Machine Learning for Predicting Postpartum Depression and Suicidal Tendencies in Bangladesh

By

Faria Tuz Zohora
ID: 012221044

Submitted in Partial Fulfillment of the Requirements For the Degree of Master of Science in Computer Science and Engineering

September 20,2024

Department of Computer Science and Engineering
United International University

Approval Certificate

This thesis titled "Machine Learning for Predicting Postpartum Depression and Suicidal Tendencies in Bangladesh" submitted by **Faria Tuz Zohora**, Student ID: **012221044**, has been accepted as Satisfactory in fulfillment of the requirement for the degree of Master of Science in Computer Science and Engineering on 31st December, 2024.

Board of Examiners

1. Supervisor

Dr. Khondaker Abdullah Al Mamun
Professor & Director - AIMS Lab
Department of CSE
United International University
Dhaka, Bangladesh

2. Head Examiner

Dr. Novia Nurain
Assistant Professor,
Department of CSE
United International University
Dhaka, Bangladesh

3. Examiner-I

Dr. Riasat Azim
Assistant Professor
Department of CSE
United International University
Dhaka, Bangladesh

4. Examiner-II

Ms. Rubaiya Rahtin Khan
Assistant Professor,
Department of CSE
United International University
Dhaka, Bangladesh

5. Ex-Officio

Prof. Dr. Md. Motaharul Islam
Professor and Director of MSCSE
Department of CSE
United International University
Dhaka, Bangladesh

Declaration

This is to certify that the work entitled **"Machine Learning for Predicting Postpartum Depression and Suicidal Tendencies in Bangladesh"** is the outcome of the research carried out by me under the supervision of Professor Dr. Khondaker Abdullah Al Mamun.

Faria Tuz Zohora
Student ID: 012221044
Department of CSE
United International University
Dhaka, Bangladesh

In my capacity as supervisor of the candidate's thesis, I certify that the above statements are true to the best of my knowledge.

Dr. Khondaker Abdullah Al Mamun
Professor & Director - AIMS Lab
Department of CSE
United International University
Dhaka, Bangladesh

Abstract

Postpartum depression (PPD) has become a significant mental health disorder affecting mothers across the world, with prevalence rates of 6% in high-income countries (HICs) and 20% in low-middle-income countries (LMICs) leading to serious consequences for maternal and neonatal health. In countries like Bangladesh, PPD remains a critical issue, contributing to maternal suicides due to the overwhelming psychological burden. The condition often goes undiagnosed in such regions due to cultural stigma surrounding mental health and a lack of education, preventing mothers from seeking help. This thesis aims to develop a machine learning model using symptom and behavioral data, to detect suicidal ideation in women with PPD, reducing diagnostic delay for early intervention and improving access to care by recommending mental health resources. Additionally, the study investigates the influence of various feature selection methods, namely, expert-selected features, Sequential Feature Selection (SFS), and Pearson's correlation (PCC) as well as balancing techniques such as undersampling and oversampling on model performance, developed using four machine learning algorithms; Decision Tree (DT), Logistic Regression (LR), Random Forest (RF) and Support Vector Machine (SVM). Random Forest, trained using the original unsampled dataset and expert-selected features, gave the best performance across other models when using various feature selection techniques and sampling approaches. Decision Trees are also reasonably accurate while Logistic Regression and Support Vector Machines, on the other hand, are less suitable for this prediction. While expert-selected features give the best outcome, it is also identified that PCC-selected features are most beneficial for LR and SVM while tree-based classifiers are benefited more with features selected by SFS. Comparisons with existing research reveal that the developed models achieve good performance, with AUC values ranging from 0.86 to 0.98 for Random Forest and Decision Tree, outperforming the PPD and suicide prediction models from existing research.

Acknowledgements

First and foremost, I want to thank my creator for allowing me to gain knowledge and cross my path with all the people supporting me in completing this thesis work.

Next, I would like to express my humble gratitude to my supervisor for his insightful feedback, motivation, and patience. He continued to encourage me to achieve better and to think like a researcher. Special thanks to the researchers and medical professionals who guided this work and generously shared their knowledge, feedback, and expertise on healthcare topics, which validated and improved the quality of this research.

I am also grateful to my husband, who is also my classmate and study partner, for his editing help, feedback, proofreading, and moral support. I am also grateful to the research assistants from the university who guided me with examples, tutorials, and pieces of advice. I'd also like to recognize the advancements made in technology such as the internet and artificial intelligent tools for research that improved access to knowledge, further supporting this work.

Finally, I want to acknowledge the support of my family, especially my parents, parents-in-law, and my friend's family who took care of my child while I attended classes and worked on this study. Their faith in me has kept my motivation high during this time and allowed me to continue my education.

Publication List

The main contributions of this research are either published or accepted or in preparation in journals and conferences as mentioned in the following list:

Journal Articles

1. The Use of Artificial Intelligence in Advancing Sexual and Reproductive Health and Rights (SRHR) Interventions: A Systematic Review

   Submitted in Heliyon (Q1 Journal)

   Manuscript number: **HELIYON-D-24-70207**

2. Identifying and predicting suicidality among postpartum mothers: a machine learning approach

   Journal in preparation

Additional Publications

Following is the list of relevant publications published in the course of the research that is not included in the thesis:

1. Predicting women's fertility using Artificial Intelligence

   (Published by Elsevier)

   Book Title: **Artificial Intelligence in e-Health Framework, Volume 1**

2. The Use of Artificial Intelligence in Advancing Mental Health Interventions: A Systematic Review

   Submitted in Heliyon (Q1 Journal)

   Manuscript number: **HELIYON-D-24-67395R1**

Table of Contents

List of Figures **viii**

List of Tables **ix**

1 Introduction **1**

1.1 Background . . . 1
1.2 Motivation . . . 2
1.3 Objectives . . . 2
1.4 Contribution . . . 3
1.5 Outline of thesis . . . 3

2 Literature Review **4**

2.1 Sexual and Reproductive Health and Rights . . . 4
2.1.1 Pregnancy health risk . . . 5
2.1.2 Childbirth complications . . . 6
2.1.3 Pregnancy loss . . . 7
2.1.4 Sexual and gender-based violence . . . 7
2.2 Mental health . . . 8
2.2.1 Anxiety disorders . . . 8
2.2.2 Mood disorders . . . 10
2.3 Postpartum depression (PPD) . . . 11
2.4 Conclusion . . . 12

3 Exploration of Postpartum depression and Suicide ideation prediction **13**

3.1 PPD data analysis . . . 13
3.2 Suicidal tendency prediction models . . . 17
3.2.1 Type of data used for suicidal ideation prediction . . . 18
3.2.2 Machine learning algorithms used to train prediction models . . . 18
3.2.3 Metrics used to evaluate prediction models . . . 19
3.3 Conclusion . . . 21

4 Methodology **22**

4.1 Data Collection . . . 22
4.2 Data Preprocessing . . . 23
4.3 Data Balancing . . . 25
4.4 Feature Selection . . . 25
4.5 Model training and evaluation . . . 26

4.6 Ethical Considerations and Data Privacy 28
4.7 Conclusion 30

5 Result Analysis **31**

6 Discussion and conclusion **35**
6.1 Model performance 35
6.2 Sampling techniques 36
6.3 Feature selection techniques 37
6.4 Comparison with existing research 38
6.5 Limitations and Challenges 39
6.5.1 Limitations 39
6.5.2 Challenges 39
6.6 Conclusion 39

References **56**

List of Figures

1 Components of Sexual and Reproductive Health. 5
2 Components of Sexual and Reproductive Health. 9
3 Chart showing the studies using different types of data to train suicidal tendency prediction models. 19
4 Correlation heatmap of dataset attributes 23
5 Complete methodology of the prediction model 29
6 Performance of the machine learning algorithms with expert selected features and unbalanced dataset. 31
7 Average accuracy of the machine learning models trained using expert selected features, represented in the form of a bar chart and box plot, different sampling methods. 32
8 Average accuracy of the machine learning models represented in the form of a bar chart and box plot, for different feature selection methods. 32
9 Average accuracy of the machine learning models represented in the form of a bar chart and box plot, for different sampling methods. 33
10 AUC under ROC curve for the best-performing model where Random forest was used with expert-selected features on the original unbalanced dataset. 33

List of Tables

1 Commonly used screening tools for PPD 14
2 Symptoms of patients with PPD 15
3 Risk factors associated with the prevalence of PPD 15
4 Studies using different types of machine learning algorithms to train suicidal tendency prediction models. 20
5 Different types of metrics used to evaluate suicide prediction models. 21

6 Characteristics of the population in the dataset in respective of symptoms and risk factors. 24
7 List of features selected by Pearson's correlation coefficient 26
8 List of features selected by expert advice 27

9 Evaluation of the machine learning models using accuracy (ACC), precision (PR), recall (RE), specificity (SP), F1 score, Cohen's Kappa (CK), and AUC. 34

Chapter 1

Introduction

This chapter introduces postpartum depression and its association with suicidal ideation in Bangladesh. This encapsulates the motivations behind this research, the aims of creating predictive models, as well as where future contributions may lie in maternal mental healthcare. The chapter also demonstrates the wider work within the context of reproductive health challenges while underscoring the transformative opportunity that machine learning affords to proactively engage vulnerable populations.

1.1 Background

Sexual and Reproductive Health and Rights (SRHR) has long been recognized as a fundamental component of global public health and gender equity initiatives [1, 2]. While significant progress has been made in this field over the years, it is crucial to acknowledge that many challenges still exist, particularly in addressing the mental health aspects of sexual and reproductive experiences. One such challenge is the occurrence of depression and the associated risk of suicidal tendencies among postpartum individuals. During the postpartum period, a mother is overwhelmed with the biological stress in her body and the new responsibility for the well-being of an infant. This makes new mothers more vulnerable to Postpartum depression (PPD) also known as postnatal depression (PND). PPD is a mental health condition that affects mothers typically within one year after childbirth [3]. Even though the exact cause of PPD is unclear, several factors can contribute to individuals having PPD, such as high expectations from society, health issues during and after pregnancy, transitioning into parenthood, newborn's health and development, getting support from friends and family, mental health and wellbeing before giving birth, sleep deprivation and getting the postpartum care new mothers need [4]. While it is a global concern, its numbers impacting mothers vary significantly across different regions of the world. Observed during a systematic review, developing countries and low-middle-income countries (LMICs) face issues with higher PPD rates, reaching up to 20% [5], in contrast to developed nations, where prevalence ranges between 6% and 13% as seen in another systematic review [6]. Bangladesh, as a densely populated South Asian nation, is no exception to this disparity. The population in Bangladesh has a widespread living condition, from rural areas to urban areas. Research within Bangladesh has unveiled alarming PPD rates, ranging from 18% in rural regions [7] to an even more concerning 22% in urban areas [8]. Conversely, the slum dwellers, despite being citizens of cities, suffer from an astonishing 39% of PPD [9]. Cohort studies have also found that biological, psychosocial, economic, hormonal, and social factors and domestic violence by husband and in-laws have been

linked with PPD in Bangladeshi mothers [10, 11].
PPD is not only detrimental to the individual but also impairs families and communities. Research on maternal and infant-related outcomes proves that PPD impacts the personality of the mother and the health of the child [12]. The findings also indicate that PPD affects the physical well-being of the infant and also disrupts the infant's emotional and cognitive development due to the decreased interaction between the mother and the infant as the mother is depressed, and has depression symptoms such as disrupted sadness, agitation, extreme fatigue, preoccupation, suicidal ideation, anxiety, mental confusion, low maternal self-esteem, limited parenting self-efficacy, intense shame and guilt, psychomotor disturbances, and cognitive impairments [11, 13]. Additionally, PPD has other related conditions including hypertension, diabetes, hyperlipidemia, stroke, along with suicidal feelings [14]. The correlation with suicide factors is in a new dimension since new mothers suffering from PPD often want to escape the situation. Evidence suggests that PPD significantly raises the risk of suicide, with a shorter time to suicide following childbirth [14, 15].

1.2 Motivation

Traditional methods of diagnosing PPD and observing suicidal tendencies depend heavily on clinical assessments by healthcare providers and self-reporting by individuals with symptoms of PPD. However, these traditional methods have limitations in the case of Bangladesh. There is a lack of awareness and education of mental health in this country, due to the cultural attributes and stigma surrounding mental health which leads to underreporting of PPD and its associated symptoms in this region [16]. Furthermore, the health system in Bangladesh faces a scarcity of skilled workforce in mental health care and there is a need for more trained professionals to provide adequate care and services, making it challenging to provide timely and effective support to all postpartum individuals who may need it [17].
Artificial Intelligence (AI) has the potential to be a useful tool in helping with the above-mentioned problems. It has been utilized in different sectors of healthcare for data analysis, improving patient engagement, building predictive models, as well as helping healthcare professionals with resource allocation and decision-making [18]. AI applies sophisticated algorithms to 'learn' features from a large volume of healthcare data and then uses the obtained insights to assist clinical practitioners in delivering improved diagnoses. It can also be equipped with learning and self-correcting abilities to improve its accuracy based on feedback [19]. These healthcare professional assisting tools developed using machine learning algorithms propose methods of early detection of PPD [20] and predicting the risk of suicidal tendencies in patients suffering from mental disorders.

1.3 Objectives

This study aims to utilize machine learning algorithms to predict PPD and suicidal tendencies in individuals after childbirth, using symptoms and behavioral data with the potential to create a patient assessment tool that clinicians and healthcare providers can use for early diagnosis of patients with PPD and suicidal ideation. The objectives of this research are as follows:

- To reduce diagnosis delay: PPD can occur anytime within the first year after childbirth. Using this prediction model, healthcare professionals can screen new mothers for PPD when they visit a clinic for their postpartum checkup.

- To improve access to care: The machine learning model will be able to identify patients who are at high risk of suicidal ideation, allowing them to prioritize these patients for immediate counseling, ensuring that they receive the necessary mental health support.
- To assist with resource allocation: Patients labeled as high-risk for suicidal ideation can be recommended to see experienced psychiatrists, ensuring that mental health resources are efficiently allocated and that high-risk individuals receive timely, personalized care for effective intervention.

1.4 Contribution

In this research, a PPD-caused suicide prediction model is proposed with the purpose of early identification of individuals at risk of suicide in Bangladesh.

- The performance of the most effective machine learning algorithm along with their hyper-parameter combination when applied to PPD data was explored.
- One systematic review is published in the Maternal Mental Health (MMH) and Sexual and Reproductive Health and Rights (SRHR) research community.
- One journal as a model being developed is in preparation to be published in the Maternal Mental Health (MMH) and Sexual and Reproductive Health and Rights (SRHR) research community.

1.5 Outline of thesis

The sections of this thesis are organized as follows:

- Chapter 1 introduces the motivation, objective, and outcome of this study.
- Chapter II provides a literature review summarizing the background of PPD with sexual health, reproductive health, and mental health context.
- Chapter III explores the previous studies on PPD data analysis, PPD prediction models, and suicidal tendency prediction models.
- Chapter IV, introduces the methodology behind our PPD-caused suicidal tendency prediction model.
- Chapter V presents the results and evaluates model performance using accuracy, precision, sensitivity, specificity, and the area under the ROC curve.
- Chapter VI, analyses the findings with comprehensive discussion, placing them in a broader context to explore their implication and explores the future scope of our contribution.

Chapter 2

Literature Review

In this chapter, the existing literature related to PPD and its dimensions is explored, with a focus on reproductive health, sexual health, and mental health. While PPD is not directly part of reproductive health or sexual health, it is very closely associated with these health sectors, either as the cause of developing this disorder or as the lead symptoms and risk factors of identifying patients who suffer from this disorder. Also, the reason for exploring these health sectors within the framework of PPD lies in their strong impact on maternal well-being during the postpartum period. Reproductive health covers complications encompassing pregnancy and childbirth. These difficulties can cause significant physical and emotional stress along with hormonal fluctuations, increasing vulnerability to PPD. Moreover, challenges in breastfeeding can contribute to the emotional burden of not being able to feed or bond with the newborn. After childbirth, many women can experience changes in their sexual health that take a toll on their overall well-being. Issues such as decreased libido, pain during intercourse, and body image concerns are some common challenges that mothers face after childbirth. The stress and frustration can increase the risk of or lead to the onset of PPD. All these physical and emotional changes after birth can also affect the relationship with their partner, adding more to the emotional vulnerabilities and potentially triggering depressive symptoms. Following this section, mental health disorders named anxiety and mood disorders will be explored to give a better understanding of how these issues are being addressed with machine learning.

After exploring the mentioned health sectors, the literature review will focus specifically on existing research about PPD, highlighting the problems associated with PPD and the solutions proposed by researchers using artificial intelligence. We will examine the existing literature that has paved the way for our novel research aimed at predicting suicidal tendencies among postpartum patients in Bangladesh.

2.1 Sexual and Reproductive Health and Rights

This section highlights the importance of addressing sexual and reproductive health and rights issues to mitigate the risk of PPD. The integration of Artificial Intelligence and machine learning in reproductive health research offers promising solutions to improve outcomes and support maternal mental health, ultimately contributing to better PPD detection and prevention strategies. This health sector covers various critical issues that women face before, during, and after pregnancy as seen in Figure 1, but for this literature review, only the components of sexual and reproductive health that can contribute to PPD are explored. Advances in technology, such as IoT devices and machine learning, are revolutionizing maternal health monitoring by enabling early detection and intervention for these complications.

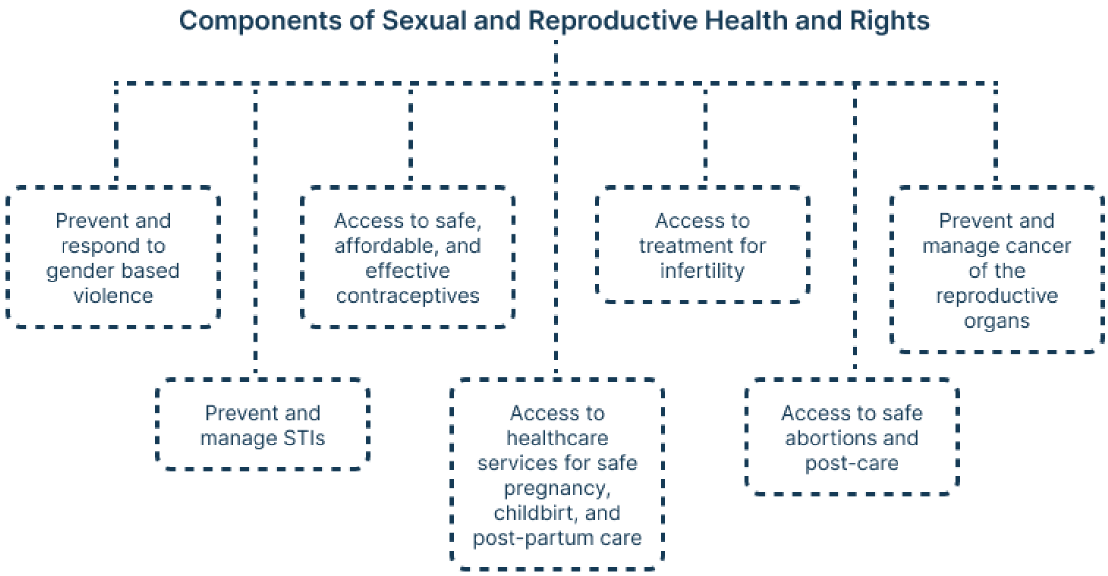

Figure 1: Components of Sexual and Reproductive Health.

2.1.1 Pregnancy health risk

During pregnancy, there are several health risks that women may experience, which necessitate careful monitoring and management. Some common health risks include gestational diabetes, gestational hypertension, preeclampsia, placental abruption, placenta previa, and preterm labor. These complications often require increased medical intervention and monitoring, adding stress and anxiety during pregnancy. Moreover, the additional physical discomfort and fear of potential adverse outcomes for both the mother and the baby can add to the stress, increasing the vulnerability to depression.

