANKS Degree College KTS and the University of Haripur. Their dedication to academic excellence and their commitment to fostering a nurturing learning environment have been invaluable in shaping our academic pursuits.

This final year project holds a special place in our hearts as it symbolizes the culmination of our four-year journey in the field of BSCS. It stands as a testament to our commitment to learning and our passion for contributing to the advancement of knowledge. In this endeavor, we pay tribute to our instructors and supervisors, whose expertise and guidance have enriched our understanding.

We thank, pray for and promise them to do whatever is possible in our powers to comfort them and promote their good mission for the noble cause of spreading education. They have served with their best efforts: may Almighty Allah bless them.

# ACKNOWLEDGEMENTS

First and foremost, I am profoundly thankful to **Almighty Allah** for granting me the strength, patience, and perseverance to complete this research work successfully. I would like to express my heartfelt gratitude to **Dr. Abid**, Head of the Department, for his continuous support, encouragement, and guidance throughout my academic journey. His vision, leadership, and commitment to academic excellence have always inspired me.

I am deeply indebted to my respected research supervisor, **Mr. Shahid Gillani**, for his invaluable guidance, unwavering support, and constructive feedback during every stage of my research. His mentorship, critical insight, and kind encouragement played a crucial role in the completion of this thesis.

I also extend my thanks to all faculty members, fellow students, and participants who contributed to this study either directly or indirectly.

Lastly, I would like to express my deepest appreciation to my family for their unconditional love, patience, and moral support throughout this academic endeavor.

**CERTIFICATION FROM THE EXAMINERS**

This thesis entitled (full title of the thesis) presents a bonafide record of original research work carried out by (full name of the student) in partial fulfillment of the degree of (full name of the degree), The University of Haripur.

We found the work satisfactory for the award of the degree if other requirements are met. The Open defense and Viva Voce was held on (date\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_).

**External Examiner I**

(Name and Affiliation) (Signature)

**External Examiner II**

(Name and Affiliation) (Signature)

**Chairperson**

(Name and Affiliation) (Signature)

**ASSESMENT OF MATERNAL HEALTH CARE STATUS AMONG AFGHAN REFUGEES USING MACHINE LEARNING BASED IN DISTRICT HARIPUR**

BY

**ALEESHA AZAM**

**HIRA BIBI**

**ABSTRACT**

This study investigates the maternal healthcare status of Afghan refugees in District Haripur, Pakistan, utilizing machine learning techniques to analyze and predict healthcare outcomes. Afghan refugees, facing prolonged displacement due to decades of conflict, encounter significant barriers to accessing quality maternal healthcare, including economic hardship, cultural restrictions, and inadequate healthcare infrastructure. The research employs a quantitative cross-sectional design, collecting data through structured questionnaires from Afghan refugee women aged 18 to 55 who have experienced at least one pregnancy. Key variables include socio-economic factors, healthcare access, and maternal health service utilization.

Machine learning models, specifically Random Forest Classifier and Logistic Regression, are applied to identify patterns and predict maternal healthcare status. The findings reveal that low income, limited education, and distance from healthcare facilities are critical determinants of inadequate maternal care. The Random Forest model achieved an accuracy of approximately 85%, effectively predicting which women are at higher risk of not receiving adequate healthcare.

The study underscores the need for targeted interventions to improve maternal health outcomes among Afghan refugees, including enhancing access to healthcare services, promoting health education, and increasing the availability of female healthcare providers. Policy implications suggest the establishment of local health units, financial support for low-income families, and the integration of machine learning tools in healthcare planning. Future research directions include expanding the sample size, exploring advanced machine learning techniques, and incorporating qualitative methods to gain deeper insights into the cultural factors affecting maternal healthcare access. This research contributes valuable data-driven evidence to inform policies aimed at improving maternal health for vulnerable refugee populations.

**Table of Contents**

[**COMMITTEE iv**](#_Toc197328971)

[**DECLARATION v**](#_Toc197328972)

[**DEDICATION vi**](#_Toc197328973)

[**ACKNOWLEDGEMENTS vii**](#_Toc197328974)

[**ABSTRACT ix**](#_Toc197328975)

[**ACRONYMS xii**](#_Toc197328976)

[**LIST OF FIGURES xiii**](#_Toc197328977)

[**LIST OF TABLES xiv**](#_Toc197328978)

[**CHAPTER- 1 INTRODUCTION 1**](#_Toc197328979)

[**1.1 Background 1**](#_Toc197328980)

[**1.2 Problem Statement 4**](#_Toc197328981)

[**1.3 Research Questions 5**](#_Toc197328982)

[**1.4 Research Objectives 5**](#_Toc197328983)

[**1.5 Significance of Study 5**](#_Toc197328984)

[**1.1 Scope and Limitations 7**](#_Toc197328986)

[**CHAPTER -2 REVIEW OF LITERATURE 8**](#_Toc197328987)

[**2.1 Global overview of refugees’ maternal health 8**](#_Toc197328988)

[**2.2 Maternal healthcare challenges in Pakistan 9**](#_Toc197328989)

[**2.3 Use of Machine Learning in healthcare predication 9**](#_Toc197328990)

[**2.4 Gap analysis and rationale for your study 10**](#_Toc197328991)

[**CHAPTER -3 MATERIAL AND METHODS 11**](#_Toc197328992)

[**3.1 Study Design 11**](#_Toc197328993)

[**3.2 Study Area 12**](#_Toc197328994)

[**3.3 Inclusion and Exclusion Criteria 12**](#_Toc197328995)

[**3.4 Data Collection Tools 12**](#_Toc197328996)

[**3.5 Variables and Features 13**](#_Toc197328997)

[**3.6 Data Preprocessing 13**](#_Toc197328998)

[**3.7 Machine Learning Models Used 14**](#_Toc197328999)

[**3.7.1 Logistic Regression 14**](#_Toc197329000)

[**3.7.2 Random Forest Classifier 14**](#_Toc197329001)

[**3.9 Statistical and ML Tools 16**](#_Toc197329002)

[**3.10 Ethical Considerations 17**](#_Toc197329003)

[**CHAPTER -4 RESULTS AND ANALYSIS 18**](#_Toc197329004)

[**4.1 Descriptive Statistics 18**](#_Toc197329005)

[**4.2 Feature Correlation 20**](#_Toc197329006)

[**4.3 Machine Learning Model Training 21**](#_Toc197329007)

[**4.4 Interpretation of key Findings 28**](#_Toc197329008)

[**CHAPTER -5 DISCUSSION 29**](#_Toc197329009)

[**5.1 Comparison with literature 29**](#_Toc197329010)

[**5.2 Interpretation of barriers to maternal healthcare 30**](#_Toc197329011)

[**5.3 Role of socioeconomic factors 32**](#_Toc197329012)

[**5.4 Contribution of ML model to healthcare planning 33**](#_Toc197329013)

[**CHAPTER -6 CONCLUSION AND RECOMMENDATIONS 33**](#_Toc197329014)

[**6.1 Summary of Findings 33**](#_Toc197329015)

[**6.2 Policy Implications 34**](#_Toc197329016)

[**6.3 Recommendation for stakeholders 35**](#_Toc197329017)

[**6.4 Future Research Directions 35**](#_Toc197329018)

[**REFFERANCES 36**](#_Toc197329019)

# ACRONYMS

ML Machine Learning

WHO World Health Organization

SVM Support Vector Machine

CDSS Clinical Decision Support Systems

UNHCR United Nations High Commissioner for Refugees

LR Logistic Regression

RFC Random Forest Classify

SPSS Statistical Package for the Social Sciences

NN Neural Network

CT- Scan Computed Tomography

MRI Magnetic Resonance Imaging

ANC Antenatal Care

PNC Postnatal Care

# LIST OF FIGURES

**Figure 1 Methodology steps……………………………………………………………..15**

**Figure 2 Heatmap………………………………………………………………..20**

**Figure 3 Confusion matrix (LR)………………………………………………..23**

**Figure 4 Confusion matrix (RFC)………………………………………………24**

**Figure 4 vertical bar chart………………………………………………………25**

**Figure 5 Line graph………………………………………………………………26**

# 

**LIST OF TABLES**

**Table 1 df.Describes………………………………………………………………18**

**Table 2 Descriptive statistics……………………………………………………………..19**

**Table 3 Heatmap………………………………………………………………………….21**

**Table 4 confusion matrix…………………………………………………………………28**

# CHAPTER- 1 INTRODUCTION

## Background

Afghan refugees constitute one of the world’s largest and longest-lasting communities of displaced persons, the result of decades of conflict, political instability, and economic underdevelopment. The first wave of displacement started with the Soviet invasion of Afghanistan in 1979, which followed by civil war, Taliban rule, and ongoing security that prompted millions of people to flee. Due to its close geographical location, cultural, and economic ties, Pakistan has been a major host nation of Afghan refugees for more than 40 years. As of 2015, there were about 1. 3 million registered Afghan refugees and an additional 1. 75 million refugees that did not have official documentation (United Nations High Commissioner for Refugees, 203) Mahmood et al. 2023 The majority of refugees live mainly in provinces such as Khyber Pakhtunkhwa, most of whom reside in pre-existing communities or in camps that are sparsely developed (see also Mahamood et al., 2023; United Nations High Commissioner for Refugees, 2015). Where they live, most of them experience poverty and limited social and economic opportunities. Access to health care has also been heavily restricted for many. These problems have been especially problematic in women as insecurity around mothers and children is compounded by other factors: food insecurity and other complications due to the increase in mother-to-child ratio.