For real-time health monitoring of mothers during pregnancy using IoT devices, [21] applies CNN and [22] tests multiple machine learning algorithms on data collected using biosensors. Similarly, [23] uses a CNN model for maternal health monitoring, specifically during ambulatory services, and compares it to other traditional machine learning algorithms. Their model alerts medical staff when high-risk health status is predicted and makes decision-supporting classifications to help clinicians make better diagnoses as well. Moreover, predictive models for assessing health risks during pregnancy using patient medical data are discussed in [24, 25] that predict potential complications that are specific to the patient. In addition to health risk prediction using random forest, [26] suggests the utilization of a blockchain-based maternal health information system to ensure timely assistance and prevent incorrect treatments by healthcare providers. The implementation of blockchain technology offers inherent immutability, preventing unauthorized editing or manipulation of data. To predict the risk of both maternal and fetal mortality by analyzing the factors leading to health risk, [27, 28] tests multiple machine learning classification algorithms on patient medical data, with Random Forest demonstrating the best performance. A similar study to predict the risk of maternal mortality is recommended by [29] using gradient boosting algorithms and [30] using a hybrid ensemble of XGboost and Catboost.

[31, 32, 33, 34] discuss gestational hypertension, which refers to high blood pressure during pregnancy, increasing the chance of preeclampsia and other complications such as damage to the kidneys, liver, and placenta. A classification model for the risk assessment of women suffering pregnancy-induced hypertension is proposed in [31] where multiple machine learning algorithms are tested on patients' medical, lifestyle, and history data. [32] focus on urinary metabolites' data for the prediction of gestational hypertension, where decision tree and gradient boost showed the highest accuracy compared to other models tested. [33] focuses on predicting the risk of women developing preeclampsia by applying multiple machine learning algorithms on patient medical data and comparing the results to help healthcare providers with early detection of gestational hypertension, treatment, and recovery. [34] proposes a similar model using patients' blood work data for the assessment of pre-eclampsia. For studies focusing on the placenta, an organ that develops in the uterus during pregnancy that provides the fetus with oxygen and nutrients while removing waste, Alexander D. Gleed et al. suggested a deep learning-based assessment tool to determine the position of the placenta from ultrasonogram video clips [35]. A placenta-related complication is placenta accreta, also known as placental invasion, which is a condition where the placenta attaches abnormally deep into the uterine wall and even into nearby organs such as the bladder or bowel, depending on the severity. [36] focuses on diagnosing placental accreta and the level of invasion from MR radionics images using CNN, and [37] recommended a similar model using deep dynamic convolutional neural network (DDCNN). Munetoshi Akazawa et al. predict the risk of postpartum hemorrhage, by building a deep learning model comprised of two layers of neural network and an ensemble machine learning using five different classification algorithms for early detection of postpartum hemorrhage and stratification of high-risk patients to help health care providers with timely prevention and treatment [38].
Preterm birth is defined as birth before completing 37 weeks of pregnancy, which can increase the risk of neonatal mortality and morbidity. Yina Wu et al. created a time series hybrid model using prenatal examination and medical records, to predict the estimated delivery date, by applying Gradient Boosting Decision Tree with Gated Recurrent Unit (GBDT–GRU) [39]. Similarly, to predict the estimated week of delivery and classify high-risk pregnancies for preterm birth, [40] tested multiple types of machine learning algorithms, where extreme gradient boosting showed the best results. [41] also applied multiple machine learning algorithms on prenatal patient data, medical history, and family history to develop a preterm birth prediction model where multi-layer perceptron (MLP) was the best-performing classification algorithm. [42] took a different approach to predict the risk of spontaneous preterm birth by applying Deep Neural Networks (DNN) on cervix ultrasound Shear Wave Elasticity (SWE) imaging, a technique used to measure cervix elasticity.

2.1.2 Childbirth complications

The experience of childbirth can be a significant factor in the onset of PPD. Emergency C-sections, induced labor, or other unforeseen complications can lead to feelings of loss of control and fear during delivery. Additionally, the physical pain and recovery process, combined with the responsibilities of caring for a newborn, can be overwhelming, leaving new mothers vulnerable to PPD, as they may struggle to cope with the demands of their new role and the recovery process simultaneously.
Childbirth can take place by normal delivery, by induced delivery using hormone injection, and by cesarean section. [43] focuses on birth mode prediction using bagging ensemble classifiers on survey data to help healthcare providers with early detection of patients who may need cesarean section

and reduce the number of unnecessary surgeries that can lead to complications for both the mother and child. Similarly, [44] applies Random Forest with hyperparameter optimization tuning using the Henry Gas Solubility Optimization (HGSO) algorithm to predict the chances of a cesarean section. Tingting Hu et al. propose a prediction model for early classification of the success of induced birth within 3 days of administering Oxytocin [45]. They applied multiple machine learning algorithms on maternal and fetal characteristics data, where logistic regression showed the best performance and accuracy. Through the use of deep learning on ultrasound SWE imaging, an attempt to predict the success of induced labor has been discussed in [42]. A high percentage of maternal mortality rate is due to the absence of skilled assistance during childbirth. [46] and [47] suggest a classification model by applying and testing multiple machine learning algorithms for early stratification of women who need skilled assistance during childbirth to allocate resources in situations with scarcity of skilled professionals. Another predictive model presented by Brooke Tesfaye et al. presents the use of multiple classification algorithms to identify the determinants of the use of skilled delivery services from survey data, and they found that a decision tree framework, namely, J48, performed the best in identifying women who were more likely to give birth outside a health facility [48]. [49] uses LASSO regularized logistic regression to develop a predictive model for assessing the vulnerability of machine learning models to adversarial attacks by testing it using One-At-a-Time (OAT) adversarial attacks.

2.1.3 Pregnancy loss

Pregnancy loss, including miscarriage and stillbirth, has a profound emotional impact. For women who experience a subsequent pregnancy, the fear of another loss can overshadow the joy of the new pregnancy, leading to chronic anxiety and stress. Additionally, societal pressures and a lack of understanding about the depth of such loss can leave mothers feeling isolated and unsupported, further increasing the risk of depression after childbirth.

Miscarriage is known as the loss of pregnancy within 20 weeks of gestational age. Jingying Huang et al. proposed a model that applies and tests six different machine learning algorithms to predict threatened miscarriage in the first trimester of pregnancy through the assessment of patients' AEA, P4, and β-hCG hormone levels [50]. Logistic regression gave the most accurate results to help clinicians provide active treatments for avoiding miscarriage. Alternatively, the authors of [51] focus on the use of the Artificial Neural Network (ANN) for predicting the risk of spontaneous miscarriage from Raman spectroscopy data of endometrial tissue, since it performed best compared to other models. [52] suggests a similar predictive model in an IoT mobile application where a K-means algorithm is applied to user-provided data and real-time sensor data to predict the risk of miscarriage. Stillbirth is known as a fetus having no signs of life before extraction from the mother after 20 weeks of gestational age. [53] presents multiple machine learning algorithms as a classification model along with ensemble learning to predict the risk of early stillbirth, late stillbirth, and preterm birth. Light gradient boosting machine (LGBM) and weighted average (WA) ensemble gave the best results on both the datasets used.

2.1.4 Sexual and gender-based violence

Sexual violence is an alarming and widespread problem that tragically impacts people from various backgrounds, irrespective of socioeconomic standing. This distressing issue encompasses a spectrum of harmful acts, such as physical assault, domestic violence, and harassment, which harm the physical, emotional, and psychological state of survivors, including the onset of PPD.

Physical harassment refers to any form of harmful or abusive behavior that involves the use of physical force or violence toward individuals, with a particular focus on women. This includes physical assault, domestic violence, and workplace harassment that cause physical harm or injury. In addressing physical mistreatment, several research papers have explored the application of machine learning techniques to predict and detect instances of mistreatment. For instance, [54] and [55] utilize Keras, a deep-learning API, to analyze video data and predict occurrences of sexual assault. In [54] a classification model that combines Convolutional Neural Networks (CNN), Recurrent Neural Networks (RNN), and Long Short Term Memory (LSTM) is employed to classify and predict instances of sexual harassment. On the other hand, [55] focuses on using various CNN architectures to identify sexual harassment in the workplace based on video footage.

Furthermore, [56], and [57] emphasize the prevalence of using multiple classifiers in addressing physical mistreatment. Lateef et al. utilize four machine learning algorithms to classify women who are more vulnerable to intimate partner violence [56]. Another model discussed in [57] employs three types of machine learning classifiers to detect instances of physical mistreatment against women.

To detect various forms of online harassment, Emad et al. built a model using eight machine learning classifiers and sentiment analysis techniques to detect cyber harassment and cyberbullying [58]. Additionally, Islam et al. introduced eight machine-learning classifiers for detecting sexual harassment in Bangla text and achieved improved accuracy [59]. Finally, Alam et al. presented four combinations of machine-learning techniques to detect hate speech related to sexual violence and produced the classification model with the highest accuracy [60].

Victims of sexual violence can access necessary support for harassment through counseling, support groups, legal advocacy, medical care, crisis hotlines, shelter services, and other online support. Conversational agents are a form of online support that exists to maintain privacy and provide appropriate information remotely. Fixed responses from conversational agents sometimes leave the user without the answer they are looking for. Artificial Intelligence solves the problem of fixed responses in conversational agents by leveraging advanced Natural Language Processing (NLP) techniques, allowing for more dynamic and context-aware interactions. [61] introduces the use of NLP as a conversational agent, namely, Law-U, to communicate with sexual violence victims and survivors and provide them with legal guidance.

2.2 Mental health

While mental health is a vast topic with a diverse set of disorders as shown in Figure 2, this section highlights some common mental health issues such as anxiety and mood disorders because of their similarity in symptoms with PPD. Anxiety disorders, characterized by pervasive worry and fear, can significantly contribute to PPD, especially when pre-existing conditions intensify during and after pregnancy. Mood disorders, such as major depressive disorder and bipolar disorder, similarly intersect with PPD, highlighting the necessity for advanced detection and treatment methods. The integration of AI and machine learning in mental health research offers solutions to improve PPD detection and prevention strategies.

2.2.1 Anxiety disorders

Anxiety disorders, characterized by excessive worry and fear, can be a major contributing factor to PPD. Women with pre-existing anxiety disorders may experience heightened anxiety during preg-

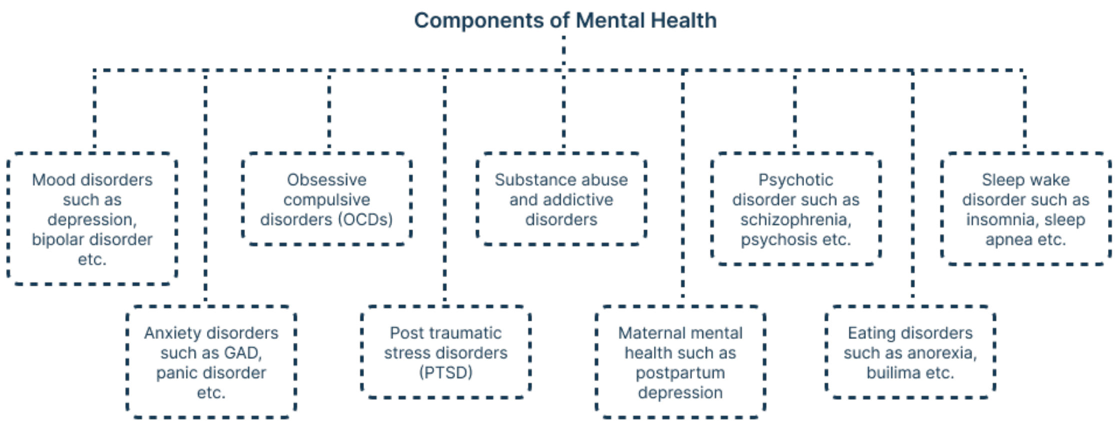

Figure 2: Components of Sexual and Reproductive Health.

nancy and postpartum due to concerns about the baby's health, parenting abilities, and changes in their own identity and lifestyle, making it difficult for new mothers to manage daily tasks and bond with their babies.

Concerning available literature solely dedicated to anxiety, the selected research demonstrates a rather dissimilar focus and significance of the selected area of research. Using a Hidden Markov Model (HMM) and Logistic Regression (LR), Ryu et al. try to predict the binary outcome of whether a patient is clinically or non-clinically anxious [62]. Another investigation utilized MLP, LR, and RF on EEG data, with RF demonstrating superior accuracy in detecting trait anxiety [63]. In a different context, SVM and KNN algorithms were applied to an EEG dataset, to detect four levels of anxiety [64]. Various machine learning (ML) approaches have been employed for anxiety diagnosis. Meng et al. developed a TakagiSugeno-Kang (TSK) fuzzy system to automatically identify anxiety in college students by extracting deep features using CNN from input data [65]. Alternatively, a Bayesian Neural Network (FE-BNN) anxiety prediction model was proposed and evaluated using data collected by the online YODA tool [66]. Additionally, a study used LR, RF, and naive Bayes (NB) to predict specific anxiety symptoms and virtual reality sickness during treatment, demonstrating the models' ability to support diagnosis with minimal participant bias [67]. A range of machine learning models has been utilized across studies focusing on mental health prognosis, risk prediction, and screening. For predicting panic attacks, a proposed model utilized multiple algorithms, and focused particularly on regularized greedy forests on a combination of questionnaire data, environmental, and real-time physiological data from wearable sensors [68]. Lastly, a hybrid machine learning model for predicting anxiety in conflict zones used SVM, MLP, and RF algorithms [69].

Some studies were conducted focused on comorbidities alongside anxiety disorder. One study developed a prediction model comparing the performance of various ML algorithms, for assessing school children for depression and anxiety, where Random Forest achieved the highest accuracy [70]. Another investigation employed LASSO and SVM to analyze EEG signals, developing a diagnostic tool capable of accurately identifying patients with panic disorder and MDD [71]. LSTM was used in yet another study where Watanabe et al. attempted to predict depression and anxiety in the working population using digital biomarkers derived from physical activity and working conditions [72]. A novel approach utilized Haar-cascades for face detection and CNN for emotion classification from image data to predict mental disorders [73]. Furthermore, another study compared the

performance of multinomial LR and NB classifiers in predicting the onset and course of depression and anxiety [74]. A patient assessment model was developed using multiple algorithms to classify participants based on anxiety, depression, and stress severity levels [75]. In a screening context, to identify anxiety, along with depression, one study introduced the DAC Stacking model, utilizing LSTM and CNN and their comorbidity from multi-label text data extracted from Reddit [76], while another study employed LR, NN, and RF for immune-mediated inflammatory disease (IMID) patients [77].
Two studies were found focused on social anxiety. One study attempted to identify social anxiety in young adults where multiple algorithms were applied to a multi-sensor wristband dataset [78], while another study explored the use of smartphone sensors to predict social anxiety symptom severity, employing XGBoost on accelerometer data, text message biomarkers, and call biomarkers [79]. Only one study was found regarding Obsessive-compulsive disorder (OCD), a hybrid model incorporating LSTM, Hopfield Network, and Bi-Directional RNN that was developed to predict the risk of OCD by learning semantic characteristics from clinical data items [80].

2.2.2 Mood disorders

PPD is a type of mood disorder and having a history of other mood disorders such as bipolar disorder or major depressive disorder before childbirth can also increase the risk of getting PPD. The outspread of depression shows the importance of possible causes and factors that require more attention in comprehending and managing different types of the disorder among various target groups. However, even where depression has been inferred through other methods, to the best of our knowledge, researchers have used diverse approaches to detect depression, by analyzing data from social media platforms [81, 82, 83, 84, 85, 86, 87], where Random Forest [81] and NLP approaches [82, 87] were used. Moreover, research also focused on the different styles of using deep learning including utilizing both bi-LSTM and CNN [83, 84, 85], using questionnaires with CNN [88] and stacking algorithms [89], using combined CNN and ANFIS for the data collected from wearable sensors [90], using different algorithms to recognize the genes related to depression [91]. Survey data was identified to be employed for identifying depression through GBDT methods [92], Random Forest [93], as well as the use of LSTM with CNN and NLP methods [94, 95].
Some studies leveraged the use of wearable sensors and mobile applications for assessing the level of depression [96, 97, 98]. Asare et al. used self-reported and wearable data with multiple AI algorithms for depression classification in one study [96] and utilized smartphone data and questionnaires with various AI algorithms for predicting depression in another study [98] while Dogrucu et al. employed the Moodable app for distinguishing depression and suicidal ideation with diverse AI algorithms [97]. Furthermore, some studies have used questionnaires and survey data to assess the depression levels of participants [99, 100, 101, 102, 103, 104]. Lyall et al. employ different regression algorithms to identify sleep and circadian predictors [99], Zhang et al. compare multiple algorithms for depression detection [100], and Siddiqui et al. used the combination of ten machine learning and two deep learning algorithms to predict depression reasons among students [101]. Two studies utilize Random Forest to assess depression levels, with one focusing on health-seeking behavior in major depressive disorder patients [102], and the latter on factors of depressive symptoms in Chinese elderly individuals [103]. Conversely, Uddin et al. employ LSTM and RNN to detect self-perceived symptoms in text [104].
Identifying depression is a somewhat complex task that holds significant interest and catches the attention of curious researchers and while developing different methods, some researchers have turned to experimental AI models to diagnose depression in individuals. In some studies, EEG

datasets were utilized [105, 106, 107]. One study utilized linear discriminant analysis (LDA) and linear support vector machine (SVM) to predict depression in drug-naive patients [105], one employed multiple algorithms [107], while another study introduced a hybrid model with bi-LSTM for depression diagnosis [106]. Furthermore, researchers have also leveraged data extracted from wearable devices and mobile applications for diagnostic purposes [108, 109, 110]. They individually utilized different approaches like DCNN [108], a combination of multiple algorithms [109], and a neural network with weighted fuzzy membership functions (NEWFM) [110] on data extracted from wearable sensors to make accurate diagnostic decisions.

Screening for depression is a critical process aimed at early identification and intervention for individuals. Thus, studies have made use of various types of data to be precise with its identification, by incorporating AI models on them [111, 112, 113, 114, 115]. Wani et al. focused on using deep learning to identify depressive symptoms in netizens through social media content [111], while Geng et al. optimized initial Major Depressive Disorder (MDD) screening using 5-minute ECG signals during sleep through multiple algorithms [112]. Another study aimed to detect depression risk in a working population with digital biomarkers from wearables [113]. Additionally, [114] predicted depression scores in adolescents using a mix of linear and non-linear machine learning techniques. In contrast, [115] enhanced MDD screening using multiple algorithms in the EarlyDetect mobile app.

The treatment of depression is usually a combination of different techniques due to its complex nature. Accordingly, studies have shown promising results in tracking, monitoring, and providing optimized treatment [116, 117, 118]. One study utilizes multiple algorithms to predict MDD using app, sensor, and questionnaire data [116], while another study observes ketamine effects on MDD patients by applying multiple AI algorithms on genome-wide sequencing data [117]. Additionally, Cousins et al. applied LSTM and a fully connected network to predict optimal medication for individuals with diagnosed depression [118].

In the context of measuring the level of depression with Bengali text data, one of the studies employ quite a diverse scenario of multiple instantiations of machine learning and deep learning [119]. On the other hand, another type of study refines the PSYCHE-D model using various algorithms in real-time data from a person for predicting future escalation of the depression severity up to one year [120].

Additionally, a separate investigation utilizes a Gated Recurrent Unit (GRU) within a deep multitask framework, employing machine learning models to detect depression, sentiment, and multi-label emotion from suicide notes [121].

2.3 Postpartum depression (PPD)

This section reviews the existing research around PPD detection using machine learning and artificial intelligence approaches.

Andersson et al. sought to use multiple machine-learning techniques to forecast the likelihood of postnatal depression. The study used a broad array of variables, demographics, medical history past and present, psychosocial factors, and pregnancy attributes. Specific examples of applied algorithms include ridge regression, LASSO regression, distributed random forest, extreme randomized forest, gradient-boosted machines, stacked ensemble, and naive Bayes. PPD was noted to be a challenging task for risk assessment focusing on the multi-modal patient data incorporation and the use of the ensemble approaches to enhance the prediction accuracy and overcome the concerning issue of overfitting [122].