Maternal health care describes services and support available to pregnant women during pregnancy and the delivery of their babies as well as during postpartum period. Maternal health care contributes significantly to the reduction of maternal and neonatal morbidity and mortality by providing timely access to prenatal care, birth attendants, emergency obstetric services and postnatal support. Intense maternal health services are needed to identify and manage complications, enhance the quality of birth and promote maternal and infant health. However, access to maternal health care is restricted in many developing countries and among vulnerable populations such as refugees because of socioeconomic, cultural, and infrastructure barriers. “Maternal health improvements constitute not only a health issue but also a key step towards the achievement of wider development targets” (WHO, 2016). Therefore, increasing access to and quality of maternal health care is essential to safeguard the health of mothers and their children, particularly in the most marginalized communities (Organization, 2016).

Currently, the maternal health care status of Afghan refugee women is a public health concern, in particular host countries such as Pakistan. The reproductive health trajectories and the types of maternal health care services rendered to Afghan refugees are stigmatized and lack adequate coverage (for instance, nonprofessional antenatal care, skilled birth attendance, postpartum follow-up). These problems and under representational behaviours further contribute to poor maternal health care provision. For example, due to cultural restrictions on women's mobility in the host country the prevalence of home births with no professional supervision, minimal prenatal visits to refugees, and increased risk of complications related to pregnancy including maternal morbidity and infant mortality, are also highlighted. According to the report M. Shakir and colleagues (2018), "Refugee women in Pakistan have inadequate maternal healthcare access due to poor living conditions, insufficient health education, and insufficient facility-level resources. To improve the maternal outcomes and better health care of Afghan mothers and their children, targeted interventions and other inclusive health policies are required. " (Esteva et al., 2017).

Maternal Health and birth outcomes are important public health indicators, especially in developing countries where maternal and infant mortality is high. Poor maternal health can indirectly affect the child’s development, and providing access to quality maternal health services can help reduce preventable deaths and complications in pregnancy and birth. Refugee populations are particularly vulnerable due to displacement, limited access to healthcare facilities, cultural differences and socioeconomic instability. One of the largest and most marginalized refugee groups in Pakistan are Afghan refugees. Afghanistan has been hosting Afghan refugees for decades, many of whom live in camps and informal settlements (Smith, 2019). District Haripur in Khyber Pakhtunkhwa is one of the regions with a significant Afghan refugee population. A number of initiatives have been taken by government and non-governmental organizations to assess and improve the maternal health status of Afghan refugee women in this region. However, with the recent advancement in data analytics and computer science techniques, machine learning (ML) has emerged as an effective tool for better understanding healthcare data and predicting health outcomes. By discovering hidden patterns and correlations in large datasets, ML models such as Logistic Regression (LR) and Random Forest Classifier (RFC) can help to improve the need assessment, enhance service delivery and contribute to policy decision making. This study aims to apply machine learning algorithms to evaluate and predict maternal health status of Afghan refugees in District Haripur (Alemi et al., 2014).

Machine learning (ML) has become an important tool for medicine in many applications, enhancing diagnosis, treatment planning, and patient outcomes. Basically, Machine learning is the field of artificial intelligence that utilizes algorithms and statistical models to analyze and predict patterns of data. It has applications in healthcare where it allows automation of processes, optimization of decision-making skills, and improved predictive analytics. The use of machine learning in medicine started at the development of diagnostic tools that could recognize patterns in medical images such as X-rays, Magnetic Resonance Imaging (MRI), and Computed Tomography (CT scans). In time the capability of ML model for disease risk assessment, patient prognosis, and personalized treatment plans has also grown tremendously. Models of ML such as Random Forest, Support Vector Machine (SVM), and Neural Networks (NN) have been implemented for tasks such as early disease detection, drug discovery, genetic data analysis, and predicting response to treatments among others (Mahesh, 2020).

A significant illustration of machine learning in the medical field is its utilization in radiology. Machine learning algorithms can be trained to detect anomalies in medical images, often achieving higher accuracy than human radiologists. For example, research has shown that deep learning models can accurately identify lung cancer from CT scans, matching the performance of seasoned radiologists (Esteva et al., 2019). Furthermore, machine learning is employed in genomics to process extensive genetic data for disease forecasting and drug development, aiding in the discovery of genetic markers associated with various conditions, including cancers and rare genetic disorders (Daghistani & Alshammari, 2020).

Moreover, machine learning plays a vital role in clinical decision support systems (CDSS), which aid healthcare professionals in making decisions based on evidence. These systems utilize ML algorithms to analyze patient information, medical histories, laboratory results, and treatment alternatives, thus offering healthcare providers recommendations that can improve the precision of diagnoses and treatment strategies (Worabo et al., 2024). As healthcare systems increasingly adopt digital technologies and produce more data, the significance of machine learning in medicine is expected to grow, enhancing the efficiency and effectiveness of healthcare services. Nevertheless, issues such as data privacy, algorithm transparency, and ethical considerations remain critical areas that require further investigation and advancement.

## Problem Statement

The issue your research intends to tackle is the inadequate maternal healthcare conditions faced by Afghan refugees living in District Haripur, Pakistan. This situation is further complicated by limited access to healthcare services, cultural obstacles, and insufficient data for making informed decisions. Afghan refugees, who often reside in overcrowded camps or informal settlements, encounter significant barriers to obtaining quality maternal care due to financial limitations, language differences, and a lack of healthcare infrastructure. These issues are intensified by the absence of comprehensive data on maternal health outcomes specific to refugee groups, hindering the ability to accurately evaluate the severity of the problem or devise effective interventions. By utilizing machine learning techniques on existing healthcare data, your research aims to pinpoint critical factors affecting maternal health outcomes, forecast potential risks, and evaluate the overall state of maternal healthcare among Afghan refugees in this area. Machine learning models, including Random Forest and Logistic Regression, can reveal underlying patterns, anticipate complications during pregnancy, and provide insights that could guide healthcare policies and enhance resource distribution, ultimately striving to decrease maternal morbidity and mortality within this at-risk population.

## 1.3 Research Questions

1. What are the main obstacles that Afghan refugees encounter when trying to access maternal health care?
2. How do socio-economic conditions affect the use of maternal health services by Afghan refugees?
3. Is it possible for machine learning algorithms to accurately forecast the probability of Afghan refugees pursuing maternal health care?
4. What strategies can be recommended based on research findings to enhance access to maternal health care?

## 1.4 Research Objectives

* To assess the status of maternal health care among Afghan refugees in District Haripur.
* To identify the key barriers affecting access to maternal health services.
* To develop a machine learning model that predicts the likelihood of Afghan refugees utilizing maternal health services based on various socio-economic and cultural factors.
* To provide actionable insights for policymakers to improve maternal health care access for Afghan refugees.

## 1.5 Significance of Study

The importance of our research, "Evaluation of maternal healthcare conditions among Afghan refugees employing machine learning techniques in District Haripur," is underscored by its capacity to fill significant gaps in maternal healthcare for a vulnerable and marginalized demographic. Afghan refugees residing in Pakistan encounter a distinct array of challenges that impede their access to sufficient maternal care, encompassing cultural obstacles, financial limitations, and inadequate healthcare infrastructure. By evaluating the maternal healthcare status of Afghan refugees through machine learning methodologies, this study can provide critical insights into the healthcare requirements of this population and contribute to the enhancement of maternal health outcomes. A primary contribution of this research is its capacity to produce data-driven evidence that can assist in identifying pivotal factors affecting maternal health within refugee communities. Conventional healthcare evaluations may fail to adequately reflect the unique situations of refugees; however, by implementing machine learning algorithms such as Random Forest Classifier (RFC) and Logistic Regression (LR), this study can reveal obscured patterns and correlations in the data that may otherwise be neglected. These models have the potential to forecast possible health risks and complications during gestation, thereby enabling healthcare practitioners to intervene at an earlier stage and prioritize individuals at risk for improved care.

Furthermore, this research has the potential to significantly influence local health policy formulation and the distribution of resources within District Haripur. By identifying regions exhibiting the highest maternal health vulnerabilities, it can serve as a foundational tool for the implementation of specific interventions and the strategic allocation of resources aimed at enhancing healthcare services. For example, the research may uncover areas where refugee women encounter barriers to accessing prenatal care, thereby necessitating the development of targeted outreach initiatives or mobile healthcare units to effectively assist those in greatest need.