Similarly, Y. Zang et al. concentrated on creating and validating machine learning-based risk assessment models of PPD among pregnant women. This research also incorporated many aspects of participants such as their health history, pregnancy characteristics, and demographic details. Some of the feature selection approaches included were Sequential Feature Selection and expert selection. Various algorithms, including logistic regression, random forest, decision trees, XGBoost, and multilayer perceptron, were evaluated. The findings emphasize the importance of comprehensive feature selection for reliable predictions [123]. Moreover, another study by W. Zang et al used both expert selection and random forest-based filter feature selection methods to optimize feature selection to train their machine learning models to predict the risk of developing PPD. The study employed random forest and support vector machine (SVM) algorithms for the classification and prediction [124]. Furthermore, Wakefield et al. aimed to predict patients at risk of PPD using electronic medical record (EMR) data collected during pregnancy. Features incorporated various factors such as pre-pregnancy mental health history, medication use, socioeconomic status, and healthcare-provider interactions. Employing random forest and logistic regression algorithms, the research demonstrated the feasibility of utilizing routine EMR data for early identification of PPD risk [125].

With a different approach, Betts et al. attempted to predict postpartum psychiatric admissions using machine learning techniques, namely boosted trees and elastic net algorithms, the study highlighted the significance of mental health history, prenatal care, and sociodemographic factors in predicting postpartum psychiatric admissions. Such findings underscore the potential of machine learning methods in identifying high-risk individuals, thereby facilitating targeted interventions and resource allocation to mitigate postpartum psychiatric complications and improve maternal mental health outcomes [126].

2.4 Conclusion

In this thesis, the relationship between PPD and other fields of health, including sexual, reproductive and mental health is discussed and the different ways through which these domains affect maternal health in the postpartum period is analyzed. Some of the many reproductive health issues include infertility, pregnancy health risks, and childbirth complications among other issues that can force a mother towards PPD. Likewise, sexual health problems such as loss of sexual desire, and body image indicators also create emotional vulnerabilities to PPD. Additionally, the exploration of mental health, particularly anxiety and mood disorders, reveals that pre-existing mental health issues increase greatly the risk of getting this particular disorder and if not treated in time, can lead to self-harm and suicide. The integration of artificial intelligence (AI) and machine learning (ML) into these health sectors presents scope for positive results and provides support for maternal mental health. These technologies have the potential that enable healthcare providers to make informed decisions, ensuring efficient interventions that can mitigate the risks associated with PPD. Focusing on PPD, the literature also highlights significant advancements in supporting patients using AI and ML for its detection, and prediction for severity, treatment, or outcomes. The advancements in this field pave the way for this novel research, which aims to predict suicidal tendencies among postpartum patients in Bangladesh, utilizing AI techniques.

Chapter 3

Exploration of Postpartum depression and Suicide ideation prediction

In this chapter, the data associated with PPD is explored, such as symptoms and risk factors along with the scales researchers have used to detect these disorders in new mothers. In doing so, we got an understanding of the trajectory of PPD and identified key variables that must be a part of any predictive model. After analyzing PPD data, the following section will explore machine learning models for predicting suicide, where the type of data used, the types of machine learning algorithms, and evaluation metrics are reviewed for studies conducted over the last 5 years. In that respect, this section is going to help us dive into what are the current state-of-the-art techniques for predicting suicidal tendencies and inform our choices on algorithms and models. This will help us develop an accurate methodology to portray PPD, and its associated Suicidal ideation using data from Bangladesh.

3.1 PPD data analysis

Understanding the complexity of factors contributing to PPD is essential for developing targeted interventions and support systems to mitigate the burden of PPD. In this section, we review existing literature on PPD screening tools, the symptoms of patients with PPD, and the risk factors associated with this disorder based on countries and their income levels.

According to a literature review of screening and prevention, Hana Andrina extensively reviewed 7 screening tools for PPD as mentioned in Table 1. In this research, it has been observed that the Edinburgh Postnatal Depression Scale (EPDS) and the Postpartum Depression Screening Scale (PDSS) outperform other scales in their sensitivity and specificity. It is also recommended to use EPDS as it has a lower number of questionnaire items and takes a fairly shorter time to complete during surveys [127]. Most of the PPD prevalence and data analysis studies discussed in this section also use EPDS as their PPD screening tool. Another advantage of the tool is its availability in multiple languages, including Bangla.

The decision was made to explore these data on the economic state of these countries because LICs - LMICs tend to have a similar cultural setting as Bangladesh and would give a better prediction when used on patients with the suicidal tendency model that this study is developing. All the

Table 1: Commonly used screening tools for PPD

Screening Tools	No. of items	Time to complete (min)
Edinburgh Postnatal Depression Scale	10	<5
Postpartum Depression Screening Scale	35	<10
Patient Health Questionnaire - 9	9	<5
Beck Depression Inventory	21	<10
Beck Depression Inventory-II	21	<10
Center for Epidemiologic Studies Depression Scale	20	<10
Zung Self-Rating Depression Scale	20	<10

symptoms associated with PPD are provided in Table 2 and risk factors associated with PPD are provided in Table 3.

In LMICs like Bangladesh, studies have identified various risk factors contributing to PPD. These include demographic factors such as age, education level, employment status, household income, household resources, and neighborhood characteristics. Additionally, experiences of domestic violence, history of child loss, inadequate social support, and stressful life events have been associated with higher PPD prevalence. Lack of access to healthcare services and limited resources exacerbate the burden of PPD in these settings [9, 128]. In LMICs like India, numerous risk factors have been identified, including maternal age, history of mental disorder, stress, and unwanted pregnancy. Cultural factors such as low self-esteem and negative birth experiences also contribute to PPD prevalence. Additionally, socioeconomic factors such as education level, financial status, and social support play significant roles [129, 130]. In Vietnam, another LMIC, studies have focused on symptoms of PPD observed by family members. These include anxiety, changes in eating habits, emotion dysregulation, and sleep disturbances. Understanding these symptoms is vital for early detection and intervention [131].

Moving to upper-middle-income countries (UMICs) like China and Iran, PPD risk factors encompass a broad range of biological, psychological, and social factors. These include factors such as the baby's gender, history of mental disorders, domestic violence, and hormonal imbalances. Socioeconomic factors like education level, employment status, and social support continue to be significant contributors to PPD prevalence [132, 133]. In UMICs like Malaysia and Turkey, lifestyle factors such as consumption of unhealthy food, high BMI, and lifestyle changes have been associated with PPD symptoms. Additionally, factors like antenatal care, domestic support, and family history of mental disorders play crucial roles in PPD prevalence [134, 135]. In HICs like the USA, studies have identified a wide array of risk factors for PPD, including socio-economic factors such as education level and employment status. Social support and recent life events also influence PPD prevalence. Moreover, psychological factors like stress and emotion dysregulation are significant contributors to PPD symptoms [136, 137].

Table 2: Symptoms of patients with PPD

Symptoms	Income level			
	LIC	LMIC	UMIC	HIC
Agitated				[136]
Anger				[136]
Anxiety		[131]		
Consumption of unhealthy food			[134]	
Eating habit		[131]		[136]
Emotion dysregulation		[131]		[136]
Fatigue				[136]
Guilt				[136]
Having illusions		[131]		
High BMI			[134]	
Ignoring children		[131]		
Lifestyle changes			[134]	
Concerned about children		[131]		
Sad without reason		[131]		
Separated from outside life		[131]		
Sleep disturbance		[131]		
Stress	[9], [128]	[131]		[136]
Suicidal ideation				[136]
Unusual Behavior			[134]	[136]

Table 3: Risk factors associated with the prevalence of PPD

Risk factors	Category	Income level			
		LIC	LMIC	UMIC	HIC
Age	Demographic	[9], [128]	[129], [130]	[132]	
Age of last-child	Demographic	[9]			
Baby gender	Demographic		[129], [130]	[132], [133]	
Immigration status	Demographic		[130]		
Marital status	Demographic		[130]		
Multiple children	Demographic	[9]	[129], [130]	[132]	
Cortisol	Health			[133]	
Genetics	Health		[130]		
High oxytocin during pregnancy	Health			[132]	
Low estrogen	Health			[132], [133]	
Anaemia	Health		[129]	[132]	[137]
Antenatal checkup	Health			[135]	
C-section	Health		[130]	[132], [133]	[137]
Chronic illness	Health		[130]		

Risk factors	Category	Income level			
		LIC	LMIC	UMIC	HIC
Gestational diabetes	Health		[130]	[132]	[137]
High-risk pregnancy	Health			[132]	
Neonatal complication	Health	[9]			
Not breastfeeding	Health			[132]	
Nutrition deficiency	Health			[132]	
Obesity	Health		[130]		
Premenstrual syndrome	Health			[132]	[137]
Preterm birth	Health	[9]	[130]	[132], [133]	[137]
Sleep deprivation	Health			[132], [133]	
Thyroid dysfunction	Health			[132]	
Vitamin B6 deficiency	Health			[132]	
Vitamin D deficiency	Health		[130]		[137]
Diet	Lifestyle			[132]	
Physical activity	Lifestyle			[132]	
Substance abuse	Lifestyle		[130]	[132]	
Body image dissatisfaction	Psychological		[130]	[133]	
History of Child Loss	Psychological	[9]			
History of mental disorder	Psychological	[9]	[129], [130]	[132], [133], [135]	
History of sexual abuse	Psychological		[129]	[132]	
Low self-esteem	Psychological		[129], [130]	[132]	
Negative birth experience	Psychological		[129], [130]	[132]	
Recent life events	Psychological	[128]			
Sexual abuse	Psychological				[137]
Stress	Psychological		[129], [130]		
Family history of mental disorder	Psychological			[135]	
Unwanted pregnancy	Psychological	[9]	[129]	[132], [133]	[137]
Domestic support	Social	[128]		[135]	
Domestic violence	Social	[9]	[130]	[133], [135]	[137]
Household members	Social	[9]			
Social support	Social	[9], [128]	[130]	[132], [133], [135]	
Neighborhood	Socioeconomic	[128]			

Risk factors	Category	Income level			
		LIC	LMIC	UMIC	HIC
Education level	Socioeconomic	[9], [128]	[130]	[132]	
Employment status	Socioeconomic	[9], [128]		[132]	
Financial status	Socioeconomic		[130]	[132]	
Household resources	Socioeconomic	[128]			
Husband education level	Socioeconomic	[128]			
Husband occupation	Socioeconomic	[128]		[135]	
Monthly income	Socioeconomic	[9], [128]			
Number of living rooms	Socioeconomic	[9]			

3.2 Suicidal tendency prediction models

In this section, recent studies in the context of suicidal prediction on machine learning-based models are reviewed based on the data type used, algorithms applied, and evaluation metrics considered. Understanding these elements will help to develop the methodology to predict PPD and suicidal ideation. For assessing the risk of suicide, researchers have employed various machine learning algorithms on linguistic features extracted from social media. One study utilized three machine learning algorithms to classify individuals with suicidal tendencies based on these linguistic features [138]. Another investigation focused on distinguishing between self-harm and suicide attempts, employing Random Forest to identify unique risk profiles of individuals [139]. Text data from Twitter were also processed by mental health professionals, applying multiple machine learning algorithms to predict suicide ideation [140]. In a different study, Wang et al. utilized the UK Biobank to predict short-term and long-term suicidal behavior in the middle-aged population using Light Gradient Boosting Machine [141]. The impact of COVID-19 on individuals prompted a study that employed multiple machine-learning algorithms on survey data to predict suicidal ideation [142]. Min et al. took a unique approach by utilizing acoustic voice features of psychiatric patients to assess suicidal tendencies [143]. Cho et al. used national medical checkup data from the National Health Insurance Sharing Service in South Korea, running Random Forest to predict suicide risks among individuals [144]. Another study analyzed the UK Mental Health Services Dataset to find the correlation between hospital stays with risks of self-harm among young adults, implementing and comparing multiple machine learning algorithms [145]. Furthermore, researchers have evaluated children and patients with chief complaints based on the ICD-10-CM code and to predict the risk of suicide using Lasso, Logistic Regression, and Random Forest [146]. In another study, researchers intended to propose an early detection of vulnerable soldiers before deployment and thus contribute to the formation of a risk model for suicide attempts during and immediately following military deployment using a machine learning approach [147]. Finally, unstructured data was used to enhance the prediction of suicidal and self-injurious events in correctional settings through machine learning techniques [148].

To detect the risk of suicide, researchers have implemented machine learning algorithms to enhance text representation and identify suicidal ideation and mental disorders from social content [149]. Another study utilized machine learning to classify social media posts into high, moderate, or no risk of suicide [150]. Kodati et al. ran an investigation that focused on identifying negative emotions in suicidal postings on social media to describe an individual's mental health, employing machine learning and deep learning techniques [151]. Wu et al. developed a suicidal ideation detection model using Long Short-Term Memory (LSTM) and Bidirectional Encoder Representations from Transformers (BERT) by examining text data from social media articles [152]. In a similar study, machine learning and neural networks were used to identify those at risk of Opioid Use Disorder (OUD) if they were thinking about suicide [153]. One more study proposed an automated detection model of self-harm using deep learning to identify people likely to re-engage in self-harm episodes in the future [154]. Additionally, a study applied machine learning algorithms to increase the rate of correct identification of suicidal thoughts in Korean adolescents [155].

For screening the risk of suicide, researchers utilized voice data from patients to label them as Healthy Controls (HC), Suicidal Ideation (SI), or Suicide Attempt (SA) using machine learning techniques [156]. Another study attempted to predict suicidal ideation among parents, by using clinical notes from Electronic Health Records (EHR) [157]. Additionally, one study identified patients with a higher risk for hospital readmission due to self-harm and suicidal behavior. This was achieved by predicting suicidal behavior using machine learning on historical data collected up to one year before and including the hospital admission index [158]. Additionally, to provide support regarding suicide, researchers developed a suicide prevention and support system named Thoth, which is a mobile application. Thoth utilizes wearable sensor data, social media data, and questionnaire information to predict the risk of suicidal ideation. This was achieved by applying and comparing multiple machine learning algorithms, among which the Gated Recurrent Unit (GRU) showed the best performance [159].

3.2.1 Type of data used for suicidal ideation prediction

Mental health research with artificial intelligence uses a variety of data including questionnaires, surveys, social media data, medical data, wearable sensor data, audio speech data, images, videos, internet search activity, notes or journals, and clinical data from EEG, ECG, and MRI. For assessing and detecting risk of suicide, most commonly used data according to the last 5 years of research is social media data [138, 140, 149, 150, 151, 152, 153, 159], medical data [144, 145, 146, 148, 154, 157, 158] questionnaire or survey data [139, 141, 142, 147, 155] and speech data in the form of audio and text [143, 156] as seen in figure 3.

3.2.2 Machine learning algorithms used to train prediction models

Many machine learning (ML) algorithms have been used to detect, assess and screen suicidal tendencies such as traditional techniques, including Random Forest, SVM, Logistic Regression and Decision Trees; advanced algorithms like Gradient Boosting, Naive Bayes, Bayesian Networks; deep learning algorithms such as Gates Recurrent Units (GRU), Convolutional Neural Networks (CNN), Long Short-Term Memory (LSTM), and Transformer encoders.

The type of data used significantly influences the choice of machine learning algorithms as some algorithms work better with one data type than another. It is observed questionnaire or survey data, which is structured and often numeric or categorical, is commonly analyzed using algorithms like Logistic Regression, Random Forest, Decision trees, Gradient boosting, and SVM. Social media data, which is unstructured text, uses deep learning algorithms like CNN, LSTM, and GRU which can process text. Medical data, which can include structured clinical data to health records, is used with both traditional and deep learning algorithms, including Random Forest, SVM, Logistic

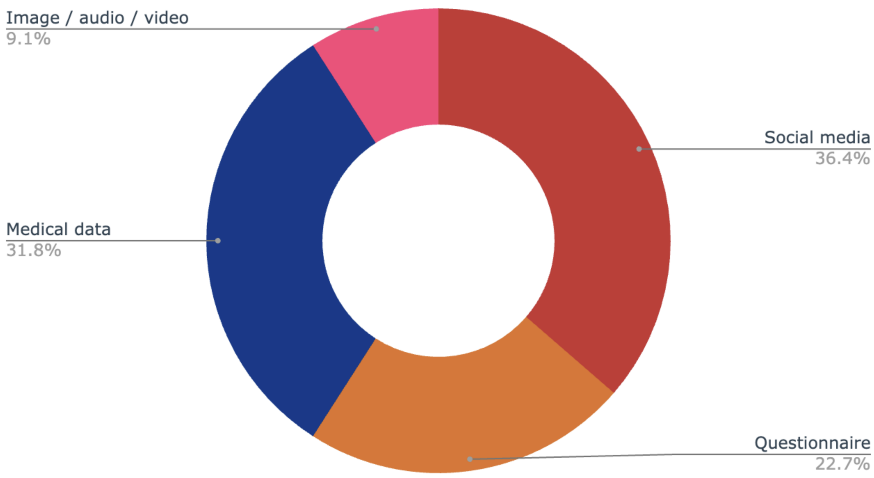

Figure 3: Chart showing the studies using different types of data to train suicidal tendency prediction models.

Regression, CNN, Transformer Encoders, and LSTM. A detailed representation of machine learning algorithms used to train suicidal tendency prediction models discussed above is provided in Table 4.

3.2.3 Metrics used to evaluate prediction models

The most commonly used metrics to evaluate machine learning prediction models are accuracy, recall, precision, specificity, F1 score, and area under the ROC curve as seen in Table 5.
Accuracy measures the proportion of correctly classified instances out of the total instances by providing a straightforward interpretation of overall model performance. It might not be suitable for imbalanced datasets, where it may overestimate performance, but is still widely used and can be misleading when classes are imbalanced.

$$\text{Accuracy} = \frac{TP + FP}{TP + FP + TN + FN} \tag{3.1}$$

Recall measures the proportion of true positive instances that were correctly identified. It is useful when minimizing false negatives is important but may not be sufficient as a standalone metric, especially when false positives need to be minimized as well. Recall is important in applications where missing positive instances are costly, for example, in medical diagnosis, it's crucial to correctly identify all instances of a disease.

$$\text{Sensitivity / Recall} = \frac{TP}{TP + FN} \tag{3.2}$$

Precision measures the proportion of true positive instances out of all instances predicted as positive. It is useful when minimizing false positives is important. It may not be sufficient as a standalone metric, especially when false negatives need to be minimized as well. Precision is important in applications where false positives are costly.

Table 4: Studies using different types of machine learning algorithms to train suicidal tendency prediction models.

Reference	RF	SVM	LR	DT	NB	BN	GB	RNN	CNN	AE
[138]	✓	✓					✓			
[139]	✓									
[140]	✓	✓								
[141]							✓			
[142]	✓	✓	✓				✓			
[143]	✓	✓					✓			
[144]	✓									
[145]	✓	✓				✓		✓	✓	
[146]	✓		✓							
[147]	✓					✓	✓			
[148]										✓
[149]								✓	✓	
[150]	✓	✓					✓			
[151]								✓	✓	
[152]								✓		
[153]	✓	✓	✓						✓	
[154]								✓	✓	
[155]			✓	✓	✓					
[156]		✓								
[157]		✓	✓		✓				✓	
[158]				✓						
[159]								✓		

$$\text{Precision} = \frac{TP}{TP + FP} \tag{3.3}$$

Specificity measures the proportion of true negative instances that were correctly identified and is particularly useful when minimizing false positives is important. Specificity complements sensitivity (recall), providing insights into the model's ability to correctly identify negative instances .

$$\text{Specificity} = \frac{TN}{TN + FP} \tag{3.4}$$

F1 Score is the harmonic mean of precision and recall, providing a balance between the two metrics. It is useful when there's an uneven class distribution but may not be ideal for situations where costs are associated with false positives and false negatives.

$$\text{F1 Score} = 2 * \frac{Precision * Recall}{Precision + Recall} \tag{3.5}$$

Table 5: Different types of metrics used to evaluate suicide prediction models.