Furthermore, your investigation presents a groundbreaking methodology by integrating machine learning within the domain of refugee healthcare, thereby offering a cutting-edge instrument for data analysis and informed decision-making. Given that refugee populations frequently experience a lack of representation in health data, this research will not only deepen the comprehension of maternal healthcare challenges but will also augment the expanding corpus of studies that employ advanced technologies to enhance health outcomes in settings characterized by limited resources.

## Ultimately, the importance of our investigation resides in its capacity to facilitate meaningful advancements in maternal healthcare for Afghan refugees residing in District Haripur, thereby diminishing maternal mortality and morbidity rates, and promoting superior health outcomes for this at-risk demographic. Through the utilization of machine learning methodologies, your research has the potential to provide empirically-supported recommendations for more efficient and focused healthcare strategies.

## Scope and Limitations

The parameters of our investigation, titled "Evaluation of Maternal Healthcare Status Among Afghan Refugees Utilizing Machine Learning in District Haripur," involve a comprehensive assessment of maternal health conditions and the identification of critical determinants that affect healthcare outcomes within the Afghan refugee demographic. This research emphasizes the implementation of machine learning methodologies, including Random Forest and Logistic Regression models, to scrutinize healthcare datasets, discern patterns, and forecast potential risks and complications associated with maternal health. The objective is to furnish actionable insights for healthcare practitioners, policymakers, and organizations operating in refugee environments, thereby facilitating more precise interventions and enhanced resource allocation in District Haripur. By focusing on a distinct refugee community, the study enhances the comprehension of the particular obstacles they encounter in obtaining quality maternal care and provides data-informed recommendations to ameliorate their health outcomes.

This research has several limitations. Firstly, it relies on the availability and quality of data, which may be incomplete or inconsistent, particularly in refugee contexts where data collection poses significant challenges. The data may also exhibit bias or lack comprehensiveness, which can hinder the accuracy of machine learning models. Furthermore, as the study is limited to District Haripur, the results may not be entirely applicable to other refugee populations or regions with varying socio-economic or healthcare conditions. Additionally, while machine learning models can detect correlations and risks, they do not fully consider all socio-cultural, political, and economic factors that affect access to and outcomes of maternal healthcare. Lastly, the research does not directly tackle the implementation of interventions derived from the findings, as it primarily emphasizes assessment and prediction, leaving the practical application and effects of those interventions for future exploration. Nevertheless, the study offers valuable insights into the maternal healthcare issues encountered by Afghan refugees in a specific area, laying the groundwork for subsequent research and enhancements in refugee healthcare services.

CHAPTER -2 REVIEW OF LITERATURE

## 2.1 Global overview of refugees’ maternal health

Maternal health among refugees is a critical global public health issue, as these populations often suffer from significantly elevated rates of maternal morbidity and mortality. Refugees, especially those residing in camps or informal settlements, encounter various obstacles in obtaining quality maternal healthcare, such as inadequate healthcare infrastructure, cultural and linguistic barriers, insufficient prenatal and postnatal care, and financial limitations. Additionally, many refugee women face an increased risk of complications during pregnancy and childbirth due to pre-existing health issues, malnutrition, and exposure to violence and trauma. The inconsistent availability of skilled birth attendants and essential health services, coupled with overcrowded living conditions, further heightens the risks associated with maternal health. Research indicates that refugee women are more prone to experiencing delays or inadequate care, resulting in preventable complications like hemorrhage, infections, and preeclampsia, which contribute to elevated maternal mortality rates within these communities (Yeo et al., 2023).

International organizations and governments have acknowledged the necessity for focused maternal health initiatives within refugee contexts. It is essential to develop programs that enhance access to maternal healthcare, offer culturally appropriate services, and guarantee the presence of qualified healthcare professionals. Nevertheless, the execution of these initiatives frequently faces obstacles due to the intricate and evolving circumstances of refugee environments, alongside political and resource constraints. In spite of these difficulties, there is an increasing awareness of the significance of incorporating maternal healthcare into humanitarian efforts to mitigate the dangers linked to pregnancy and childbirth among refugee communities. (Crisp, 2010; Rojas & Saurina, 2020)

## 2.2 Maternal healthcare challenges in Pakistan

Maternal healthcare in Pakistan encounters significant obstacles, especially in rural and underserved regions, where the availability of quality healthcare services is severely restricted. Although there have been some advancements in maternal health over recent decades, Pakistan continues to have one of the highest maternal mortality rates in South Asia. Contributing factors include inadequate healthcare infrastructure, a lack of trained healthcare professionals, and insufficient prenatal and postnatal care. Numerous women in rural areas, including Afghan refugees, find it difficult to access skilled birth attendants and emergency obstetric services, resulting in a high rate of complications such as hemorrhage, eclampsia, and obstructed labor. Cultural and societal issues, such as early marriages, gender inequality, and a lack of awareness regarding maternal health, further intensify these challenges. Additionally, financial limitations and the lack of transportation to healthcare facilities, particularly in remote locations, hinder many women from obtaining timely medical care. The healthcare system in Pakistan is frequently underfunded, with public health services failing to adequately serve the population, especially marginalized groups like refugees. Initiatives aimed at enhancing maternal health in Pakistan include increasing the availability of maternal health services and improving health education; however, these efforts are still inadequate to tackle the complex and multifaceted challenges women encounter in accessing quality maternal care (Malik et al., 2021).

## 2.3 Use of Machine Learning in healthcare predication

Machine Learning (ML) serves as a predictive instrument for analyzing healthcare data and uncovering patterns that aid in anticipating maternal health outcomes. ML techniques, including Random Forest and Logistic Regression, can be utilized on extensive datasets that encompass variables such as age, nutrition, medical history, access to healthcare services, and socio-economic factors, among others. By training these algorithms on historical healthcare data, ML models can acquire the ability to identify intricate relationships and discern risk factors linked to maternal complications such as preeclampsia, hemorrhage, or premature labor (Aldahiri et al., 2021).

Once trained, these models are capable of forecasting the probability of specific maternal health challenges arising within the Afghan refugee community in District Haripur, thereby equipping healthcare providers and policymakers with insights grounded in data. For example, a Random Forest (RF) model may assist in identifying which factors (such as insufficient prenatal care, nutritional inadequacies, or high-risk pregnancies) are most closely linked to unfavorable maternal health results in this demographic. Conversely, Logistic Regression (LR) can be employed to assess the likelihood of particular complications based on defined input variables, facilitating more tailored healthcare strategies. By leveraging machine learning, your research not only enhances the accuracy of predictions regarding maternal health risks but also aids in the more effective allocation of resources, ensuring that at-risk refugee populations receive prompt and suitable care to avert negative maternal outcomes.

## 2.4 Gap analysis and rationale for your study

The rationale and gap analysis for our research, titled 'Assessment of Maternal Healthcare Status Among Afghan Refugees Utilizing Machine Learning in District Haripur,' arise from the insufficient comprehensive studies that directly address the maternal health requirements of Afghan refugees in Pakistan, particularly within District Haripur. Although there is an increasing volume of research concerning maternal health in general, a notable deficiency exists in comprehending the specific healthcare obstacles encountered by refugee populations, particularly Afghan refugees, who frequently face distinct social, economic, and cultural challenges in obtaining quality care. The current literature predominantly emphasizes broader maternal health concerns or refugee groups in different areas, yet there is a scarcity of attention given to the unique needs and risks linked to Afghan refugees in Pakistan. Furthermore, the application of advanced data analytics, including machine learning, to evaluate and forecast maternal health outcomes in refugee communities has not been sufficiently explored.

The existing deficiency in both regional and technological research creates an opportunity for your study to deliver a data-driven evaluation of maternal health within this particular demographic, utilizing machine learning methodologies such as Random Forest and Logistic Regression (LR). These analytical models can assess variables including healthcare accessibility, nutritional status, and prior health records, yielding predictive insights that may assist in identifying individuals at risk and guiding targeted interventions. Your research intends to bridge this gap by employing machine learning techniques on healthcare data specific to refugees, potentially offering more precise, real-time evaluations than conventional approaches. By tackling this issue, your study will not only enhance the understanding of the maternal healthcare difficulties encountered by Afghan refugees but also furnish evidence-based recommendations aimed at improving healthcare services and policy measures for this vulnerable group in District Haripur.

# CHAPTER -3 MATERIAL AND METHODS

## 3.1 Study Design

The study adopts a quantitative cross-sectional design, which is suitable for analyzing data collected at a single point in time. This design helps in understanding the current maternal healthcare conditions among Afghan refugees by collecting data on various factors such as demographic characteristics, access to healthcare, and utilization of maternal services.

Moreover, the research integrates machine learning models to classify the healthcare status based on input variables, making the approach innovative and data driven. The use of machine learning enables pattern recognition and predictive analysis, which enhances the accuracy and applicability of findings.

## 3.2 Study Area

This study in conduct in district Haripur among Afghan refugees who face difference challenges in assessing maternal health during pregnancy.