Reference	Accuracy	Recall	Precision	Specificity	F1 score	AUC -ROC
[138], [140], [151], [153], [154], [157]	✓	✓	✓		✓	✓
[139], [141], [144], [148]		✓		✓		✓
[145], [158]	✓	✓		✓		✓
[149], [150]	✓	✓	✓		✓	
[146], [152]	✓	✓		✓		
[143]	✓	✓	✓	✓	✓	
[147]		✓				✓
[155]		✓		✓		
[156]	✓				✓	
[159]	✓					
[142]						✓

Area Under the Receiver Operating Characteristic (ROC) Curve summarizes the model's performance across various threshold settings by providing a comprehensive evaluation of the model's performance across all possible classification thresholds.

3.3 Conclusion

This chapter provided an extensive analysis of PPD data and explored various machine learning models used to predict suicidal tendencies. By examining the symptoms, risk factors, and screening tools associated with PPD, we identified critical variables essential for developing effective predictive models. The Edinburgh Postnatal Depression Scale (EPDS) emerged as a particularly effective screening tool due to its sensitivity, specificity, and availability in multiple languages, including Bangla. We reviewed literature from different income-level countries to understand the varying risk factors and symptoms of PPD based on country, culture, and economy. This cross-country analysis highlighted demographic, health, psychological, social, and socioeconomic factors that influence PPD prevalence. Shifting the focus to suicidal tendency prediction, the chapter explored recent studies using various machine learning algorithms. These studies used various data types, including social media data, medical records, questionnaires, and speech data, to predict suicidal ideation and behavior. The use of algorithms such as Random Forest, Logistic Regression, SVM, and advanced techniques like LSTM, BERT, and CNN performed well in this context. Evaluation metrics like accuracy, recall, precision, specificity, F1 score, and AUC-ROC were used to assess the performance of these predictive models. These metrics provided a balanced view of model effectiveness, especially in handling imbalanced datasets and minimizing false positives and negatives.

By incorporating the information gained from PPD data analysis and machine learning techniques for predicting suicidal tendencies, this chapter sets the structure for developing a robust methodology for this thesis, specific to the Bangladeshi context.

Chapter 4

Methodology

This chapter presents a detailed methodology to achieve the objectives of predicting that PPD causes suicidal tendencies among patients. First, the data collection process is discussed and the description of the postpartum survey dataset is represented, followed by an explanation and steps taken to ensure data quality through preprocessing. Additionally, the techniques used to address class imbalance using sampling are explored to ensure that the models do not favor the majority class. Then three different types of feature selection methods are discussed to identify the most relevant attributes, which are crucial for building efficient and interpretable prediction models. Then the application of machine learning algorithms—Logistic Regression (LR), Support Vector Machine (SVM), Decision Tree (DT), and Random Forest (RF)—is thoroughly explained, including the role of hyperparameter optimization using grid search, k fold cross validation and LASSO regularization in model training. Finally, the chapter outlines the evaluation metrics used to assess model performance, ensuring that the developed models are both accurate and reliable in predicting PPD.

4.1 Data Collection

To develop the prediction model, a postpartum dataset has been used. The dataset has been sourced from Kaggle, contributed by Md. Parvez Mosaraf, and consists of survey data from participants in Bangladesh from 1503 individuals, collected via a questionnaire administered through a Google Form in collaboration with a medical hospital. The questionnaire covered various aspects of PPD symptoms and was designed to gather insights into the experiences of individuals following childbirth [160]. Figure 4 represents the correlation heatmap of these attributes, while Table 6 shows the characteristics of these symptoms and risk factors among the dataset sample. The validation of the dataset relies solely on its availability on Kaggle, as no associated publications support its credibility. The data items included in the dataset are as follows:

- Age
- Feeling sad or Tearful
- Irritable towards baby & partner
- Trouble sleeping at night
- Problems concentrating or making a decision

- Overeating or loss of appetite
- Feeling anxious
- Feeling of guilt
- Problems of bonding with baby
- Suicide attempt

Where suicide attempt is the target variable with 459 positive and 1044 negative instances in the 1503 rows of collected data.

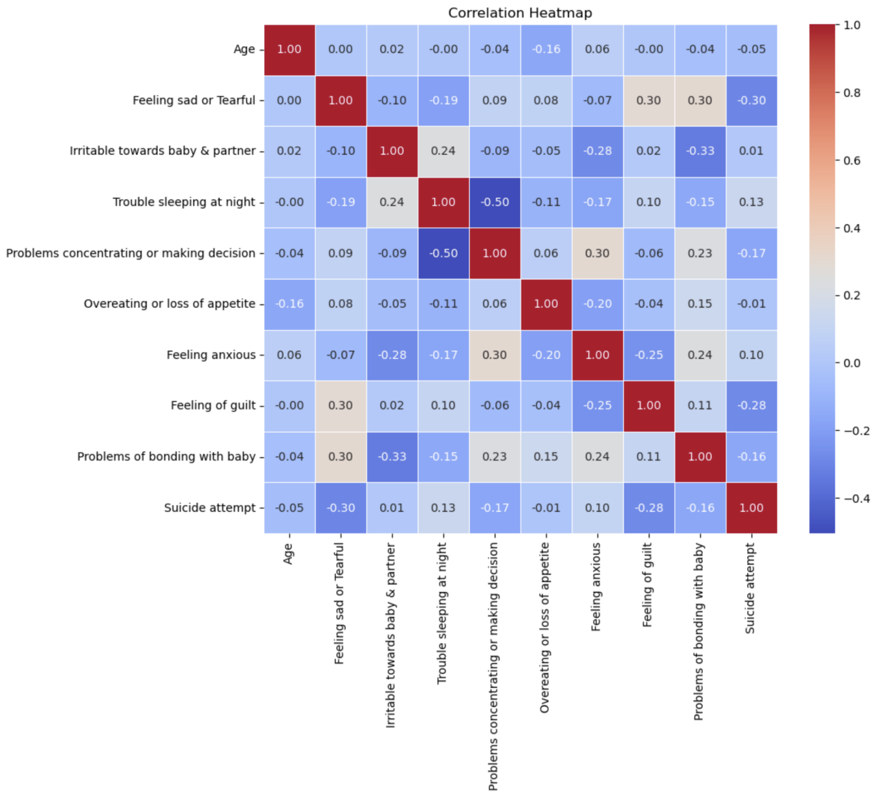

Figure 4: Correlation heatmap of dataset attributes

4.2 Data Preprocessing

After data collection, data preprocessing is a crucial step in the machine learning pipeline because it directly impacts the quality of the data used for training models. Properly preprocessed data leads to more accurate and efficient models.

Firstly, the dataset is inspected for any missing information, as some machine learning algorithms cannot handle missing values and those that do can give skewed results, reducing accuracy. Missing data can lead to potential bias in the classification model due to the information loss, hence imputation has been performed to preserve the statistical significance of the model. There are several methods of handling missing values such as using mean, median, or mode, using regression of classification models to predict missing values, or removing the rows with missing values in them. For this research, the missing values were replaced with the most frequent value within

Table 6: Characteristics of the population in the dataset in respective of symptoms and risk factors.

Variable	Category	Frequency (Yes)	Frequency (No)	Frequency Total	P-Value
Age	35-40	121 (42.5%)	164 (57.5%)	285 (24.4%)	0.003
	40-45	118 (41.4%)	167 (58.6%)	285 (24.4%)	
	30-35	99 (38.7%)	157 (61.3%)	256 (21.9%)	
	25-30	61 (46.9%)	69 (53.1%)	130 (11.1%)	
	45-50	60 (28.3%)	152 (71.7%)	212 (18.2%)	
Feeling sad or Tearful	No	254 (61.4%)	160 (38.6%)	414 (35.4%)	<0.001
	Yes	140 (32.3%)	294 (67.7%)	434 (37.2%)	
	Sometimes	65 (20.3%)	255 (79.7%)	320 (27.4%)	
Irritable towards baby & partner	Yes	260 (58.7%)	183 (41.3%)	443 (37.9%)	<0.001
	No	105 (24.6%)	321 (75.4%)	426 (36.5%)	
	Sometimes	94 (31.4%)	205 (68.6%)	299 (25.6%)	
Trouble sleeping at night	Two or more days a week	227 (52.3%)	207 (47.7%)	434 (37.2%)	<0.001
	Yes	146 (41.5%)	206 (58.5%)	352 (30.1%)	
	No	86 (22.5%)	296 (77.5%)	382 (32.7%)	
Problems concentrating or making a decision	Yes	207 (62%)	127 (38%)	334 (28.6%)	<0.001
	No	193 (45.1%)	235 (54.9%)	428 (36.6%)	
	Often	59 (14.5%)	347 (85.5%)	406 (34.8%)	
Overeating or loss of appetite	No	270 (40.5%)	396 (59.5%)	666 (57%)	0.181
	Yes	103 (38.3%)	166 (61.7%)	269 (23%)	
	Not at all	86 (36.9%)	147 (63.1%)	233 (19.9%)	
Feeling anxious	Yes	331 (38.9%)	519 (61.1%)	850 (72.8%)	<0.001
	No	128 (40.3%)	190 (59.7%)	318 (27.2%)	
Feeling of guilt	No	267 (47.1%)	300 (52.9%)	567 (48.5%)	<0.001
	Yes	124 (59.9%)	83 (40.1%)	207 (17.7%)	
	Maybe	68 (17.3%)	326 (82.7%)	394 (33.7%)	
Problems of bonding with baby	No	209 (46.9%)	237 (53.1%)	446 (38.2%)	<0.001
	Yes	141 (43.7%)	182 (56.3%)	323 (27.7%)	
	Sometimes	109 (27.3%)	290 (72.7%)	399 (34.2%)	

each respective column. This approach was deemed appropriate given the categorical nature of the attributes in the selected dataset. By replacing missing values with the most common value, we ensured that the integrity of the dataset was maintained while retaining as much information as possible.

Given that the majority of the data comprised categorical variables rather than numerical values, Label Encoding was employed to convert ordinal categorical data to numeric data. This process involved assigning numerical labels to categorical variables, enhancing the computer's ability to process and interpret the data accurately. This approach was used based on its compatibility with earlier analysis of similar models.

Lastly, to ensure that the attributes contribute equally during classification, Min-Max normalization was performed on each column of data. This technique involves scaling numerical values to a predetermined range, between 0 and 1. By doing so, the risk of any single feature influencing the analysis due to its magnitude is solved, ensuring that all features contribute meaningfully to the final results and reduce overfitting.

4.3 Data Balancing

To train machine learning prediction models, balanced data are very important because they can affect the model that we train by overfitting. An imbalanced dataset means that one class is observed in some instances more than the other class, which will skew the model performance since it can perform poorly on minority classes. To address this issue, two balancing techniques were used in this research, namely oversampling using SMOTE and undersampling using Random UnderSampler. A third set of prediction models was trained using the original unsampled dataset to evaluate the effect of sampling and nonsampling on predictive classification models.

SMOTE generates synthetic examples of the minority class by creating new samples rather than simply duplicating them. This approach not only balances the dataset but also introduces more variability within the minority class, leading to better generalization of the model. SMOTE is considered one of the most influential data preprocessing and sampling algorithms in machine learning and data mining [161]. In this research, the initial dataset contained 459 instances of the positive class and 1044 instances of the negative class for the target variable column. After applying SMOTE, the dataset was balanced, containing 1044 instances of the positive class and 1044 instances of the negative class, resulting in a final dataset with 2088 rows.

Random Undersampling is a technique used to balance the dataset by reducing the number of majority class instances by randomly removing samples from the majority class to match the number of samples in the minority class. While this approach can lead to a significant loss of information from the majority class, it simplifies the model training process and can prevent the model from becoming biased towards the majority class [162]. In this research, the Random Undersampler was applied to the training data to create a balanced dataset, reducing the positive class to 459 instances and matching 459 instances of negative class for the target variable, resulting in a final dataset of 918 rows.

As a baseline, the study also involved training models on the original, imbalanced dataset without applying any sampling technique, to provide a reference point to evaluate the effectiveness of the sampling methods used. Here the total number of rows is 1503 with 1044 instances of positive class and 459 instances of negative class for the target variable column.

4.4 Feature Selection

In machine learning, feature selection is an important step as it helps in improving model performance by selecting the most relevant features while reducing noise and redundancy, it also improves accuracy and reduces training time by removing misleading and unimportant data [163]. In this research, three different types of feature selection methods are used, namely the wrapper method (sequential/forward feature selection), filter method (Pearson's correlation coefficient), and expert-selected features.

Forward/Sequential Feature selection initializes with an empty set of features and gradually adds features up until the best combination of features is selected. This approach ensures that the final model is simplified, hence increasing interpretability and computational efficiency [164]. For this research, the Forward/Sequential feature selection function was used with the Random forest classifier. This repeated through each feature, including one at a time while evaluating the model's performance, and continued until no combination of features added resulted in a higher accuracy score. Features selected using the forward feature selection method are listed below:

- Age
- Feeling sad or Tearful
- Irritable towards baby & partner
- Trouble sleeping at night
- Problems concentrating or making a decision
- Feeling of guilt
- Problems of bonding with baby

Pearson's correlation coefficient was used to select features as a filter method. This method evaluates the linear correlation between each feature and the target variable, where features with a high correlation coefficient **(r)** are retained for model training. For all features, **r** were calculated against suicide attempts. The statistical significance was calculated using p-values, with n=1503 and α=0.01, ensuring the selected features has a confidence level of 99%. This method offers a statistical way to identify features that have a significant impact on suicidal ideation. Features selected using Pearson's correlation are listed in Table 7.

Table 7: List of features selected by Pearson's correlation coefficient

Attribute name	Pearson's Correlation Coefficient (r)	P-values
Feeling sad or Tearful	-0.3	7.00E-32
Trouble sleeping at night	0.13	6.35E-07
Problems concentrating or making a decision	-0.17	8.17E-11
Feeling of guilt	-0.28	1.55E-27
Problems of bonding with baby	-0.16	3.26E-10
Feeling anxious	0.10	1.88E-4

Finally, the 3rd set of features was selected by an expert in the field of maternal mental health. For this approach, a medical professional, provided her expertise to select symptoms and risk factors associated with PPD, ensuring that the resulting model captured the factors contributing to PPD-related suicidal tendencies with greater accuracy and clinical relevance. Features selected using expert advice are listed in Table 8.

4.5 Model training and evaluation

To develop the machine learning model, Python's libraries, namely pandas and sci-kit-learn (sklearn), were used. The pandas library allows efficient data manipulation and preprocessing, while sklearn provides a comprehensive, open-source set of functions for model development and evaluation.
To reduce overfitting, feature regularization was used, and the L1 regularization technique, commonly known as LASSO (Least Absolute Shrinkage and Selection Operator), was applied. Lasso regularization assigns weights to features and reduces overfitting by optimizing the model complexity [165], it regularizes important features by estimating model coefficients simultaneously [166]. Additionally, the most optimal hyperparameter combination for each supervised learning algorithm

Table 8: List of features selected by expert advice

Attribute name	Category
Age	Risk factor
Feeling sad or Tearful	Symptom
Irritable towards baby & partner	Symptom
Trouble sleeping at night	Symptoms, Risk factor
Problems concentrating or making a decision	Symptom
Overeating or loss of appetite	Symptom
Feeling anxious	Symptom
Feeling of guilt	Symptom
Problems of bonding with baby	Symptoms, Risk factor

has been sought through gridsearchCV. Grid search systematically explores a predefined parameter grid to find the combination that provides the best performance. Moreover, 5-fold cross-validation is performed to partition the dataset into five subsets, iteratively using four subsets for training and one for validation. The average performance across the five folds is calculated to provide a robust evaluation metric for model performance.

Four simple supervised learning algorithms: Logistic Regression (LR), Support Vector Machines (SVM), Decision Trees (DT), and Random Forest (RF) were selected to test and compare the performance of the PPD-caused suicidal tendency prediction. These algorithms have been chosen for their simplicity, and effectiveness in handling binary classification tasks as explored in the previous chapter.

Logistic Regression is a flexible tool to understand the connection between many descriptive features in a binary response [167]. It is widely used to classify inputs by modeling the probability that they belong to a class through the use of a logistic function. In this research, regularization was achieved using an L1 penalty known as Lasso regularization. Additionally, the 'liblinear' solver was used, which was selected for its efficacy with small datasets and compliance with L1 regularization, making it perfect for the dataset used. Finally, the maximum number of iterations for the algorithm was set to 1000. These hyperparameters collectively influence the model's complexity and performance, allowing for fine-tuning to achieve optimal results.

When it comes to binary classification, the support vector machine has been a good mechanism for the past couple of years with good generalization properties [168]. It operates by identifying the hyperplane that provides the widest margin of separation between the classes within the feature space. For this study, the linear SVM kernel was used. This is because the selected dataset was found to be linearly separable through the use of a scatter plot diagram. Moreover, the radial bias kernel requires the computation of a gram matrix which grows with the size of the dataset and due to resource limitation, the linear kernel was chosen. Regularization was also achieved with L1 regularization, and finally, the maximum number of iterations for the algorithm has been set to 1000 to let the algorithm have enough time to find a solution.

Decision trees are hierarchical structures used for binary classification, where each internal node represents a decision based on a feature, leading to the next nodes representing class labels. They are effective for binary classification due to their ability to handle nonlinear relationships, and feature interactions efficiently. Moreover, decision trees are interpretable, enabling insights into decision-making processes, and making them a suitable choice for binary classification tasks. Furthermore, their ability to handle non-linear relationships and interactions between features makes them a preferred choice in various domains, including healthcare [169]. In this research, optimal

hyperparameters for decision trees were selected by testing several combinations, with 3 options for maximum depth (tree depth: 3, 5, 7), 3 options for minimum samples required to split a node: 2, 5, 10, and 3 options for minimum samples required at a leaf node: 1, 2, 4. This results in $3 \times 3 \times 3 = 27$ combinations of hyperparameters to be tested. Finally, the selected hyperparameters were, maximum depth of 7, minimum samples required to split a node as 2 and minimum samples required at a leaf node as 1, giving the optimum performance.
Random Forest, is adept at binary classification tasks because of its ability to process complex data and reduce overfitting. It is a combination of multiple decision tree outputs where each tree is trained using a random subset of the dataset sample. The hyperparameters of Random Forest include the maximum depth of each tree, the minimum number of samples, and the number of trees in the forest required to split a node. These control the model's complexity, aiding in regularization and preventing overfitting [170], making it efficient for binary classification tasks where robustness and generalization are required. In this research, optimal hyperparameters for random forests were selected by testing several combinations, with 3 options for number of trees in the forest: 100, 200, 300, 4 options for tree depth: 10, 20, 30, or unlimited, 3 options for minimum samples to split a node: 2, 5, 10, and 3 options for minimum samples required in a leaf: 1, 2, 4. This creates $3 \times 4 \times 3 \times 3 = 108$ combinations of hyperparameters. Finally, the selected hyperparameters were, number of trees in the forest being 100, maximum depth of 20, minimum samples required to split a node as 2 and minimum samples required at a leaf node as 1, giving the optimum performance.
The evaluation metrics used to measure the performance of the machine learning models are accuracy, recall, precision, specificity, f1 score, and area under the receiver operating curve. These metrics are the most frequently used with binary classification models [171], and are most commonly used to evaluate machine learning models, as observed in section 3.3. Additionally, Cohen's Kappa has been used to measure the agreement between the actual and predicted labels while accounting for chance agreement, as it takes into account the possibility of the agreement occurring by chance. Cohen's Kappa is particularly useful when evaluating the performance of classifiers in situations where there might be an imbalance in the distribution of classes. A visual representation of the complete methodology is provided in Figure 5.