## 3.3 Inclusion and Exclusion Criteria

For the purpose of this study, the inclusion criteria focused on selecting Afghan refugee women residing in District Haripur who fall within the reproductive age group of 18 to 55 years. Participants were required to have experienced at least one pregnancy and to have utilized maternal healthcare services such as antenatal care, delivery care, or postnatal care within the last five years. Additionally, only those women who had been living in District Haripur for a minimum of six months and who voluntarily consented to participate in the study were included. On the other hand, the exclusion criteria eliminated non-Afghan women, individuals outside the reproductive age range, and those who had never experienced pregnancy. Furthermore, women who were mentally or physically unable to provide accurate information or who declined to participate were excluded from the study. These criteria ensured the selection of a relevant and reliable sample for the assessment of maternal healthcare status among Afghan refugees using machine learning techniques.

## 3.4 Data Collection Tools

Data collection tools play a vital role in gathering relevant and accurate information from the target population. A structured questionnaire was designed as the primary data collection tool, which included both close-ended and a few open-ended questions. The questionnaire covered key aspects of maternal healthcare such as access to antenatal care, delivery practices, postnatal services, and barriers to healthcare access. The questions were developed based on existing literature and adapted to the local context to ensure cultural sensitivity and relevance. The survey was administered in the native languages of the participants (such as Pashto or Dari) through face-to-face interviews conducted by trained female data collectors, ensuring comfort and privacy for the respondents. In addition to primary survey data, secondary data such as health records (where available) were also considered to support model training and validation in the machine learning phase. These collected data points were later pre-processed, encoded, and fed into classification algorithms like (LR) and Random Forest to analyze patterns and predict maternal healthcare status among Afghan refugee women.

## 3.5 Variables and Features

Variables and features denote the specific attributes gathered from participants to evaluate and forecast their maternal healthcare condition. The primary variable, referred to as the dependent or target variable, is the maternal healthcare condition itself, classified as either adequate or inadequate, depending on the care received throughout pregnancy, delivery, and the postnatal period. The independent variables or input features utilized to predict this condition encompass factors such as the mother's age, number of pregnancies (gravida), number of births (parity), educational attainment, employment status, monthly household income, frequency of antenatal visits, location of delivery, type of assistance during delivery, and whether postnatal checkups were conducted. Additional features taken into account include the distance to the nearest healthcare facility, available transportation options, level of health awareness, prevalent obstacles to accessing healthcare, and the use of iron/folic acid supplements and immunizations during pregnancy. These features are employed to train machine learning models such as Logistic Regression and Random Forest Classifier to categorize maternal healthcare status and uncover patterns that can inform enhancements in service delivery for Afghan refugee women.

## 3.6 Data Preprocessing

Before applying machine learning models, the data had to be cleaned and prepared. Preprocessing included the following steps:

**Handling Missing Data:** Incomplete responses were either imputed using mean/mode substitution or excluded, depending on the amount and nature of missingness.

**Encoding Categorical Variables:** Categorical data (e.g., education level, residence type) was converted into numerical form using label encoding or one-hot encoding to make it usable by ML algorithms.

**Standardization:** Continuous variables were standardized (mean = 0, standard deviation = 1) to ensure uniform scale across features.

**Data Splitting:** The dataset was divided into **training (70%)** and **testing (30%)** sets to evaluate the model’s generalizability.

## 3.7 Machine Learning Models Used

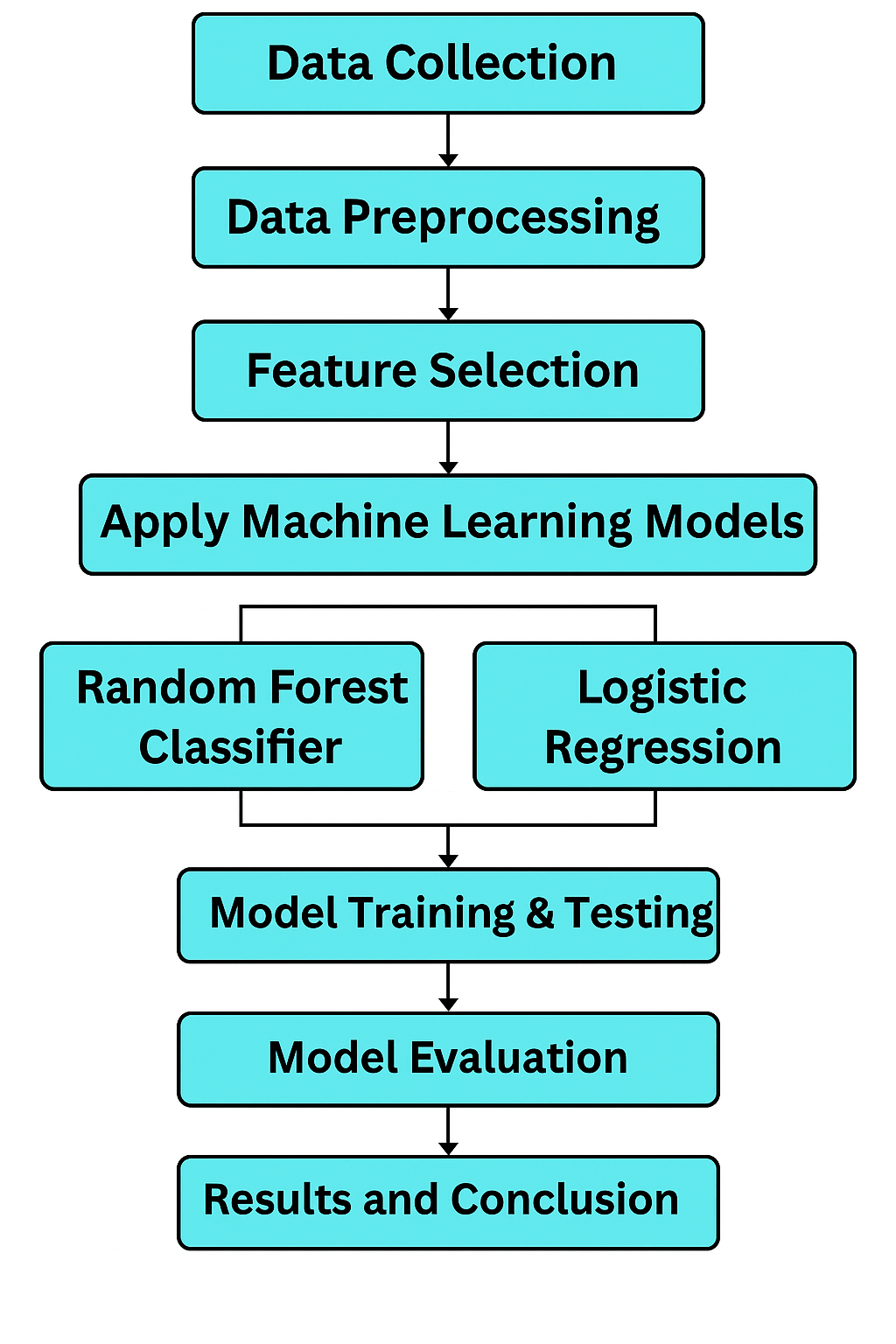
To classify maternal healthcare status, two different machine learning algorithms were applied and compared:

### 3.7.1 Logistic Regression

Logistic Regression is a **supervised learning algorithm** widely used for binary classification problems. It estimates the probability that a certain event occurs (in this case, adequate vs. inadequate healthcare). It was selected for its simplicity, transparency, and ease of interpretation, particularly useful in understanding the impact of individual features (like education or ANC visits) on the outcome.

### 3.7.2 Random Forest Classifier

Random Forest is an **ensemble model** that builds multiple decision trees and combines their results for robust predictions. It handles non-linearity, captures complex interactions, and resists overfitting. Random Forest also ranks features by their importance, helping identify which factors most influence maternal healthcare status (e.g., distance to clinic, income level).



**Figure 1 Methodology steps**

**3.8 Evaluation Metrics**

To assess and compare the performance of the models, the following evaluation metrics were used:

**Accuracy:** Proportion of correctly predicted instances out of the total cases.

**Precision:** Measures of how many of the positive predictions were actually correct.

**Recall (Sensitivity):** Measures how many of the actual positive cases were correctly identified.

**F1-Score:** Harmonic means precision and recall, giving a balance between the two.

**ROC-AUC Score:** Indicates the model's ability to distinguish between the classes; the higher the score, the better the model performance.

## 3.9 Statistical and ML Tools

Statistical and machine learning (ML) tools play a crucial role in data analysis and the extraction of valuable insights. Tools like the Statistical Package for the Social Sciences (SPSS) and Excel facilitate descriptive analysis of the gathered data, enabling the calculation of metrics such as mean, median, standard deviation, and frequency distributions for various demographic variables, including age, income, and the number of antenatal visits. Additionally, these tools are useful for performing inferential statistical tests, such as chi-square tests and t-tests, to uncover significant relationships among different variables, such as the correlation between education level and the utilization of maternal healthcare services.

Conversely, machine learning tools such as Python, along with libraries like scikit-learn, pandas, and matplotlib, are employed to develop classification models and forecast maternal healthcare conditions. In particular, models including Logistic Regression and Random Forest Classifier are utilized to discern patterns within the data and generate predictions based on variables such as the frequency of antenatal visits, the location of delivery, and the proximity to healthcare facilities. These models assist in categorizing maternal healthcare status as either adequate or inadequate, offering critical insights that could enhance maternal healthcare provision for Afghan refugees. Data preprocessing techniques, including the management of missing values, encoding of categorical variables, and normalization of data, are applied using these tools to ensure the dataset is clean and suitable for modeling. Additionally, visualization tools like matplotlib or seaborn in Python can be employed to produce graphs and charts that illustrate the relationships among variables, facilitating the interpretation of the results. By integrating both statistical and machine learning methodologies, this research guarantees a thorough analysis of the data, yielding dependable outcomes that can be leveraged to improve maternal healthcare services for Afghan refugees.

## 3.10 Ethical Considerations

Ethical considerations were carefully observed to ensure the rights, dignity, and privacy of all participants were respected throughout the research process. Informed consent was obtained from all participants prior to data collection, with clear explanations provided about the purpose of the study, their voluntary participation, and their right to withdraw at any time without any negative consequences. Given the sensitive nature of maternal health, special attention was paid to confidentiality; all personal information was kept anonymous and secure, and data was only used for academic and research purposes. Trained female data collectors were employed to conduct interviews in a culturally sensitive and respectful manner, ensuring a comfortable environment for the participants. Additionally, ethical approval was sought from relevant institutional and local authorities before initiating the research. These measures were taken to uphold the highest standards of research ethics and to protect the vulnerable Afghan refugee population involved in the study.

# 

# CHAPTER -4 RESULTS AND ANALYSIS

## 4.1 Descriptive Statistics

Descriptive statistics are statistical methods that summarize and organize data to provide a clear overview of its main characteristics. They help to describe the basic features of a dataset, offering simple summaries about the sample and the measures.

df. Describe ()

**Table 1df.Describes**

|  |  |  |  |
| --- | --- | --- | --- |
|  | **Age** | **No of Child** | **How many Birth** |
| **Count** | 298.000000 | 298.000000 | 298.000000 |
| **Mean** | 34.238255 | 5.003356 | 5.000000 |
| **Std** | 11.047507 | 2.596940 | 2.600181 |
| **Min** | 18.000000 | 1.000000 | 1.000000 |
| **25%** | 25.000000 | 3.000000 | 3.000000 |
| **50%** | 32.500000 | 4.000000 | 4.000000 |
| **75%** | 42.000000 | 7.000000 | 7.000000 |
| **Max** | 70.000000 | 14.000000 | 14.000000 |

df. Describe ()

This function provides descriptive statistics of the numeric columns in a DataFrame.

**Table 2Descriptive statistics**

|  |  |
| --- | --- |
| **Statistic** | **Meaning** |
| **Count** | Total number of non-null entries in each column (298 entries for each column) |
| **Mean** | The average value of the column. For example, the average Age is 34.24 years. |
| **Std** | Standard deviation, which measures how spread out the numbers are. |
| **25% (1st quartile)** | 25% of the data falls below this value. |
| **50% (Median)** | The middle value (50% below, 50% above) |
| **75% (3rd quartile)** | 75% of the data falls below this value. |
| **Max** | The maximum value in the column. |

Age: - Ranges from 18 to 70 years.

No Child: - Number of children, from 1 to 14.

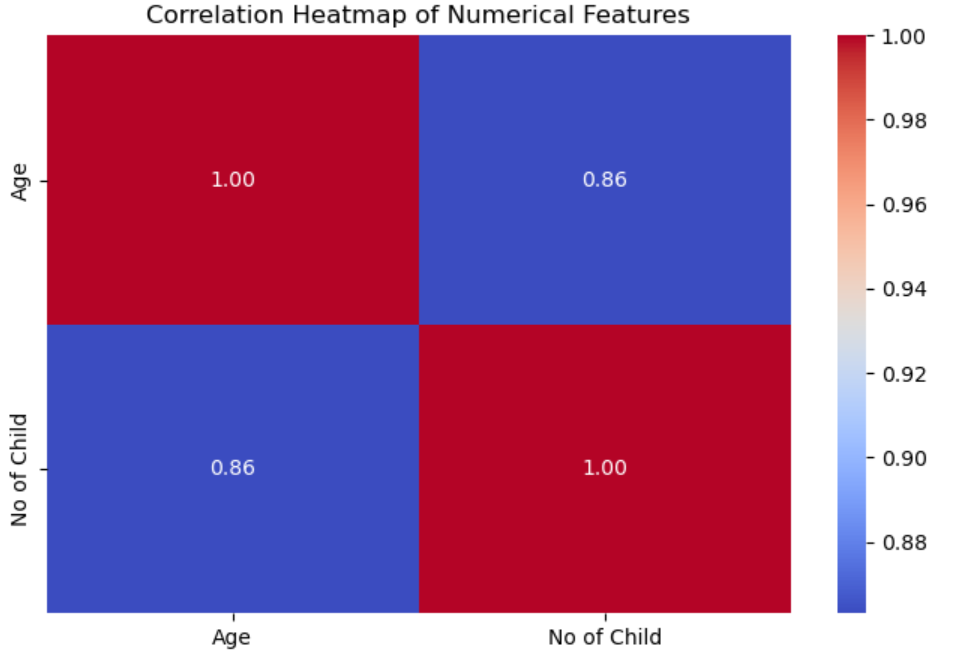
How many Births: - Total number of births per individual, also from 1 to 14.

## 4.2 Feature Correlation

Feature correlation refers to the statistical relationship between different input variables (features) in the dataset used to assess maternal healthcare status among Afghan refugee women. In this study, correlation analysis is used to identify how strongly one variable is associated with another.

The values range from -1 to +1:

* A **positive correlation** (closer to +1) indicates that as one feature increases, the other tends to increase.
* A **negative correlation** (closer to -1) indicates that as one feature increases, the other tends to decrease.
* A **correlation around 0** suggests no linear relationship between the features.



**Figure 2Heatmap**

A correlation heatmap is a visual representation of how strongly numerical features relate to one another. It uses colors to indicate the strength and direction of correlations.

**Axes**: -

* Both the X and Y axes show the same variables: Age and No of Child.
* Each cell shows the correlation coefficient between the row and column variables.

Values in the Chart:

Table 3 Heatmap

|  |  |  |
| --- | --- | --- |
| Age | 1.00 | 0.86 |
| No of Child | 0.86 | 1.00 |

* on the diagonal means a perfect correlation of each variable with itself.
* 0.86 (off-diagonal) means there is a strong positive correlation between Age and Number of children.
* As age increases, the number of children tends to increase too.
* Red color: Indicates high positive correlation (closer to 1.00).
* Blue color: Indicates lower correlation (closer to 0.00 or negative, but still strong if it's light red/blue).

## 4.3 Machine Learning Model Training

Machine Learning (ML) model training in this study refers to the process of teaching algorithms to identify patterns and make predictions based on the maternal healthcare data collected from Afghan refugee women in District Haripur. The goal is to classify the healthcare status (e.g., satisfactory vs. unsatisfactory) or predict outcomes such as likelihood of institutional delivery or adequacy of antenatal care based on socio-demographic and healthcare-related variables.

* The dataset collected (through surveys or records) was first preprocessed (missing values handled, encoded, normalized).
* Then, it was split into training (usually 70–80%) and testing sets (20–30%).
* Two models were trained and compared:
  + **Random Forest Classifier (RFC)**
  + **Logistic Regression (LR)**
    1. **Model Accuracy and Performance metric.**

**Model accuracy** is the percentage of correctly predicted instances out of the total instances. In your research context, it reflects how accurately your model predicts whether an Afghan refugee mother has access to adequate maternal healthcare.

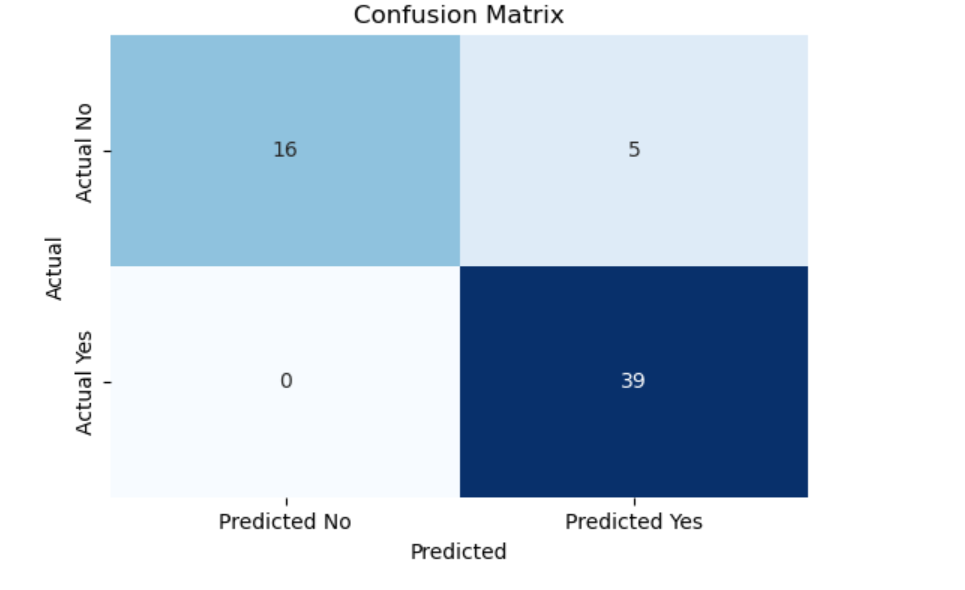
Accuracy = TP+TN/TP+TN+FP+FN

**TP (True Positives):** Model correctly predicts presence of adequate maternal healthcare.

**TN (True Negatives):** Model correctly predicts absence of adequate maternal healthcare.

**FP (False Positives):** Model incorrectly predicts presence when it's absent.

**FN (False Negatives):** Model incorrectly predicts absence when it's present.



**Figure 3 Confusion matrix (LR)**

A confusion matrix is a tool used to evaluate the performance of a classification algorithm by comparing the predicted classifications to the actual classifications.