4.6 Ethical Considerations and Data Privacy

Research using personal data requires ethical consideration to protect sensitive information. The process of data collection, data sharing, and using the data in studies, need to follow governance frameworks covering anonymization and participant consent of data usage, ensuring transparency, and data privacy [172, 173]. In this research, sensitive information related to PPD has been sourced from an online repository. As this is secondary data, there is limited direct verification of ethical practices used for data collection. No personally identifiable information was included in the dataset to anonymize the data. The data was collected from voluntary participants during a survey using google forms. Furthermore, the research outcomes are confined to academic purposes, with no immediate application in clinical or real-world settings. Any future implementation of these models will necessitate comprehensive ethical governance and compliance with data protection laws to safeguard participants' rights. This research is conducted as a subcomponent of the larger research project titled "SuSastho.AI: An AI-enabled solution to improve the health and wellbeing of the population in Bangladesh" which has received ethical approval from the Bangladesh Medical Research Council (Approval Reference: BMRC/NREC/2022-2025/368, Registration Number: 515122jQ212Q2k).

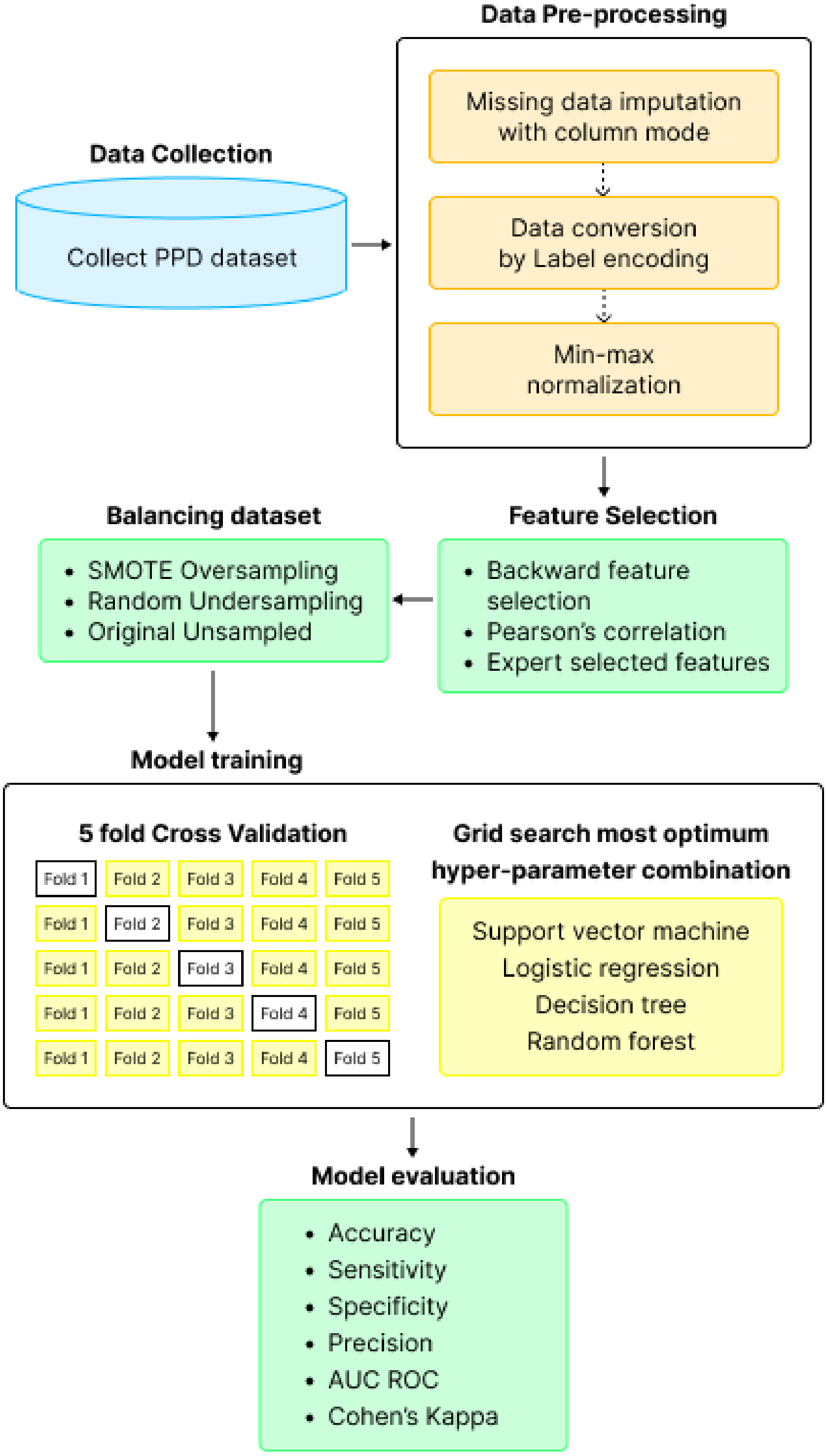

Figure 5: Complete methodology of the prediction model

4.7 Conclusion

In this chapter, the methodology used to develop and evaluate the prediction models for PPD and associated suicidal tendencies is outlined. The process began with the collection of a postpartum survey dataset from Kaggle, followed by data cleaning and preprocessing to handle missing values using mode, categorical to numeric data conversion using label encoding, and min-max normalization. These steps were essential to prepare the dataset for effective model training. Data balancing techniques, namely SMOTE and Random UnderSampling, were used to address class imbalance, to ensure that the machine learning models could perform well on both the minority and majority classes, thereby improving the generalization of the models. Feature selection was done using three different methods, forward feature feature selection, Pearson's correlation, and expert advice to identify the most relevant features for predicting PPD and associated suicidal tendencies. The core of our methodology involved training and evaluating four machine learning algorithms: Logistic Regression (LR), Support Vector Machine (SVM), Decision Tree (DT), and Random Forest (RF). Each algorithm was chosen for its unique strengths in handling binary classification tasks. The hyperparameters of these models were optimized using grid search combined with 5-fold cross-validation, ensuring that the models were fine-tuned to achieve the best performance. To evaluate these models, accuracy, sensitivity, specificity, precision, F1 score, area under the ROC curve (AUC), and Cohen's Kappa were used. In the following chapter, the results of these predictive models will be discussed in detail.

Chapter 5

Result Analysis

This chapter presents the evaluation of various machine learning models applied to the postpartum dataset, focusing on the effects of different feature selection methods and sampling techniques on model performance. The aim is to identify the most effective combinations of feature selection, sampling, and algorithms to optimize the predictive accuracy and reliability of the postpartum suicidality prediction model, providing a detailed understanding of how each model performs under varying conditions. Initially, the four machine learning algorithms were trained using the expert-selected features and the original unbalanced dataset, results are shown in Figure 6. Random forest demonstrated the best performance across all metrics, with 99.73% accuracy, 99.14% precision, and 100% recall, indicating strong effectiveness in identifying both positive and negative cases. The Decision tree also performed well, showing balanced results with 92.68% accuracy, 90.49% precision, and 84.97% recall. On the other hand, both Logistic regression and SVM had lower accuracy, with LR achieving 76.38% accuracy and SVM 75.45%. These results highlight Random forest as the most reliable model, followed by the Decision tree, with Logistic regression and SVM showing weaker performance.

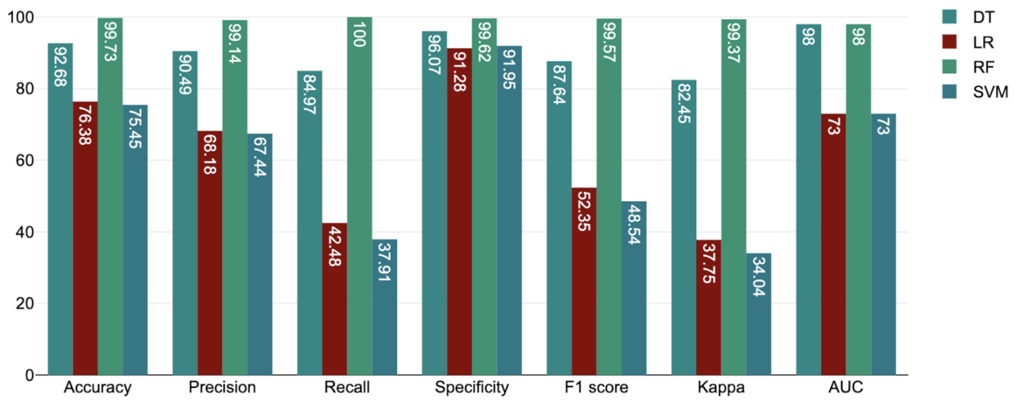

Figure 6: Performance of the machine learning algorithms with expert selected features and unbalanced dataset.

Additionally, to assess the influence of data sampling on model performance, two sampling techniques were applied, SMOTE oversampling and random undersampling. The performance of the models using these sampling methods was compared to the results obtained without sampling. For all algorithms except the Decision tree, the unsampled data showed better model performance.

Specifically, oversampling using SMOTE slightly improved the performance of algorithm Decision trees, but this improvement was not observed for algorithms Random forest, Logistic regression, and Support vector machine. Figure 7 shows a bar chart comparing the performance of different sampling techniques.

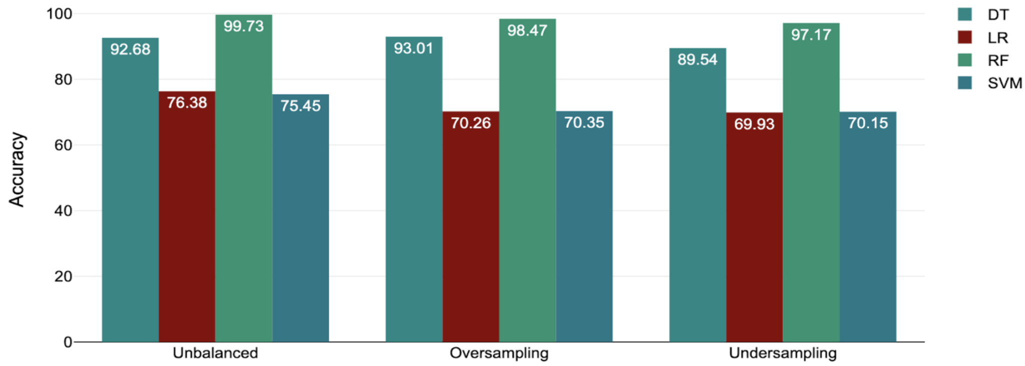

Figure 7: Average accuracy of the machine learning models trained using expert selected features, represented in the form of a bar chart and box plot, different sampling methods.

Moreover, to evaluate the impact of alternative feature selection methods on performance, the model was trained using Pearson's Correlation Coefficient (PCC) and Sequential Forward Selection (SFS). Across all algorithms, the expert-selected features consistently provided the best results. However, there were some differences in performance when using PCC and SFS. SVM and Logistic regression performed slightly better with PCC-selected features compared to SFS, whereas algorithms Random forest and Decision tree demonstrated superior performance with SFS-selected features over PCC. This result is represented in Figure 8 using a bar chart and box plot to compare the average accuracy.

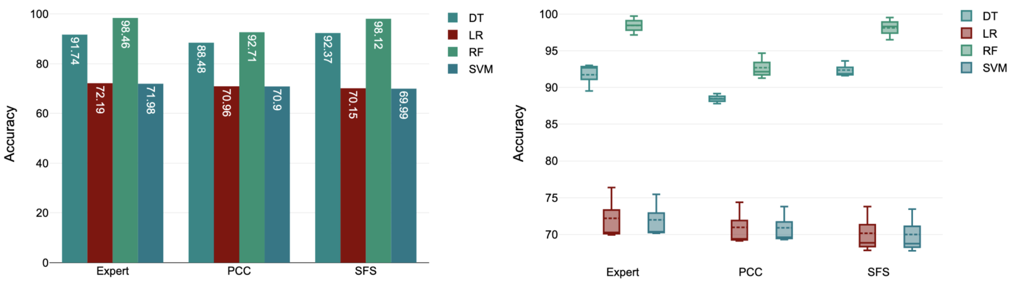

Figure 8: Average accuracy of the machine learning models represented in the form of a bar chart and box plot, for different feature selection methods.

While expert-selected features worked better with unsampled data, in cases of using PCC and SFS for feature selection, using oversampled data slightly improved performance for Random forest and Decision tree, but no such improvement is seen with SVM and Logistic regression. Undersampled data consistently gave the lowest accuracy with all algorithms and feature selection methods, as shown in Figure 9.

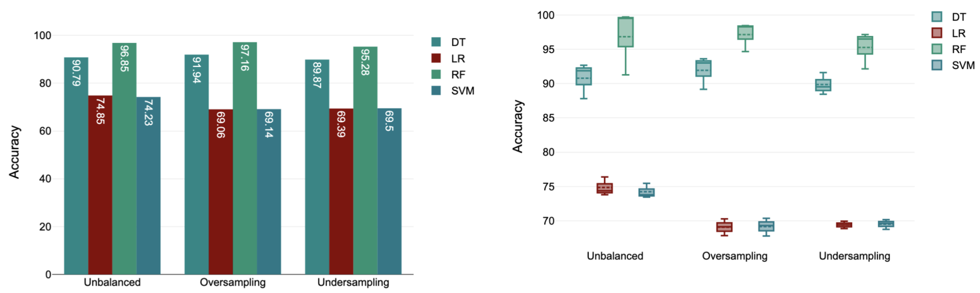

Figure 9: Average accuracy of the machine learning models represented in the form of a bar chart and box plot, for different sampling methods.

Overall, the best model was identified as a Random forest, trained using the expert-selected features on the unsampled dataset, and consistently outperformed all others, with an accuracy of 99.73% and AUC of 0.98. An area under the receiver operating curve graph of the best-performing model is shown in Figure 10. Finally, table 9 summarizes the results of all the developed machine learning models in different setups and their evaluations.

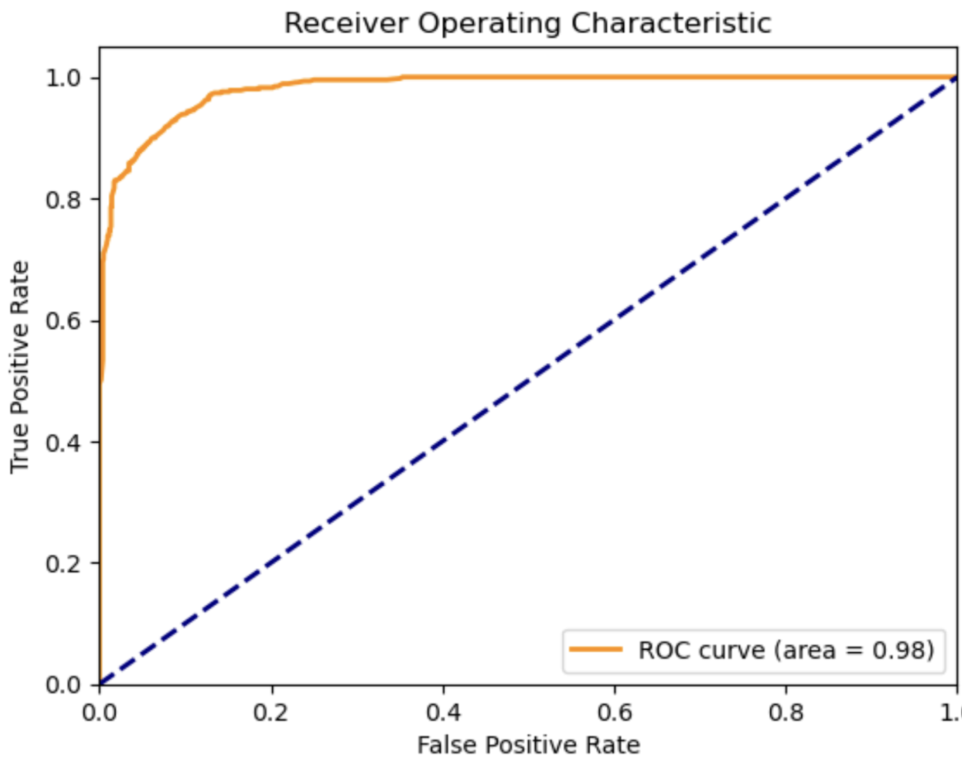

Figure 10: AUC under ROC curve for the best-performing model where Random forest was used with expert-selected features on the original unbalanced dataset.

Table 9: Evaluation of the machine learning models using accuracy (ACC), precision (PR), recall (RE), specificity (SP), F1 score, Cohen's Kappa (CK), and AUC.

FS method	Sampling	Model	ACC	PR	RE	SP	F1	CK	AUC
Expert selected features	No	LR	76.38	68.18	42.48	91.28	52.35	37.75	0.73
		RF	99.73	99.14	100.00	99.62	99.57	99.37	0.98
		SVM	75.45	67.44	37.91	91.95	48.54	34.04	0.73
		DT	92.68	90.49	84.97	96.07	87.64	82.45	0.98
	Over	LR	70.26	70.05	70.79	69.73	70.41	40.52	0.75
		RF	98.47	99.41	97.51	99.43	98.45	96.93	0.98
		SVM	70.35	70.10	70.98	69.73	70.54	40.71	0.75
		DT	93.01	92.20	93.97	92.05	93.07	86.02	0.97
	Under	LR	69.93	70.38	68.85	71.02	69.60	39.87	0.74
		RF	97.17	94.64	100.00	94.34	97.25	94.34	0.97
		SVM	70.15	70.51	69.28	71.02	69.89	40.31	0.74
		DT	89.54	87.27	92.59	86.49	89.85	79.08	0.94
Pearson's Correlation selected features	No	LR	74.38	62.67	39.87	89.56	48.74	32.77	0.74
		RF	91.28	86.12	85.19	93.97	85.65	79.39	0.87
		SVM	73.79	62.08	36.38	90.23	45.88	30.10	0.73
		DT	87.89	81.99	77.34	92.53	79.60	70.99	0.95
	Over	LR	68.06	68.39	67.15	68.97	67.76	36.11	0.74
		RF	94.25	95.56	92.82	95.69	94.17	88.51	0.9
		SVM	68.20	68.52	67.34	69.06	67.92	36.39	0.74
		DT	89.66	89.81	89.46	89.85	89.64	79.31	0.96
	Under	LR	67.32	67.95	65.58	69.06	66.74	34.64	0.73
		RF	91.94	90.19	94.12	89.76	92.11	83.88	0.85
		SVM	66.88	68.24	63.18	70.59	65.61	33.77	0.73
		DT	86.49	86.98	85.84	87.15	86.40	72.98	0.94
Sequential Forward feature selection	No	LR	73.79	61.02	39.22	88.98	47.75	31.34	0.73
		RF	99.53	98.92	99.56	99.52	99.24	98.90	0.93
		SVM	73.45	61.11	35.95	89.94	45.27	29.27	0.73
		DT	91.88	86.55	86.93	94.06	86.74	80.89	0.98
	Over	LR	67.82	67.99	67.34	68.30	67.66	35.63	0.74
		RF	98.32	99.41	97.22	99.43	98.31	96.65	0.94
		SVM	67.77	67.79	67.72	67.82	67.75	35.54	0.74
		DT	93.63	92.29	95.21	92.05	93.73	87.26	0.97
	Under	LR	68.85	68.93	68.63	69.06	68.78	37.69	0.74
		RF	96.51	93.48	100.00	93.03	96.63	93.03	0.91
		SVM	68.74	68.86	68.41	69.06	68.63	37.47	0.73
		DT	91.61	88.20	96.08	87.15	91.97	83.22	0.96

Chapter 6

Discussion and conclusion

Considering the extent of this topic which has limited research, this chapter has been kept separate to provide an in-depth understanding of PPD caused suicidal ideation prediction. Accordingly, the research findings such as the performance of the prediction models, why one was better than the other in predicting PPD and suicidal tendencies, the variability, and analyzing key performance metrics like precision, recall, F1 score, Kappa, and AUC will be discussed in this chapter. Furthermore, this study applied oversampling and undersampling techniques which led to specific changes in the outcome which will be discussed in this chapter. Moving forward, from the many techniques of feature selection, the purpose of selecting the specific techniques chosen for this study and their analysis will be presented here. Many researchers have adopted different approaches to similar solutions, a detailed comparison of their research emphasizing how they implemented their model in search of a specific goal would bring more clarity to the topic and subtopics of this research.
Through the course of this study, many limitations and challenges have been identified, increasing the time taken to complete this research or deviating from the initial approach. To overcome such limitations and challenges, specific methods and approaches were adopted which will be discussed in this chapter.