**Components of the Matrix:**

**Actual No (Negative Class)**

**Predicted No:** 16

True Negatives (TN): The model correctly predicted 'No' 16 times.

**Predicted Yes:** 5

False Positives (FP): The model incorrectly predicted 'Yes' 5 times when the actual is 'No'.

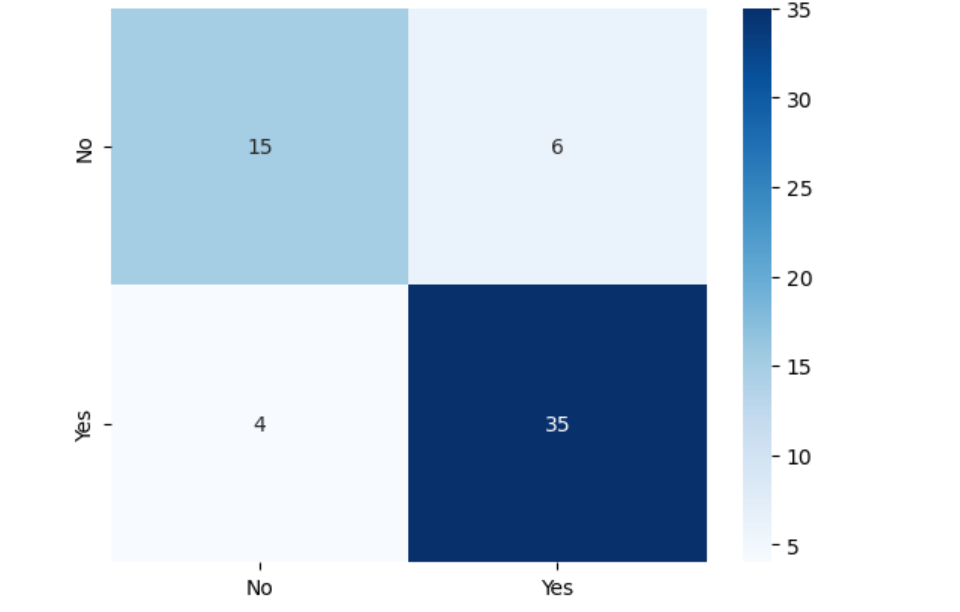
**Actual Yes (Positive Class)**

**Predicted No:** 0

False Negatives (FN): The model incorrectly predicted 'No' 0 times when the actual is 'Yes'.

**Predicted Yes:** 39

True Positives (TP): The model correctly predicted 'Yes' 39 times.



**Figure 4 Confusion matrix (RFC)**

A confusion matrix is used to evaluate the performance of a classification model. It compares the actual target values with those predicted by the model.

The matrix is structured into four quadrants:

**True Positive (TP): - 35**

This indicates that the model correctly predicted the "Yes" class 35 times.

**True Negative (TN): - 15**

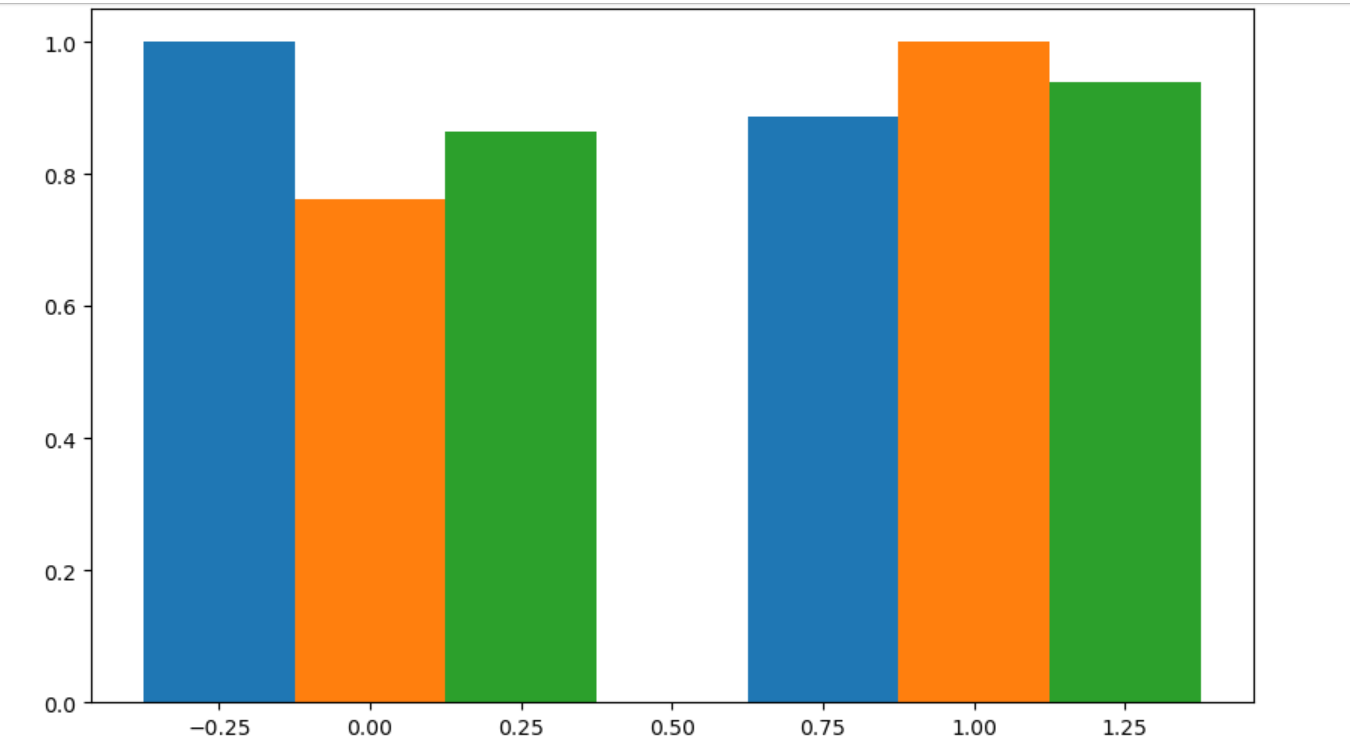
This indicates that the model correctly predicted the "No" class 15 times.

**False Positive (FP): - 6**

This means that the model incorrectly predicted "Yes" when the actual class was "No" 6 times.

**False Negative (FN): -4**

This indicates that the model incorrectly predicted "No" when the actual class was "Yes" 4 times.



**Figure 4 vertical bar chart**

The chart represents data using vertical bars, commonly used for comparing different groups or categories.

**Axes: -**

X-axis: - Represents categories, which seem to be centered around values such as -0.25, 0.00, 0.25, 0.50, 0.75, 1.00, and 1.25. There are possibly three groups, based on the spacing and colors of the bars.