6.1 Model performance

The performance of various machine learning algorithms in predicting PPD-related suicidal ideation was examined, focusing on how well each model balanced accuracy, recall, and specificity. Each machine learning algorithm's strengths and limitations are explored in terms of key performance metrics such as accuracy, recall, specificity, F1 score, and kappa values.
Among all the machine learning algorithms, Random Forest consistently outperformed other models across different feature selection (FS) methods and sampling techniques, achieving the highest accuracy and F1 score. Random Forest was able to maintain high values for both recall and specificity, indicating less of a trade-off and a more balanced performance, suggesting its effectiveness in identifying suicidality cases and being able to distinguish between positive and negative cases well. Kappa scores are consistently high for Random Forest indicating stronger agreement model reliability.
Decision Trees also performed well, especially in terms of recall and specificity, showing strong predictive power next to Random Forests with the highest AUC for models trained using SFS and PCC selected features regardless of the sampling technique used. Kappa values for Decision trees are also high, with the highest at 87.26%, indicating good model reliability. This indicates

tree-based models like Decision tree and Random Forest are able to work well with categorical features, capturing complex patterns and relationships between symptom and behavioral features even when converted to numeric data using one hot encoding. With slightly lower recall values, Decision Tree potentially lacks the robustness ensemble methods like Random forests can achieve, which leads to missing some positive suicidal ideation cases.
Support Vector Machines (SVM) and Logistic Regression generally show lower performance compared to the tree-based models, with logistic regression being slightly better than SVM in most cases. Both show a clear trade-off between recall and specificity, where high recall often comes at the cost of lower precision, and vice versa. This indicates the need for threshold tuning if such a model is to be applied in the medical system. Logistic Regression and SVM also show lower Kappa scores, often below 50%, reflecting less reliable performance compared to Random Forest and Decision Tree models. Moreover, Logistic Regression and SVM show lower AUC values, generally around 0.73-0.76, indicating moderate predictive power. Consistent low performance of these models may be due to their linear nature, making them unable to capture complex, nonlinear interactions between predictors. It also shows these models face difficulty in handling categorical data. The numeric transformation may not align well with the linear decision boundaries that these models use, leading to poorer classification outcomes.

6.2 Sampling techniques

The influence of different sampling methods (Unbalanced, Oversampling, and Undersampling) on machine learning algorithms were examined, with the goal of predicting Postpartum Depression (PPD) leading to suicidal ideation. The findings show that the impact of these sampling methods varied across models, highlighting their sensitivity to class imbalance.
When trained on an unbalanced dataset, where the minority class were individuals with suicidal ideation, the models exhibited a range of behaviors depending on their ability to manage imbalanced data. Random Forest demonstrated exceptional performance on the unbalanced data, achieving high accuracy and recall. This suggests that Random forest, due to its ensemble nature and robustness to overfitting, is highly effective at capturing patterns in both the majority and minority classes, even without balancing the data. Decision Tree also performed relatively well but showed a drop in recall, indicating that some suicidal ideation cases were misclassified. This suggests that Decision Trees struggle more than RF when dealing with imbalanced data, but still performs adequately. Logistic Regression and SVM, on the other hand, struggled significantly on the unbalanced dataset, with very low readings for recall, F1 score and Cohen's kappa. Both models exhibited poor performance in identifying the minority class, which is a common challenge for algorithms that assume balanced class distributions. Their high specificity and low recall indicate a bias toward the majority class, making them less effective in predicting the critical minority cases of suicidal ideation.
Oversampling, which artificially increases the number of minority class instances, generally led to improvements in model performance by providing a more balanced representation of classes during training. Random Forest performance remained high with oversampling though slightly lower than on the unbalanced dataset. This small decrease suggests that oversampling slightly increased model complexity but did not significantly impact the algorithm's ability to generalize across classes. Decision Tree improved significantly with oversampling, with increase in accuracy, recall and F1 score. The structure of Decision Trees is likely a contributing factor here, since Decision Trees tend to split the data based on maximizing information gain at each node, the imbalance

in data can result in skewed splits. Oversampling through SMOTE provides more minority class instances, improving the chance that the Decision Tree can make more balanced decisions. Logistic Regression and SVM showed the most improvement, even though accuracy was low, there was a significant increase in recall, F1 score and Cohen's kappa with oversampling compared to use of unbalanced dataset. These results indicate that oversampling helped these linear models overcome its limitations on imbalanced data, allowing for more balanced predictions between classes.
Undersampling, which reduces the majority class to match the size of the minority class, resulted in more balanced predictions but often at the expense of overall accuracy. Using undersampled data consistently resulted in the lowest accuracy for all the models. While undersampling balances the class distribution, it also discards potentially valuable data, which can lead to a loss of important information, hence reducing model performance.

6.3 Feature selection techniques

The study evaluated the impact of three feature selection (FS) methods namely, Expert, Pearson's Correlation Coefficient (PCC), and Sequential Feature Selection (SFS) on the performance of four machine learning algorithms to determine how well these combinations predict postpartum depression related suicidal ideation and to understand the interaction between FS methods and algorithms.
Expert FS prioritizes features that are theoretically linked to suicidal ideation based on clinical knowledge. Random forest and Decision trees showed high performance with these features possibly due to their ability to handle complex patterns effectively. These models are able to handle non-linear and hierarchical relationships, so they naturally benefit from this FS method. High precision and recall of these models with expert-selected features suggest good balance, which is critical in identifying suicidal ideation without too many false positives or negatives. In clinical or sensitive domains like PPD-related suicidal ideation, using expert-selected features may provide more interpretable results. On the other hand, Logistic Regression struggled with classification when paired with Expert FS, leading to lower accuracy and poor precision-recall trade-offs. Logistic Regression is a linear model, and the features selected by experts might not all exhibit a linear relationship with the target variable. This could explain the moderate performance, particularly in terms of precision and recall. Similarly, SVM consistently performed poorly with Expert FS, showing the lowest precision and recall metrics. SVM performs best when the data has clear boundaries between classes. The expert-selected features may not provide the necessary structure for SVM to establish such boundaries effectively.
PCC FS selects features based on linear correlations with the target variable. Random Forest and Decision Tree again showed significantly lower performance with PCC FS compare to other methods. The reason behind this might be because these tree-based algorithms are nonlinear models that can handle complex interactions between features, which statistically selected features cannot capture, as these methods focus on high correlation over complex interaction between predictors. In clinical datasets with complex relationships (like PPD-related suicidal ideation), FS methods like PCC may oversimplify the feature interactions, leading to lower performance. Logistic Regression and SVM performed reasonably well with PCC FS, with moderate precision, recall, and F1 scores across all sampling techniques, due to the features' strong linear correlation with the target variable.
SFS works by iteratively selecting features that maximize performance. Random Forest and Decision Tree showed excellent results with SFS, with near-perfect scores in oversampling and high

precision-recall balance. SFS can capture non-linear interactions between features that may have been missed in other FS methods, this particularly benefits tree-based model due to their ability to capture non linear, complex interactions between predictors. Logistic regression and SVM showed low performance with SFS, especially in precision and recall. While SFS introduced extra features such as “age” and “irritability towards the baby” because they are not linearly correlated but are significant predictors of the target variable when combined with the interactions of other features, linear machine learning models could not process these complex interaction and treated these features as noise, hence Logistic regression and SVM showed lowest performance with features selected by SFS These algorithms are not suited to non-linear relationships, and fail to make use of the complex interactions between predictors when making classifications.
The choice of FS method directly impacts the model’s ability to identify PPD-induced suicidal ideation. For clinical applications to predicting suicidal ideation in PPD patients, tree-based models like RF, paired with feature selection methods that emphasize both expert knowledge and performance optimization, like Expert FS or SFS, provide the best results. These models offer both high precision and recall, ensuring fewer missed cases and false alarms, which is critical in a clinical setting. PCC, though useful for linear models, may not suffice for complex clinical datasets.

6.4 Comparison with existing research

Since this PPD-caused suicidal prediction model is a novel approach and no other similar model exists, the results will be first compared with existing PPD-related prediction models and suicidality prediction models in mental health research. The evaluation metric was used to compare the results in the area under the receiver operating curve (AUC) since all the related studies shared their results using this metric. From the two algorithms used by Betts et al. to predict PPD-related psychiatric admissions, Gradient Boosted Trees showed the best performance with an AUC of 0.80 [126]. Zhang et al. implemented their model on questionnaire and scale data to conclude SVM as the best predictor of PPD with an AUC of 0.81 [124]. Similarly, a study conducted by Andersson et al. using Extreme Randomized Forests (ERT) performed best in predicting PPD with an AUC of 0.81. [122]. Wakefield et al. reached even better performance using Random forest to predict women requiring treatment for PPD with an AUC of 0.91 [125] and Zhang et al. reached an AUC of 0.93 to predict the risk of pregnant women developing PPD in the future, using Logistic regression [123]. Compared to these studies, the PPD-caused suicidal tendency prediction model developed in this research provided comparatively better AUC as a performance measure ranging from 0.86 to 0.98 using Random forest.
For studies focusing on predicting suicide attempts using questionnaire data, Su et al. utilized a Random Forest (RF) model to predict suicide attempts in adolescents, achieving an AUC of 0.72 [139], Zuromski et al. predicted suicide attempts among soldiers after deployment using ensemble methods achieving an AUC of 0.77 [147]. Additionally, for studies focusing on suicidal ideation Donnely et al. and Rus Prelog et al. were able to achieve an AUC of 0.83 using Logistic regression [142, 155]. Lastly, Wang et al. used a LightGBM (LGB) model and achieved an AUC of 90.1 to predict suicidal behavior among the middle-aged population using Light gradient boosting [141]. Compared to these studies, the PPD caused suicidality prediction model developed in this research showed a better AUC ranging from 0.86 to 0.98 using Random forest.

6.5 Limitations and Challenges

Predicting the risk of suicide due to PPD is an important area to explore, given its potential impact on maternal mental health and infant wellbeing. However, research in this domain faces several limitations and challenges that can affect the validity and reliability of the findings.

6.5.1 Limitations

- **Data acquisition limits:** No previous studies on PPD-related suicide risk prediction; and data curation was not possible due to time and funding constraints
- **Use of online datasets:** Absence of comprehensive data from varied geographic locations and socioeconomic backgrounds in Bangladesh.
- **Missing critical factors:** Important characteristics were missing such as sociodemographic data, family history, economic status, past medical history, and past mental health history
- **Bias in dataset:** Dataset may be biased because it is based on online survey, which cannot be validated with psychiatric professionals and may be inaccurate due to underreporting, bias, misinterpretation and cultural sensitivity.
- **Small dataset size:** Limited number of samples reduced statistical power and generalizability of the machine learning model
- **Impact on advanced modeling:** The absence of diverse data limits the application of deep learning models like CNN and LSTM, which perform better with unstructured data.

6.5.2 Challenges

- **Scarcity of datasets:** Limited availability of online datasets hinders research on PPD and suicide risk prediction.
- **Access to diverse data sources:** Medical records, clinical data, and social media data are difficult to obtain due to privacy regulations and accessibility constraints.
- **Data collection hurdles:** Gathering comprehensive datasets is challenging due to privacy issues, lack of institutional support, and difficulties in standardizing data across sources.

6.6 Conclusion

This research highlights the need for innovative approaches to address PPD and associated suicidal tendencies among new mothers in Bangladesh, where traditional diagnostic methods face significant challenges. By utilizing the rising potential of artificial intelligence (AI) and machine learning (ML) algorithms, this study has successfully developed a predictive model to aid in the early detection of PPD-related suicidal ideation, reducing diagnostic delay, and provide improved acces to patient mental healtcare with resource recommendation, potentially transforming patient assessment and intervention strategies.

A postpartum survey dataset collected from Bangladeshi participants, consisting of symptom and behavioral data, which was preprocessed to ensure quality. Techniques such as missing value imputation, data conversion, and normalization were performed on the dataset to ensure proper training of machine learning models. Different feature selection methods and sampling techniques were also applied to find the most optimal combination of features and a balanced dataset that created a robust model. Multiple machine learning algorithms, including Logistic Regression (LR), Support Vector Machine (SVM), Decision Tree (DT), and Random Forest (RF), were used to develop the PPD-related suicidal ideation prediction model. The best hyperparameter combinations were determined using a grid search technique with K-fold cross-validation. Finally, the developed models were evaluated using key metrics such as accuracy, sensitivity, specificity, precision, F1 score, Cohen's kappa and the area under the ROC curve (AUC).

Our findings show that Random Forest consistently outperformed other models across different feature selection methods and sampling techniques, achieving the highest accuracy, precision, recall, specificity, F1 score, and Kappa values. Decision Tree models also demonstrated strong performance, with high AUC values. On the other hand, logistic regression and SVM showed lower performance, with highly varying results depending on the sampling techniques and feature selection methods used with these algorithms. This can be explained by the tree-base nature of Decision trees and ensemble technique of Random Forest, giving these an advantage in making more accurate predictions.

When trained on an unbalanced dataset, Random Forest performed exceptionally well, achieving high accuracy and recall, demonstrating its ability to handle imbalanced data effectively. Decision Tree also performed adequately but had lower recall, misclassifying some cases of suicidal ideation. Logistic Regression and SVM struggled with low recall and F1 scores, reflecting difficulty in identifying the minority class. Oversampling improved performance across all models, particularly for Decision Tree as they are prone to overfitting, and linear models like Logistic Regression and SVM as they are more sensitive to class imbalance, though undersampling led to balanced predictions at the cost of overall accuracy due to the loss of valuable data.

The analysis also revealed that Decision Tree and Random Forest models benefitted from SFS and expert-selected features due to their ability to manage complex interactions between features. In comparison, expert feature selection and Pearson's correlation yielded the best results for logistic regression and SVM due to high correlation of predictors with target variables, while SFS selected features with complexity that these linear models could not effectively handle.

When comparing the results with existing research on PPD and suicidal ideation prediction models, this study outperformed existing models in both PPD and suicide prediction domains with Random Forest and Decision Tree models achieving AUCs ranging from 0.86 to 0.98.

Despite the results, this study faced several limitations. The use of online datasets without the use of validated scales or clinical diagnoses for PPD and suicidal ideation, the limited sample size and missing attributes such as symptoms, risk factors and demographic information as predictors, lack of different data types such as clinical data or social media data, all limited the scope of this study and introduced challenges. Additionally, the preprocessing methods and the choice of machine learning algorithms limited by computational power may have impacted the study's outcomes. Finally the model's true efficacy, reliability, and adaptability in practical scenarios remain uncertain for not being able to validate the results in a clinical setting with real patient data.

The outcomes of this research hold significant implications for public health, particularly in maternal mental health in Bangladesh. By allowing early identification of individuals at risk of suicide due to PPD, this study contributes to the prevention of severe mental health outcomes and suicide attempts among new mothers. The publication of research findings through two journal papers, one as a systematic review and another detailing the developed model, contributes to advancing maternal mental health (MMH) and sexual and reproductive health and rights (SRHR) research in Bangladesh. The practical applications of this research include Healthcare practitioners using the developed predictive models to assess postpartum women for the risk of PPD, enabling early intervention and support. Additionally, the early detection of suicidal tendencies among women suffering from PPD gives time to use targeted interventions, such as psychiatrist allocation and tailored treatment plans. By integrating AI-driven patient screening tools into healthcare systems, clinicians can provide timely support and prevent PPD-related suicides, hence improving diagnosis outcomes and treatment in Bangladesh's mental health sector.

Moving forward, future research efforts should focus on expanding the scope and impact of predictive models for PPD in Bangladesh. This includes the creation of an extensive PPD dataset incorporating clinical data during pregnancy, real-time sensor data, demography-specific information of patients, patient social media data, etc, to diversify data collection and allow the use of advanced machine learning algorithms. Additionally, further investigation on effects of other feature selection and sampling methods, including hybrid, adaptive, and other alternative techniques should be explored in the context of suicidality prediction. Moreover, the effectiveness of deep learning approaches, such as neural networks or attention-based models for suicidal ideation prediction, should be investigated. Real-world testing in clinical settings is also essential to validate the models' effectiveness and adaptability under practical conditions, addressing complexities like incomplete or noisy data. The development of a Computer-Aided Diagnostic (CAD) system will further facilitate healthcare providers in patient screening and early detection of PPD, while development of self diagnosing assessments through mobile application can help patients who are unwilling to seek help due to cultural stigma. Nationwide interventions in hospital gynecology departments to implement AI-based patient screening tools can also help with the early detection and prevention of PPD-related suicides. By addressing these future research directions, we can advance the field of maternal mental health and improve outcomes for postpartum women in Bangladesh and beyond.

References

[1] Lavanya Pillay, Jenni Smit, Mags Beksinska, and Chelsea Morroni. Sexual and reproductive health. *Oxford Textbook of Global Health of Women, Newborns, Children, and Adolescents*, 2018.

[2] Ebenezer Oluwole Akande. Sexual and reproductive health and rights: An overview. *Contemporary Obstetrics and Gynecology for Developing Countries*, pages 3–6, 2021.

[3] Uriel Halbreich and Sandhya Karkun. Cross-cultural and social diversity of prevalence of postpartum depression and depressive symptoms. *Journal of affective disorders*, 91(2-3):97–111, 2006.

[4] Catherine Habel, Nancy Feeley, Barbara Hayton, Linda Bell, and Phyllis Zelkowitz. Causes of women s postpartum depression symptoms: Men s and women s perceptions. *Midwifery*, 31(7):728–734, 2015.

[5] Jane Fisher, Meena Cabral de Mello, Vikram Patel, Atif Rahman, Thach Tran, Sara Holton, and Wendy Holmes. Prevalence and determinants of common perinatal mental disorders in women in low-and lower-middle-income countries: a systematic review. *Bulletin of the World Health Organization*, 90:139–149, 2012.

[6] Norma I Gavin, Bradley N Gaynes, Kathleen N Lohr, Samantha Meltzer-Brody, Gerald Gartlehner, and Tammeka Swinson. Perinatal depression: a systematic review of prevalence and incidence. *Obstetrics & Gynecology*, 106(5 Part 1):1071–1083, 2005.

[7] HE Nasreen, M Edhborg, M Petzold, Y Forsell, and ZN Kabir. Incidence and risk factor of postpartum depressive symptoms in women: a population based prospective cohort study in a rural district in bangladesh. *J Depress Anxiety*, 4(1000180):2167–1044, 2015.

[8] Kazi Nazira Sharmin, Nazmul Sarwar, Shirin Jahan Mumu, DR Abu Taleb, and Meerjady Sabrina Flora. Postnatal depression and infant growth in an urban area of bangladesh. *Midwifery*, 74:57–67, 2019.

[9] Rashidul Azad, Rukhshan Fahmi, Sadichhya Shrestha, Hemraj Joshi, Mehedi Hasan, Abdullah Nurus Salam Khan, Mohiuddin Ahsanul Kabir Chowdhury, Shams El Arifeen, and Sk Masum Billah. Prevalence and risk factors of postpartum depression within one year after birth in urban slums of dhaka, bangladesh. *PloS one*, 14(5):e0215735, 2019.

[10] Fowzia Tasnim, Mymuna Rahman, Md Monimul Islam, Mahmudul Hasan, Md Golam Mostofa, and Md Mosfequr Rahman. Exposure to domestic violence and the risk of developing depression within 6 months postpartum in bangladesh. *Social psychiatry and psychiatric epidemiology*, 56:1189–1200, 2021.

[11] Md Jahirul Islam, Lisa Broidy, Kathleen Baird, and Paul Mazerolle. Intimate partner violence around the time of pregnancy and postpartum depression: The experience of women of bangladesh. *PloS one*, 12(5):e0176211, 2017.