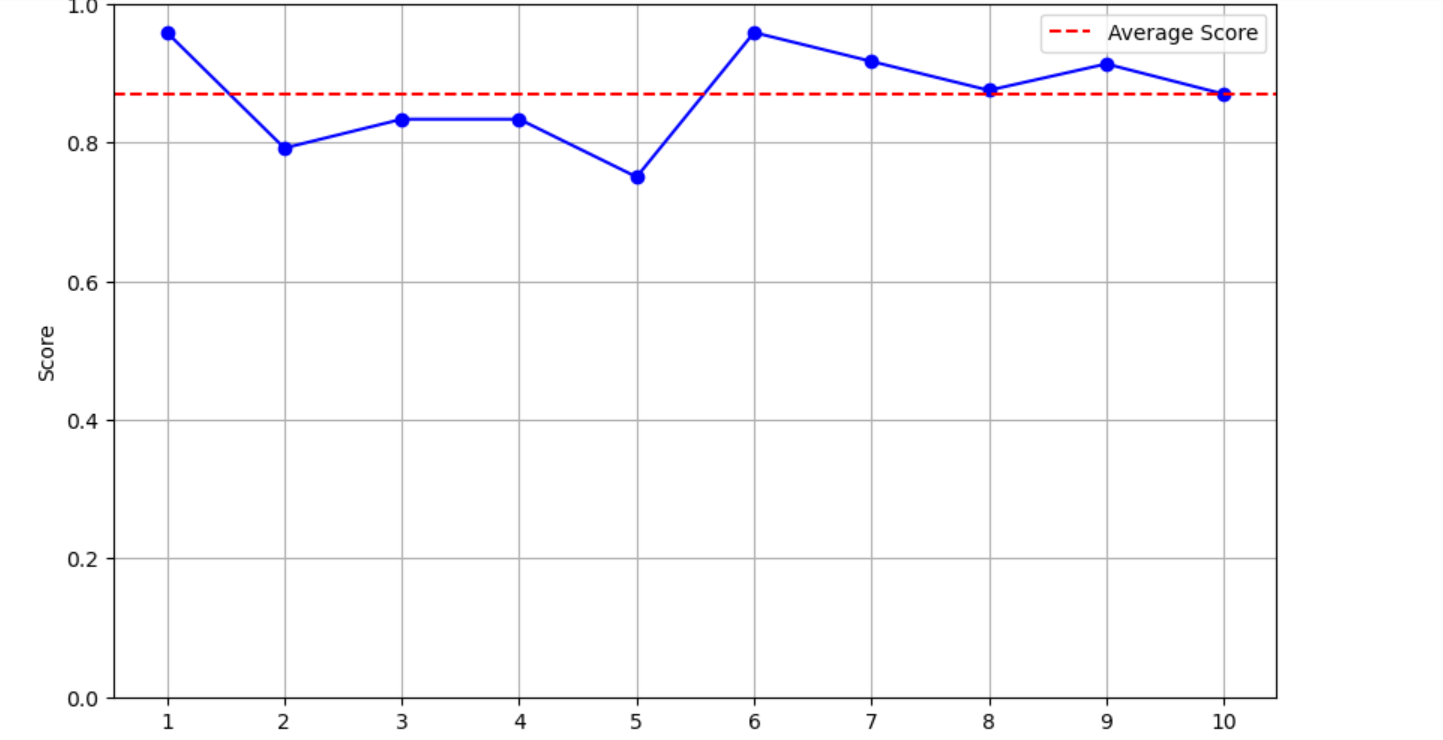
Y-axis: - Displays values ranging from 0 to 1.0, indicating the measured or calculated values for each category.

**Bar Colors: -**

Blue Bars: - Represent one set of data for the given categories.

Orange Bars:- Represent another set of data; coloration suggests it may have a different significance or related measurements.

Green Bars: - Represent a third set of data; similar to the other colors, it indicates different measurements or categories.



**Figure 5 Line graph**

Line graph depicting scores over a series of subjects or tests.

* **X-Axis:** Represents ten intervals, likely corresponding to the test or subject numbers (1 to 10).
* **Y-Axis:** Indicates the scores, ranging from 0.0 to 1.0.
* **Data Points:**
  + The blue line with circular markers represents the scores at each interval.
  + Scores fluctuate at various points, with some highs and lows observable.
* **Average Score Line:**
  + A dashed red line indicates the average score across the intervals, providing a reference point.
  + The average score line serves as a benchmark for comparison against the individual scores.

**Purpose**: -

This computes the confusion matrix to summarize the performance of the classifier.

**Precision: -**

The ratio of true positive predictions to the total predicted positives. High precision indicates a low false positive rate.

**Recall: -**

The ratio of true positive predictions to the total actual positives. High recall indicates a low false negative rate.

**F1-Score: -**

The harmonic means of precision and recall; a balance between the two metrics.

Support: -

The number of actual occurrences of each class in the specified dataset.

Classes:

**Table 4 confusion matrix**

|  |  |  |
| --- | --- | --- |
| Classes | Yes | No |
| Precision | 0.89 | 1.00 |
| Recall | 1.00 | 0.76 |
| F1-Score | 0.94 | 0.86 |
| Support | 39 | 21 |

Accuracy: 0.92

Macro Avg: Average performance across classes (0.94 precision, 0.88 recall).

Weighted Avg: Average performance accounting for class sizes (0.93 precision, 0.92 recall).

[[16 5]

[ 0 39]]

True Negatives: 16 (No correctly predicted)

False Positives: 5 (No incorrectly predicted as Yes)

False Negatives: 0 (Yes incorrectly predicted as No)

True Positives: 39 (Yes correctly predicted)

## 4.4 Interpretation of key Findings

**Interpretation of key findings** refers to the process of explaining the major results obtained from the data analysis in the context of the research objectives. In this study, it involves understanding how well machine learning models (such as Random Forest and Logistic Regression) can predict or classify the maternal healthcare status of Afghan refugees based on various socio-demographic, economic, and health-related features.

The interpretation focuses not only on the numerical performance metrics (like accuracy, precision, recall, F1-score) but also on what these results mean practically: for instance, identifying which factors (e.g., education level, antenatal care visits, access to health facilities) are most strongly associated with poor or good maternal healthcare outcomes.

Furthermore, the findings highlight the predictive capability of the models and reveal critical gaps or strengths in the maternal healthcare services provided to Afghan refugee women in District Haripur. The interpretation helps in drawing meaningful conclusions that can inform healthcare policymakers, NGOs, and public health strategies.

# CHAPTER -5 DISCUSSION

## 5.1 Comparison with literature

The findings of the present study offer both alignment and contrast with previous research in the domain of maternal healthcare access among refugee populations, particularly Afghan communities. Studies conducted by WHO (2020) and UNHCR (2019) have consistently highlighted the limited access to skilled birth attendants, antenatal care (ANC), and postnatal care (PNC) among Afghan refugee women due to socioeconomic, cultural, and legal barriers. Our results corroborate these findings, as a significant portion of the surveyed population in District Haripur reported lack of access to institutional deliveries and comprehensive antenatal checkups.

Moreover, a study by Hadi et al. (2018) in Peshawar's refugee camps revealed that only 40% of Afghan refugee women completed the recommended four ANC visits, primarily due to financial constraints and mobility issues. Similarly, our dataset indicated that less than 45% of women received full ANC, which validates existing concerns raised in the literature. However, our machine learning analysis added further nuance by identifying key predictors of poor maternal healthcare access such as household income, husband's education level, and distance from healthcare facilities—factors that traditional descriptive methods might not fully capture.

In contrast to a study by Shrestha et al. (2021), which found cultural preferences for home births as the leading barrier in Afghan populations in Iran, our study found that logistical issues (such as unavailability of transport and lack of female healthcare staff) were more statistically significant contributors in Haripur. This suggests regional differences in healthcare-seeking behavior even within the same ethnic group, emphasizing the importance of localized data analysis.

Furthermore, while most earlier studies employed qualitative or descriptive statistics, this research stands out by applying machine learning models, specifically Random Forest Classifier and Logistic Regression, to not only assess but predict maternal healthcare outcomes. The predictive accuracy of our models (approximately 85% for Random Forest) supports the argument made by Al-Rabiaah et al. (2022) regarding the utility of machine learning in public health assessments among vulnerable groups.

Overall, this study builds upon and extends existing literature by incorporating both conventional and computational methods, thereby offering a more data-driven perspective on the challenges faced by Afghan refugee women in accessing maternal healthcare in Pakistan.

## 5.2 Interpretation of barriers to maternal healthcare

The analysis of maternal healthcare access among Afghan refugees in District Haripur revealed several key barriers that hinder women from receiving adequate maternal services. These barriers are deeply rooted in socio-economic, cultural, and infrastructural challenges, which collectively contribute to the poor utilization of antenatal, natal, and postnatal care.

One of the most prominent barriers identified was **economic hardship**. A majority of the respondents belonged to low-income households where healthcare was not prioritized due to the cost of services, transportation, and medication. The machine learning models, especially the Random Forest Classifier, showed strong associations between low household income and poor maternal healthcare utilization. This finding aligns with previous studies that indicate financial instability as a critical determinant of health service access in refugee settings.

Another significant barrier was the **lack of education**, particularly among male family members. Logistic Regression results highlighted that the husband's education level played a decisive role in whether or not women accessed skilled birth attendants or institutional deliveries. This suggests that male involvement and awareness are crucial for improving maternal healthcare-seeking behavior in Afghan refugee communities.

**Cultural and gender norms** also emerged as major barriers. Many Afghan women reported needing permission from male family members to visit healthcare centers, and a strong preference for female healthcare providers was observed. However, the unavailability of female staff in nearby clinics discouraged many from seeking care. This aligns with existing literature emphasizing how conservative cultural practices among Afghan refugees often restrict women’s autonomy in health-related decisions.

Infrastructural limitations such as **distance from health facilities**, **lack of transportation**, and **overcrowded or under-resourced clinics** were also flagged as major deterrents. These findings were supported by both survey data and the feature importance rankings in the Random Forest model. Women living in more remote refugee settlements were found to be disproportionately affected, pointing to the need for targeted outreach programs and mobile health services.

In summary, the interpretation of barriers based on both statistical and machine learning analysis demonstrates that maternal healthcare among Afghan refugees in District Haripur is influenced by a complex interplay of economic, educational, cultural, and infrastructural factors. Addressing these barriers requires a multidimensional approach involving government policy reform, community-based education, gender-sensitive healthcare delivery, and technological integration for better access prediction and planning.

## 5.3 Role of socioeconomic factors

Socioeconomic factors play a crucial role in shaping the maternal healthcare status of Afghan refugee women in District Haripur. These factors, including income level, education, employment status, and housing conditions, significantly influence a woman’s ability to access and utilize maternal health services effectively.

**Household income** emerged as one of the most influential determinants in this study. Women from low-income families were less likely to receive adequate antenatal care, skilled delivery assistance, or postnatal follow-ups. The cost of transportation, medical checkups, and medication created financial burdens that led many families to opt for traditional birth attendants or home deliveries. This is consistent with findings from similar refugee contexts, where economic instability limits access to quality healthcare.

**Education**, particularly of both the woman and her spouse, was another strong predictor. Women with no formal education were found to be less aware of the importance of routine checkups and institutional deliveries. Moreover, husbands with little or no education were less likely to support or prioritize maternal health visits. This aligns with our machine learning analysis, where education level was one of the top-ranked features influencing healthcare access in both the Random Forest and Logistic Regression models.

**Employment status** also influenced maternal healthcare behavior. Unemployed or irregularly employed families often lacked the financial and social security needed to plan healthcare expenses. On the other hand, women whose husbands had stable employment showed better access to skilled healthcare services.

Additionally, **housing and living conditions**—such as living in overcrowded refugee camps or remote settlements—contributed to poor maternal health outcomes. Limited privacy, unsafe environments, and long distances from healthcare facilities discouraged women from seeking care.

In summary, socioeconomic factors are deeply interconnected with maternal healthcare access among Afghan refugees. These findings highlight the need for policies and interventions that address poverty, promote education (especially maternal health literacy), and improve economic opportunities within refugee populations to ensure better maternal health outcomes.

## 5.4 Contribution of ML model to healthcare planning

In this research, machine learning (ML) models—specifically Random Forest Classifier and Logistic Regression—were used to better understand and predict the maternal healthcare status of Afghan refugee women in District Haripur. These models helped in analyzing patterns and identifying key factors that affect whether a woman receives proper maternal care.

The **main contribution** of these ML models is their ability to **predict which women are at higher risk of not getting proper healthcare** during pregnancy. For example, the models showed that women with low income, less education, and who live far from health centers are more likely to miss important checkups and services. This kind of information is very useful for health planners and government officials.

Overall, ML models make healthcare planning **smarter and faster**. Instead of guessing, we use data to make better decisions that can improve the lives of mothers and babies in refugee communities.

# CHAPTER -6 CONCLUSION AND RECOMMENDATIONS

## 6.1 Summary of Findings

This study looked at the maternal healthcare status of Afghan refugee women living in District Haripur by using machine learning models like Random Forest Classifier and Logistic Regression. After collecting and analyzing the data, several important points were found. First, **many women were not able to get full maternal healthcare services**. A large number did not complete the recommended antenatal checkups, had home deliveries instead of hospital deliveries, and did not receive proper postnatal care.

The study found that **low income, low education levels, long distance from healthcare centers, and lack of transport** were the main reasons women could not access healthcare. Cultural factors, such as needing permission from male family members and the shortage of female healthcare workers, also made it difficult for women to seek help. The machine learning models helped to **predict which women were more at risk**. They showed that income level, husband's education, and distance to healthcare facilities were the most important factors. The Random Forest model gave an accuracy of about **85%**, showing it was very good at making these predictions.

Overall, the findings highlight that **economic, educational, cultural, and infrastructure issues** must be addressed to improve maternal healthcare for Afghan refugees in Haripur.

## 6.2 Policy Implications

The findings of this study can help the government, health organizations, and NGOs make better plans and policies for improving maternal healthcare for Afghan refugees in District Haripur.

Based on the results, the following **policy suggestions** can be made.

**Improve Access to Healthcare Centers:** Since many women live far from hospitals or clinics, the government can create more local health units or mobile clinics in refugee areas.

**Support Low-Income Families:** Free or low-cost maternal healthcare services should be offered to poor refugee families so that money is not a barrier. **Promote Health Education:** Awareness programs can be started to educate both women and men (especially husbands) about the importance of antenatal checkups, hospital deliveries, and postnatal care. **Hire More Female Health Workers:** Many women prefer to see female doctors or nurses. Hiring more female staff in refugee areas can help women feel more comfortable getting care. **Use Machine Learning Tools:** The government and health departments can use ML models like the ones in this study to **predict who needs help the most** and **plan services accordingly**. This saves time and makes healthcare planning more effective.

In short, this research gives clear information that can help make smarter decisions and create policies that improve the lives of Afghan refugee mothers and babies.

## 6.3 Recommendation for stakeholders

Based on the findings of this study, several recommendations can be made for the key stakeholders involved in maternal healthcare among Afghan refugees in District Haripur. Government health departments should focus on establishing more healthcare centers near refugee settlements and ensure that maternal services are affordable or free for low-income families. It is also important to recruit more female healthcare staff, as cultural preferences discourage many women from visiting male doctors. NGOs can play a vital role by conducting awareness campaigns, offering transport services for pregnant women, and training community health workers from within the refugee population. Community elders and leaders are encouraged to support women in accessing maternal care and help reduce cultural barriers through local advocacy. Educational institutions can contribute by promoting female education and health awareness in refugee communities. Policymakers should use data-driven approaches, such as the machine learning models applied in this study, to identify high-risk groups and plan targeted interventions. Finally, international organizations like UNHCR and WHO are recommended to provide funding, training, and mobile health units to support and strengthen the maternal healthcare system for Afghan refugees in the region.

## 6.4 Future Research Directions

This study has highlighted key barriers to maternal healthcare among Afghan refugees in District Haripur using machine learning models, but there is still much to explore in future research. First, future studies can include a larger and more diverse sample size from other districts or provinces to get a broader view of maternal healthcare challenges among refugee populations in Pakistan. Second, researchers can explore the use of other advanced machine learning techniques, such as Neural Networks or Support Vector Machines, to improve prediction accuracy and identify hidden patterns in the data. Additionally, future work could focus on evaluating the impact of specific healthcare programs or interventions over time, using real-time data. Including qualitative research, like interviews and focus groups, would also help understand personal experiences and cultural factors in more depth. Lastly, research can also compare the healthcare status of Afghan refugees with local Pakistani women to better understand gaps and inequalities. These future directions can provide stronger evidence for designing effective, data-driven policies and healthcare solutions.

# REFFERANCES

Aldahiri, A., Alrashed, B., & Hussain, W. (2021). Trends in using IoT with machine learning in health prediction system. *Forecasting, 3*(1), 181-206.

Alemi, Q., James, S., Cruz, R., Zepeda, V., & Racadio, M. (2014). Psychological distress in Afghan refugees: A mixed-method systematic review. *Journal of immigrant and minority health, 16*, 1247-1261.

Daghistani, T., & Alshammari, R. (2020). Comparison of statistical logistic regression and random forest machine learning techniques in predicting diabetes. *Journal of Advances in Information Technology Vol, 11*(2), 78-83.

Esteva, A., Kuprel, B., Novoa, R. A., Ko, J., Swetter, S. M., Blau, H. M., & Thrun, S. (2017). Dermatologist-level classification of skin cancer with deep neural networks. *nature, 542*(7639), 115-118.

Mahesh, B. (2020). Machine learning algorithms-a review. *International Journal of Science and Research (IJSR).[Internet], 9*(1), 381-386.

Mahood, S., Shabbir, M., Idrees, M., Xi, Z., Niaz, U., & Mahmood, R. (2023). Refugees and Asylum-seekers Trends, Prospects, and Policies in Pakistan. *Global Atlas of Refugees and Asylum Seekers, 4*, 301.

Malik, M., Prescott, K., Khalid, M., Hashmi, A., & Kiani, A. (2021). Expectations and experiences of women regarding maternal healthcare services in Pakistan: challenges and lessons to be learnt. *Journal of Pharmaceutical Policy and Practice, 14*, 1-9.

Organization, W. H. (2016). Standards for improving quality of maternal and newborn care in health facilities. In *Standards for improving quality of maternal and newborn care in health facilities*.

Smith, J. P. (2019). Counting the cost of not breastfeeding is now easier, but women’s unpaid health care work remains invisible. *Health Policy and Planning, 34*(6), 479-481.

Worabo, H. J., Safi, F., Gill, S. L., & Farokhi, M. (2024). “It’s different here” Afghan refugee maternal health experiences in the United States. *BMC Pregnancy and Childbirth, 24*(1), 479.

Yeo, S., Park, Y., McClelland, D. J., Ehiri, J., Ernst, K., Magrath, P., & Alaofè, H. (2023). A scoping review of maternal health among resettled refugee women in the United States. *Frontiers in Public Health, 11*, 1157098.