[12] Justine Slomian, Germain Honvo, Patrick Emonts, Jean-Yves Reginster, and Olivier Bruyère. Consequences of maternal postpartum depression: A systematic review of maternal and infant outcomes. *Women's health*, 15:1745506519844044, 2019.

[13] Ruth Paris, Rendelle E Bolton, and M Katherine Weinberg. Postpartum depression, suicidality, and mother-infant interactions. *Archives of Women's Mental Health*, 12:309–321, 2009.

[14] Yi-Liang Lee, Yun Tien, Yin-Shiuan Bai, Chi-Kang Lin, Chang-Sheng Yin, Chi-Hsiang Chung, Chien-An Sun, Shi-Hao Huang, Yao-Ching Huang, Wu-Chien Chien, et al. Association of postpartum depression with maternal suicide: a nationwide population-based study. *International journal of environmental research and public health*, 19(9):5118, 2022.

[15] Carlie Boisvert, Robert Talarico, Jasmine Gandhi, Mark Kaluzienski, Alysha LJ Dingwall-Harvey, Ruth Rennicks White, Kari Sampsel, Shi Wu Wen, Mark Walker, Katherine A Muldoon, et al. Screening for postpartum depression and risk of suicidality with obstetrical patients: a cross-sectional survey. *BMC pregnancy and childbirth*, 23(1):635, 2023.

[16] Anaise Williams, Malabika Sarker, and Syeda Tahmina Ferdous. Cultural attitudes toward postpartum depression in dhaka, bangladesh. *Medical anthropology*, 37(3):194–205, 2018.

[17] Anwar Islam and Tuhin Biswas. Mental health and the health system in bangladesh: situation analysis of a neglected domain. *Am J Psychiatry Neurosci*, 3(4):57–62, 2015.

[18] Sandeep Reddy. Use of artificial intelligence in healthcare delivery. *eHealth-making health care smarter*, pages 81–97, 2018.

[19] Fei Jiang, Yong Jiang, Hui Zhi, Yi Dong, Hao Li, Sufeng Ma, Yilong Wang, Qiang Dong, Haipeng Shen, and Yongjun Wang. Artificial intelligence in healthcare: past, present and future. *Stroke and vascular neurology*, 2(4), 2017.

[20] Kiran Saqib, Amber Fozia Khan, and Zahid Ahmad Butt. Machine learning methods for predicting postpartum depression: scoping review. *JMIR mental health*, 8(11):e29838, 2021.

[21] Rajkumar Ettiyan and V Geetha. Iod-nets–an iot based intelligent health care monitoring system for ambulatory pregnant mothers and fetuses. *Measurement: Sensors*, 27:100781, 2023.

[22] Marzia Ahmed and Mohammod Abul Kashem. Iot based risk level prediction model for maternal health care in the context of bangladesh. In *2020 2nd International Conference on Sustainable Technologies for Industry 4.0 (STI)*, pages 1–6. IEEE, 2020.

[23] João Alexandre Lobo Marques, Tao Han, Wanqing Wu, João Paulo do Vale Madeiro, Aloísio Vieira Lira Neto, Raffaele Gravina, Giancarlo Fortino, and Victor Hugo C de Albuquerque. Iot-based smart health system for ambulatory maternal and fetal monitoring. *IEEE Internet of Things Journal*, 8(23):16814–16824, 2020.

[24] BN Lakshmi, TS Indumathi, and Nandini Ravi. An hybrid approach for prediction based health monitoring in pregnant women. *Procedia Technology*, 24:1635–1642, 2016.

[25] Lokesh Pawar, Janvi Malhotra, Astha Sharma, Diya Arora, and Devanshi Vaidya. A robust machine learning predictive model for maternal health risk. In *2022 3rd international conference on electronics and sustainable communication systems (ICESC)*, pages 882–888. IEEE, 2022.

[26] Kausthav Pratim Kalita, Sarat Kumar Chettri, and Rup Kumar Deka. A blockchain-based model for maternal health information exchange and prediction of health risks using machine learning. In *2023 International Conference on Intelligent and Innovative Technologies in Computing, Electrical and Electronics (IITCEE)*, pages 1184–1189. IEEE, 2023.

[27] Md Alif Sheakh, Mst Sazia Tahosin, Md Maruf Hasan, Taminul Islam, Onamika Islam, and Md Masum Rana. Child and maternal mortality risk factor analysis using machine learning approaches. In *2023 11th International Symposium on Digital Forensics and Security (ISDFS)*, pages 1–6. IEEE, 2023.

[28] Md Assaduzzaman, Abdullah Al Mamun, and Md Zahid Hasan. Early prediction of maternal health risk factors using machine learning techniques. In *2023 international conference for advancement in technology (ICONAT)*, pages 1–6. IEEE, 2023.

[29] Dawit S Bogale, Tesfamariam M Abuhay, and Belayneh E Dejene. Predicting perinatal mortality based on maternal health status and health insurance service using homogeneous ensemble machine learning methods. *BMC Medical Informatics and Decision Making*, 22(1):341, 2022.

[30] Isaac Kofi Nti and Bridgitte Owusu-Boadu. A hybrid boosting ensemble model for predicting maternal mortality and sustaining reproductive. *Smart Health*, 26:100325, 2022.

[31] Sirinat Wanriko, Narit Hnoohom, Konlakom Wongpatikaseree, Anuchit Jitpattanakul, and Olarik Musigavong. Risk assessment of pregnancy-induced hypertension using a machine learning approach. In *2021 joint international conference on digital arts, media and technology with ECTI northern section conference on electrical, electronics, computer and telecommunication engineering*, pages 233–237. IEEE, 2021.

[32] Bincy Varghese, Aishwarya Jala, Soumya Meka, Deepthi Adla, Shraddha Jangili, RK Talukdar, Srinivasa Rao Mutheneni, Roshan M Borkar, and Ramu Adela. Integrated metabolomics and machine learning approach to predict hypertensive disorders of pregnancy. *American Journal of Obstetrics & Gynecology MFM*, 5(2):100829, 2023.

[33] Oknalita Simbolon, Melyana Nurul Widyawati, Kurnianingsih Kurnianingsih, Naoyuki Kubota, and Nawi Ng. Predicting the risk of preeclampsia using soft voting-based ensemble and its recommendation. In *2020 International Symposium on Community-centric Systems (CcS)*, pages 1–6. IEEE, 2020.

[34] Xinyuan Zhang, Yu Chen, Stephen Salerno, Yi Li, Libin Zhou, Xiaoxi Zeng, and Huafeng Li. Prediction of severe preeclampsia in machine learning. *Medicine in Novel Technology and Devices*, 15:100158, 2022.

[35] Alexander D Gleed, Qingchao Chen, James Jackman, Divyanshu Mishra, Varun Chandramohan, Alice Self, Shinjini Bhatnagar, Aris T Papageorghiou, and J Alison Noble. Automatic image guidance for assessment of placenta location in ultrasound video sweeps. *Ultrasound in Medicine & Biology*, 49(1):106–121, 2023.

[36] Hao Zhu, Xuan Yin, Haijie Wang, Yida Wang, Xuefen Liu, Chenglong Wang, Xiaotian Li, Yuanyuan Lu, Guang Yang, and He Zhang. A computerized diagnostic model for automatically evaluating placenta accrete spectrum disorders based on the combined mr radiomics-clinical signatures. *Scientific Reports*, 12(1):10130, 2022.

[37] Rongrong Xuan, Tao Li, Yutao Wang, Jian Xu, and Wei Jin. Prenatal prediction and typing of placental invasion using mri deep and radiomic features. *BioMedical Engineering OnLine*, 20(1):56, 2021.

[38] Munetoshi Akazawa, Kazunori Hashimoto, Noda Katsuhiko, and Yoshida Kaname. Machine learning approach for the prediction of postpartum hemorrhage in vaginal birth. *Scientific reports*, 11(1):22620, 2021.

[39] Yina Wu, Yichao Zhang, Xu Zou, Zhenming Yuan, Wensheng Hu, Sha Lu, Xiaoyan Sun, and Yingfei Wu. Estimated date of delivery with electronic medical records by a hybrid gbdt-gru model. *Scientific Reports*, 12(1):4892, 2022.

[40] Thiago Augusto Hernandes Rocha, Erika Bárbara Abreu Fonseca de Thomaz, Dante Grapiuna de Almeida, Núbia Cristina da Silva, Rejane Christine de Sousa Queiroz, Luciano Andrade, Luiz Augusto Facchini, Marcos Luiggi Lemos Sartori, Dalton Breno Costa, Marcos Adriano Garcia Campos, et al. Data-driven risk stratification for preterm birth in brazil: a population-based study to develop of a machine learning risk assessment approach. *The Lancet Regional Health–Americas*, 3, 2021.

[41] Kingsley Wong, Gizachew A Tessema, Kevin Chai, and Gavin Pereira. Development of prognostic model for preterm birth using machine learning in a population-based cohort of western australia births between 1980 and 2015. *Scientific Reports*, 12(1):19153, 2022.

[42] Felix Q Jin, Lindsey C Carlson, Timothy J Hall, Helen Feltovich, and Mark L Palmeri. Deep learning based quantitative uncertainty estimation for ultrasound shear wave elasticity imaging. In *2021 IEEE International Ultrasonics Symposium (IUS)*, pages 1–4. IEEE, 2021.

[43] Md Sakib Bin Alam, Muhammed JA Patwary, and Maruf Hassan. Birth mode prediction using bagging ensemble classifier: A case study of bangladesh. In *2021 International conference on information and communication technology for sustainable development (ICICT4SD)*, pages 95–99. IEEE, 2021.

[44] Tingting Hu, Sisi Du, Xiaoyan Li, Fang Yang, Shanshan Zhang, Jingjing Yi, Birong Xiao, Tingting Li, and Lin He. Establishment of a model for predicting the outcome of induced labor in full-term pregnancy based on machine learning algorithm. *Scientific Reports*, 12(1):19063, 2022.

[45] Nasratullah Nasrat and Kittichai Lavangnananda. Application of machine learning in assignment of child delivery service in afghanistan. In *2021 18th International Conference on Electrical Engineering/Electronics, Computer, Telecommunications and Information Technology (ECTICON)*, pages 1172–1175, 2021.

[46] Nasratullah Nasrat, Mohammad Dawood Babakerkhell, Gawhar Shah Gawhari, and Abdul Rahim Ahmadi. Implementation of a predictive model for skilled child delivery service use in afghanistan. In *2021 IEEE Intl Conf on Dependable, Autonomic and Secure Computing, Intl Conf on Pervasive Intelligence and Computing, Intl Conf on Cloud and Big Data Computing, Intl Conf on Cyber Science and Technology Congress (DASC/PiCom/CBDCom/CyberSciTech)*, pages 249–255. IEEE, 2021.

[47] Brook Tesfaye, Suleman Atique, Tariq Azim, and Mihiretu M Kebede. Predicting skilled delivery service use in ethiopia: dual application of logistic regression and machine learning algorithms. *BMC medical informatics and decision making*, 19:1–10, 2019.

[48] Md Saiful Islam, Md Abdul Awal, Jinnaton Nessa Laboni, Farhana Tazmim Pinki, Shatu Karmokar, Khondoker Mirazul Mumenin, Saad Al-Ahmadi, Md Ashfikur Rahman, Md Shahadat Hossain, and Seyedali Mirjalili. Hgsorf: Henry gas solubility optimization-based random forest for c-section prediction and xai-based cause analysis. *Computers in Biology and Medicine*, 147:105671, 2022.

[49] Yi-Ting Tsai, Isabel R Fulcher, Tracey Li, Felix Sukums, and Bethany Hedt-Gauthier. Predicting facility-based delivery in zanzibar: The vulnerability of machine learning algorithms to adversarial attacks. *Heliyon*, 9(5), 2023.

[50] Jingying Huang, Ping Lv, Yunzhi Lian, Meihua Zhang, Xin Ge, Shuheng Li, Yingxia Pan, Jiangman Zhao, Yue Xu, Hui Tang, et al. Construction of machine learning tools to predict threatened miscarriage in the first trimester based on aea, progesterone and β-hcg in china: a multicentre, observational, case-control study. *BMC Pregnancy and Childbirth*, 22(1):697, 2022.

[51] Dadoma Sherpa, Rajwade Dhruva Abhijit, Imon Mitra, Dhruba Dhar, Sunita Sharma, Pratip Chakraborty, and Koel Chaudhury. Prediction of idiopathic recurrent spontaneous miscarriage using machine learning. In *2023 International Conference on Computer, Electrical & Communication Engineering (ICCECE)*, pages 1–8. IEEE, 2023.

[52] Hiba Asri and Zahi Jarir. Toward a smart health: big data analytics and iot for real-time miscarriage prediction. *Journal of Big Data*, 10(1):34, 2023.

[53] Aki Koivu and Mikko Sairanen. Predicting risk of stillbirth and preterm pregnancies with machine learning. *Health information science and systems*, 8(1):14, 2020.

[54] Md Shamimul Islam, Md Mahedi Hasan, Sohaib Abdullah, Jalal Uddin Md Akbar, NHM Arafat, and Saydul Akbar Murad. A deep spatio-temporal network for vision-based sexual harassment detection. In *2021 Emerging Technology in Computing, Communication and Electronics (ETCCE)*, pages 1–6. IEEE, 2021.

[55] MD Tanzil Shahria, Farhat Tasnim Progga, Saif Ahmed, and Asma Arisha. Application of neural networks for detection of sexual harassment in workspace. In *2021 international conference on advances in electrical, computing, communication and sustainable technologies (ICAECT)*, pages 1–4. IEEE, 2021.

[56] Lateef B Amusa, Annah V Bengesai, and Hafiz TA Khan. Predicting the vulnerability of women to intimate partner violence in south africa: Evidence from tree-based machine learning techniques. *Journal of interpersonal violence*, 37(7-8):NP5228–NP5245, 2022.

[57] Nemias Saboya, A Angel Sullon, and Omar L Loaiza. Predictive model based on machine learning for the detection of physically mistreated women in the peruvian scope. In *Proceedings of the 3rd International Conference on Advances in Artificial Intelligence*, pages 18–23, 2019.

[58] Emad Alawneh, Mohammad Al-Fawa'reh, Mousa Tayseer Jafar, and Mustafa Al Fayoumi. Sentiment analysis-based sexual harassment detection using machine learning techniques. In *2021 international symposium on electronics and smart devices (ISESD)*, pages 1–6. IEEE, 2021.

[59] Mujahidul Islam, Maqsudur Rahman, Md Tofael Ahmed, Abu Zafor Muhammad Islam, Dipankar Das, and Mohammed Moshiul Hoque. Sexual harassment detection using machine learning and deep learning techniques for bangla text. In *2023 International Conference on Electrical, Computer and Communication Engineering (ECCE)*, pages 1–6. IEEE, 2023.

[60] Muammar Notareza Ramadhan, Indra Budi, Aris Budi Santoso, and Ryan Randy Suryono. Sexual violence classification as hate speech using indonesian tweet. In *2022 International Symposium on Information Technology and Digital Innovation (ISITDI)*, pages 114–120. IEEE, 2022.

[61] Vorada Socatiyanurak, Nittayapa Klangpornkun, Adirek Munthuli, Phongphan Phienphanich, Lalin Kovudhikulrungsri, Nantawat Saksakulkunakorn, Phonkanok Chairaungsri, and Charturong Tantibundhit. Law-u: Legal guidance through artificial intelligence chatbot for sexual violence victims and survivors. *IEEE Access*, 9:131440–131461, 2021.

[62] Jihan Ryu, Emese Sükei, Agnes Norbury, Shelley H Liu, Juan José Campaña-Montes, Enrique Baca-Garcia, Antonio Artés, and M Mercedes Perez-Rodriguez. Shift in social media app usage during covid-19 lockdown and clinical anxiety symptoms: machine learning–based ecological momentary assessment study. *JMIR mental health*, 8(9):e30833, 2021.

[63] Aamir Arsalan and Muhammad Majid. A study on multi-class anxiety detection using wearable eeg headband. *Journal of Ambient Intelligence and Humanized Computing*, 13(12):5739–5749, 2022.

[64] Asma Baghdadi, Yassine Aribi, Rahma Fourati, Najla Halouani, Patrick Siarry, and Adel Alimi. Psychological stimulation for anxious states detection based on eeg-related features. *Journal of Ambient Intelligence and Humanized Computing*, 12:8519–8533, 2021.

[65] Xiangmin Meng and Jie Zhang. Anxiety recognition of college students using a takagi-sugeno-kang fuzzy system modeling method and deep features. *IEEE Access*, 8:159897–159905, 2020.

[66] Hao Xiong, Shlomo Berkovsky, Mia Romano, Roneel V Sharan, Sidong Liu, Enrico Coiera, and Lauren F McLellan. Prediction of anxiety disorders using a feature ensemble based bayesian neural network. *Journal of Biomedical Informatics*, 123:103921, 2021.

[67] Joo Young Chun, Hyun-Jin Kim, Ji-Won Hur, Dooyoung Jung, Heon-Jeong Lee, Seung Pil Pack, Sungkil Lee, Gerard Kim, Chung-Yean Cho, Seung-Moo Lee, et al. Prediction of specific anxiety symptoms and virtual reality sickness using in situ autonomic physiological signals during virtual reality treatment in patients with social anxiety disorder: mixed methods study. *JMIR serious games*, 10(3):e38284, 2022.

[68] Chan-Hen Tsai, Pei-Chen Chen, Ding-Shan Liu, Ying-Ying Kuo, Tsung-Ting Hsieh, Dai-Lun Chiang, Feipei Lai, Chia-Tung Wu, et al. Panic attack prediction using wearable devices and machine learning: development and cohort study. *JMIR Medical Informatics*, 10(2):e33063, 2022.

[69] E Syed Mohamed, Tawseef Ahmad Naqishbandi, Syed Ahmad Chan Bukhari, Insha Rauf, Vilas Sawrikar, and Arshad Hussain. A hybrid mental health prediction model using support vector machine, multilayer perceptron, and random forest algorithms. *Healthcare Analytics*, 3:100185, 2023.

[70] Radwan Qasrawi, Stephanny Vicuna Polo, Diala Abu Al-Halawa, Sameh Hallaq, and Ziad Abdeen. Schoolchildren'depression and anxiety risk factors assessment and prediction: Machine learning techniques performance analysis. *JMIR Formative Research*, 2022.

[71] Adedoyin Aderinwale, Gemechu Bekele Tolossa, Ah Young Kim, Eun Hye Jang, Yong-il Lee, Hong Jin Jeon, Hyewon Kim, Han Young Yu, and Jaeseung Jeong. Two-channel eeg based diagnosis of panic disorder and major depressive disorder using machine learning and non-linear dynamical methods. *Psychiatry Research: Neuroimaging*, 332:111641, 2023.

[72] Kazuhiro Watanabe and Akizumi Tsutsumi. The passive monitoring of depression and anxiety among workers using digital biomarkers based on their physical activity and working conditions: 2-week longitudinal study. *JMIR Formative Research*, 6(11):e40339, 2022.

[73] Shereen A Hussein, Abd El Rahman S Bayoumi, and Ayat M Soliman. Automated detection of human mental disorder. *Journal of Electrical Systems and Information Technology*, 10(1):9, 2023.

[74] Wessel A Van Eeden, Chuan Luo, Albert M van Hemert, Ingrid VE Carlier, Brenda W Penninx, Klaas J Wardenaar, Holger Hoos, and Erik J Giltay. Predicting the 9-year course of mood and anxiety disorders with automated machine learning: A comparison between auto-sklearn, naïve bayes classifier, and traditional logistic regression. *Psychiatry Research*, 299:113823, 2021.

[75] Prince Kumar, Shruti Garg, and Ashwani Garg. Assessment of anxiety, depression and stress using machine learning models. *Procedia Computer Science*, 171:1989–1998, 2020.

[76] Vanessa Borba De Souza, Jeferson Campos Nobre, and Karin Becker. Dac stacking: a deep learning ensemble to classify anxiety, depression, and their comorbidity from reddit texts. *IEEE Journal of Biomedical and Health Informatics*, 26(7):3303–3311, 2022.

[77] Lana G Tennenhouse, Ruth Ann Marrie, Charles N Bernstein, Lisa M Lix, et al. Machine-learning models for depression and anxiety in individuals with immune-mediated inflammatory disease. *Journal of psychosomatic research*, 134:110126, 2020.

[78] Ruksana Shaukat-Jali, Nejra van Zalk, David Edward Boyle, et al. Detecting subclinical social anxiety using physiological data from a wrist-worn wearable: small-scale feasibility study. *JMIR Formative Research*, 5(10):e32656, 2021.

[79] Nicholas C Jacobson, Berta Summers, and Sabine Wilhelm. Digital biomarkers of social anxiety severity: digital phenotyping using passive smartphone sensors. *Journal of medical Internet research*, 22(5):e16875, 2020.

[80] Vikas Kamra, Praveen Kumar, and Masoud Mohammadian. An intelligent disease prediction system for psychological diseases by implementing hybrid hopfield recurrent neural network approach. *Intelligent Systems with Applications*, 18:200208, 2023.

[81] Fidel Cacheda, Diego Fernandez, Francisco J Novoa, and Victor Carneiro. Early detection of depression: social network analysis and random forest techniques. *Journal of medical Internet research*, 21(6):e12554, 2019.

[82] David Owen, Dimosthenis Antypas, Athanasios Hassoulas, Antonio F Pardiñas, Luis Espinosa-Anke, Jose Camacho Collados, et al. Enabling early health care intervention by detecting depression in users of web-based forums using language models: longitudinal analysis and evaluation. *JMIR AI*, 2(1):e41205, 2023.

[83] Tapotosh Ghosh, Md Hasan Al Banna, Md Jaber Al Nahian, Mohammed Nasir Uddin, M Shamim Kaiser, and Mufti Mahmud. An attention-based hybrid architecture with explainability for depressive social media text detection in bangla. *Expert Systems with Applications*, 213:119007, 2023.

[84] Hui Zhang, Hong Wang, Shu Han, Wei Li, and Luhe Zhuang. Detecting depression tendency with multimodal features. *Computer Methods and Programs in Biomedicine*, 240:107702, 2023.

[85] Ju Chun Cheng and Arbee LP Chen. Multimodal time-aware attention networks for depression detection. *Journal of Intelligent Information Systems*, 59(2):319–339, 2022.

[86] Lu Ren, Hongfei Lin, Bo Xu, Shaowu Zhang, Liang Yang, and Shichang Sun. Depression detection on reddit with an emotion-based attention network: algorithm development and validation. *JMIR medical informatics*, 9(7):e28754, 2021.

[87] Xiaofeng Wang, Shuai Chen, Tao Li, Wanting Li, Yejie Zhou, Jie Zheng, Qingcai Chen, Jun Yan, Buzhou Tang, et al. Depression risk prediction for chinese microblogs via deep-learning methods: Content analysis. *JMIR medical informatics*, 8(7):e17958, 2020.

[88] Ah Young Kim, Eun Hye Jang, Seung-Hwan Lee, Kwang-Yeon Choi, Jeon Gue Park, and Hyun-Chool Shin. Automatic depression detection using smartphone-based text-dependent speech signals: deep convolutional neural network approach. *Journal of medical Internet research*, 25:e34474, 2023.

[89] Alfredo Daza Vergaray, Juan Carlos Herrera Miranda, Juana Bobadilla Cornelio, Atilio Rubén López Carranza, and Carlos Fidel Ponce Sánchez. Predicting the depression in university students using stacking ensemble techniques over oversampling method. *Informatics in medicine unlocked*, 41:101295, 2023.

[90] Akshi Kumar, Saurabh Raj Sangwan, Anshika Arora, and Varun G Menon. Depress-dcnf: A deep convolutional neuro-fuzzy model for detection of depression episodes using iomt. *Applied Soft Computing*, 122:108863, 2022.

[91] Amal Bouzid, Abdulrahman Almidani, Maria Zubrikhina, Altyngul Kamzanova, Burcu Yener Ilce, Manzura Zholdassova, Ayesha M Yusuf, Poorna Manasa Bhamidimarri, Hamid A AlHaj, Almira Kustubayeva, et al. Integrative bioinformatics and artificial intelligence analyses of transcriptomics data identified genes associated with major depressive disorders including nrg1. *Neurobiology of stress*, 26:100555, 2023.

[92] Daniela Caldirola, Silvia Daccò, Francesco Cuniberti, Massimiliano Grassi, Alessandra Alciati, Tatiana Torti, and Giampaolo Perna. First-onset major depression during the covid-19 pandemic: A predictive machine learning model. *Journal of Affective Disorders*, 310:75–86, 2022.

[93] Kyoung-Sae Na, Seo-Eun Cho, Zong Woo Geem, and Yong-Ku Kim. Predicting future onset of depression among community dwelling adults in the republic of korea using a machine learning algorithm. *Neuroscience Letters*, 721:134804, 2020.

[94] Nikhil Marriwala, Deepti Chaudhary, et al. A hybrid model for depression detection using deep learning. *Measurement: Sensors*, 25:100587, 2023.

[95] Kirill Milintsevich, Kairit Sirts, and Gaël Dias. Towards automatic text-based estimation of depression through symptom prediction. *Brain informatics*, 10(1):4, 2023.

[96] Kennedy Opoku Asare, Isaac Moshe, Yannik Terhorst, Julio Vega, Simo Hosio, Harald Baumeister, Laura Pulkki-Råback, and Denzil Ferreira. Mood ratings and digital biomarkers from smartphone and wearable data differentiates and predicts depression status: A longitudinal data analysis. *Pervasive and Mobile Computing*, 83:101621, 2022.

[97] Ada Dogrucu, Alex Perucic, Anabella Isaro, Damon Ball, Ermal Toto, Elke A Rundensteiner, Emmanuel Agu, Rachel Davis-Martin, and Edwin Boudreaux. Moodable: On feasibility of instantaneous depression assessment using machine learning on voice samples with retrospectively harvested smartphone and social media data. *Smart Health*, 17:100118, 2020.

[98] Kennedy Opoku Asare, Yannik Terhorst, Julio Vega, Ella Peltonen, Eemil Lagerspetz, and Denzil Ferreira. Predicting depression from smartphone behavioral markers using machine learning methods, hyperparameter optimization, and feature importance analysis: exploratory study. *JMIR mHealth and uHealth*, 9(7):e26540, 2021.

[99] Laura M Lyall, Natasha Sangha, Xingxing Zhu, Donald M Lyall, Joey Ward, Rona J Strawbridge, Breda Cullen, and Daniel J Smith. Subjective and objective sleep and circadian parameters as predictors of depression-related outcomes: a machine learning approach in uk biobank. *Journal of Affective Disorders*, 335:83–94, 2023.

[100] Chenyang Zhang, Xiaofei Chen, Song Wang, Junjun Hu, Chunpeng Wang, and Xin Liu. Using catboost algorithm to identify middle-aged and elderly depression, national health and nutrition examination survey 2011–2018. *Psychiatry research*, 306:114261, 2021.

[101] Rokeya Siddiqua, Nusrat Islam, Jarba Farnaz Bolaka, Riasat Khan, and Sifat Momen. Aida: Artificial intelligence based depression assessment applied to bangladeshi students. *Array*, 18:100291, 2023.

[102] Lvchun Cui, Chenglei Wang, Zhiguo Wu, Daihui Peng, Jingjing Huang, Chen Zhang, Jia Huang, Wu Hong, Yong Wang, Jun Chen, et al. Symptomatology differences of major depression in psychiatric versus general hospitals: A machine learning approach. *Journal of affective disorders*, 260:349–360, 2020.

[103] Yu Wu, Binbin Su, Chen Chen, Yihao Zhao, Panliang Zhong, and Xiaoying Zheng. Urban-rural disparities in the prevalence and trends of depressive symptoms among chinese elderly and their associated factors. *Journal of Affective Disorders*, 340:258–268, 2023.

[104] Md Zia Uddin, Kim Kristoffer Dysthe, Asbjørn Følstad, and Petter Bae Brandtzaeg. Deep learning for prediction of depressive symptoms in a large textual dataset. *Neural Computing and Applications*, 34(1):721–744, 2022.

[105] Kuk-In Jang, Sungkean Kim, Jeong-Ho Chae, and Chany Lee. Machine learning-based classification using electroencephalographic multi-paradigms between drug-naïve patients with depression and healthy controls. *Journal of Affective Disorders*, 338:270–277, 2023.

[106] Danniel Shazmeer Bin Abdul Hamid, SB Goyal, and Pradeep Bedi. Integration of deep learning for improved diagnosis of depression using eeg and facial features. *Materials Today: Proceedings*, 80:1965–1969, 2023.

[107] Ayan Seal, Rishabh Bajpai, Mohan Karnati, Jagriti Agnihotri, Anis Yazidi, Enrique Herrera-Viedma, and Ondrej Krejcar. Benchmarks for machine learning in depression discrimination using electroencephalography signals. *Applied Intelligence*, 53(10):12666–12683, 2023.

[108] Sukit Suparatpinyo and Nuanwan Soonthornphisaj. Smart voice recognition based on deep learning for depression diagnosis. *Artificial Life and Robotics*, 28(2):332–342, 2023.

[109] Heejung Kim, SungHee Lee, SangEun Lee, Soyun Hong, HeeJae Kang, Namhee Kim, et al. Depression prediction by using ecological momentary assessment, actiwatch data, and machine learning: observational study on older adults living alone. *JMIR mHealth and uHealth*, 7(10):e14149, 2019.

[110] Jae Sung Kim, Bohyun Wang, Meelim Kim, Jung Lee, Hyungjun Kim, Danyeul Roh, Kyung Hwa Lee, Soon-Beom Hong, Joon Shik Lim, Jae-Won Kim, et al. Prediction of diagnosis and treatment response in adolescents with depression by using a smartphone app and deep learning approaches: usability study. *JMIR formative research*, 7(1):e45991, 2023.

[111] Mudasir Ahmad Wani, Mohammad A ELAffendi, Kashish Ara Shakil, Ali Shariq Imran, and Ahmed A Abd El-Latif. Depression screening in humans with ai and deep learning techniques. *IEEE transactions on computational social systems*, 10(4):2074–2089, 2022.

[112] Duyan Geng, Qiang An, Zhigang Fu, Chao Wang, and Hongxia An. Identification of major depression patients using machine learning models based on heart rate variability during sleep stages for pre-hospital screening. *Computers in Biology and Medicine*, 162:107060, 2023.

[113] Yuri Rykov, Thuan-Quoc Thach, Iva Bojic, George Christopoulos, and Josip Car. Digital biomarkers for depression screening with wearable devices: cross-sectional study with machine learning modeling. *JMIR mHealth and uHealth*, 9(10):e24872, 2021.

[114] Tahsin Mullick, Ana Radovic, Sam Shaaban, Afsaneh Doryab, et al. Predicting depression in adolescents using mobile and wearable sensors: multimodal machine learning–based exploratory study. *JMIR Formative Research*, 6(6):e35807, 2022.

[115] Yang Liu, Jeffrey Hankey, Bo Cao, and Pratap Chokka. Screening for major depressive disorder in a tertiary mental health centre using earlydetect: A machine learning-based pilot study. *Journal of affective disorders reports*, 3:100062, 2021.

[116] Ran Bai, Le Xiao, Yu Guo, Xuequan Zhu, Nanxi Li, Yashen Wang, Qinqin Chen, Lei Feng, Yinghua Wang, Xiangyi Yu, et al. Tracking and monitoring mood stability of patients with major depressive disorder by machine learning models using passive digital data: prospective naturalistic multicenter study. *JMIR mHealth and uHealth*, 9(3):e24365, 2021.

[117] Zhiwei Bao, Xinyi Zhao, Jingjing Li, Guanghua Zhang, Hairong Wu, Yuping Ning, Ming D Li, and Zhongli Yang. Prediction of repeated-dose intravenous ketamine response in major depressive disorder using the gwas-based machine learning approach. *Journal of Psychiatric Research*, 138:284–290, 2021.

[118] Aidan Cousins, Lucas Nakano, Emma Schofield, and Rasa Kabaila. A neural network approach to optimising treatments for depression using data from specialist and community psychiatric services in australia, new zealand and japan. *Neural Computing and Applications*, 35(16):11497–11516, 2023.

[119] Muhammad Khubayeeb Kabir, Maisha Islam, Anika Nahian Binte Kabir, Adiba Haque, and Md Khalilur Rhaman. Detection of depression severity using bengali social media posts on mental health: study using natural language processing techniques. *JMIR Formative Research*, 6(9):e36118, 2022.

[120] Mariko Makhmutova, Raghu Kainkaryam, Marta Ferreira, Jae Min, Martin Jaggi, Ieuan Clay, et al. Predicting changes in depression severity using the psyche-d (prediction of severity change-depression) model involving person-generated health data: longitudinal case-control observational study. *JMIR mHealth and uHealth*, 10(3):e34148, 2022.

[121] Soumitra Ghosh, Asif Ekbal, and Pushpak Bhattacharyya. A multitask framework to detect depression, sentiment and multi-label emotion from suicide notes. *Cognitive Computation*, 14(1):110–129, 2022.

[122] Sam Andersson, Deepti R Bathula, Stavros I Iliadis, Martin Walter, and Alkistis Skalkidou. Predicting women with depressive symptoms postpartum with machine learning methods. *Scientific reports*, 11(1):7877, 2021.

[123] Yiye Zhang, Shuojia Wang, Alison Hermann, Rochelle Joly, and Jyotishman Pathak. Development and validation of a machine learning algorithm for predicting the risk of postpartum depression among pregnant women. *Journal of affective disorders*, 279:1–8, 2021.

[124] Weina Zhang, Han Liu, Vincent Michael Bernard Silenzio, Peiyuan Qiu, Wenjie Gong, et al. Machine learning models for the prediction of postpartum depression: application and comparison based on a cohort study. *JMIR medical informatics*, 8(4):e15516, 2020.

[125] Colin Wakefield and Martin G Frasch. Predicting patients requiring treatment for depression in the postpartum period using common electronic medical record data available antepartum. *AJPM focus*, 2(3):100100, 2023.

[126] Kim S Betts, Steve Kisely, and Rosa Alati. Predicting postpartum psychiatric admission using a machine learning approach. *Journal of psychiatric research*, 130:35–40, 2020.

[127] Hana Andrina. Postpartum depression: A literature review of screening and prevention. *Scientia Psychiatrica*, 2(4):183–192, 2021.

[128] Viviane Valdes, Anne E Berens, and Charles A Nelson III. Socioeconomic and psychological correlates of postpartum depression at 6 months in dhaka, bangladesh. *International Journal of Psychology*, 56(5):729–738, 2021.

[129] Aditi Shelke and Swarupa Chakole. A review on risk factors of postpartum depression in india and its management. *Cureus*, 14(9), 2022.

[130] Iris Agrawal, Ashok M Mehendale, and Ritika Malhotra. Risk factors of postpartum depression. *Cureus*, 14(10), 2022.

[131] Huong Thi Thanh Nguyen, Ly Thi Kim Do, Huong Thi Thu Pham, Anh Phuong Hoang, Hoa Thi Truong, and Huyen Thi Hoa Nguyen. The symptoms of postpartum depression observed by family members: A pilot study. *Frontiers in Psychiatry*, 13:897175, 2022.

[132] Maryam Ghaedrahmati, Ashraf Kazemi, Gholamreza Kheirabadi, Amrollah Ebrahimi, and Masood Bahrami. Postpartum depression risk factors: A narrative review. *Journal of education and health promotion*, 6(1):60, 2017.

[133] Shengfei Cao, Taowenxuan Yan, and Jikai Yang. Analysis of risk factors of postpartum depression. *Theoretical and Natural Science*, 4:504–510, 2023.

[134] Che Wan Jasimah Bt Wan Mohamed Radzi, Hashem Salarzadeh Jenatabadi, and Nadia Samsudin. Postpartum depression symptoms in survey-based research: a structural equation analysis. *BMC public health*, 21:1–12, 2021.

[135] Gülay Pamuk and Yusuf Adnan Güçlü. Prevalence and accompanying factors for postpartum depression symptoms. *Family Practice and Palliative Care*, 7(1):18–23, 2022.

[136] Molly Fox, Curt A Sandman, Elysia Poggi Davis, and Laura M Glynn. A longitudinal study of women's depression symptom profiles during and after the postpartum phase. *Depression and anxiety*, 35(4):292–304, 2018.

[137] Chiara Gastaldon, Marco Solmi, Christoph U Correll, Corrado Barbui, and Georgios Schoretsanitis. Risk factors of postpartum depression and depressive symptoms: umbrella review of current evidence from systematic reviews and meta-analyses of observational studies. *The British Journal of Psychiatry*, 221(4):591–602, 2022.

[138] Cecilia Lao, Jo Lane, Hanna Suominen, et al. Analyzing suicide risk from linguistic features in social media: evaluation study. *JMIR formative research*, 6(8):e35563, 2022.

[139] Raymond Su, James Rufus John, and Ping-I Lin. Machine learning-based prediction for self-harm and suicide attempts in adolescents. *Psychiatry research*, 328:115446, 2023.

[140] Adonias C de Oliveira, Evandro JS Diniz, Silmar Teixeira, and Ariel S Teles. How can machine learning identify suicidal ideation from user's texts? towards the explanation of the boamente system. *Procedia Computer Science*, 206:141–150, 2022.

[141] Junren Wang, Jiajun Qiu, Ting Zhu, Yu Zeng, Huazhen Yang, Yanan Shang, Jin Yin, Yajing Sun, Yuanyuan Qu, Unnur A Valdimarsdóttir, et al. Prediction of suicidal behaviors in the middle-aged population: machine learning analyses of uk biobank. *JMIR public health and surveillance*, 9(1):e43419, 2023.

[142] Polona Rus Prelog, Teodora Matić, Peter Pregelj, and Aleksander Sadikov. A pilot predictive model based on covid-19 data to assess suicidal ideation indirectly. *Journal of Psychiatric Research*, 163:318–324, 2023.

[143] Sooyeon Min, Daun Shin, Sang Jin Rhee, C Hyung Keun Park, Jeong Hun Yang, Yoojin Song, Min Ji Kim, Kyungdo Kim, Won Ik Cho, Oh Chul Kwon, et al. Acoustic analysis of speech for screening for suicide risk: machine learning classifiers for between-and within-person evaluation of suicidality. *Journal of medical internet research*, 25:e45456, 2023.

[144] Seo-Eun Cho, Zong Woo Geem, and Kyoung-Sae Na. Prediction of suicide among 372,813 individuals under medical check-up. *Journal of psychiatric research*, 131:9–14, 2020.

[145] Anmol Arora, Louis Bojko, Santosh Kumar, Joseph Lillington, Sukhmeet Panesar, and Bruno Petrungaro. Assessment of machine learning algorithms in national data to classify the risk of self-harm among young adults in hospital: a retrospective study. *International Journal of Medical Informatics*, 177:105164, 2023.

[146] Juliet Beni Edgcomb, Chi-hong Tseng, Mengtong Pan, Alexandra Klomhaus, and Bonnie T Zima. Assessing detection of children with suicide-related emergencies: evaluation and development of computable phenotyping approaches. *JMIR mental health*, 10:e47084, 2023.

[147] Kelly L Zuromski, Samantha L Bernecker, Carol Chu, Chelsey R Wilks, Peter M Gutierrez, Thomas E Joiner, Howard Liu, James A Naifeh, Matthew K Nock, Nancy A Sampson, et al. Pre-deployment predictors of suicide attempt during and after combat deployment: Results from the army study to assess risk and resilience in servicemembers. *Journal of psychiatric research*, 121:214–221, 2020.

[148] Hongxia Lu, Alex Barrett, Albert Pierce, Jianwei Zheng, Yun Wang, Chun Chiang, and Cyril Rakovski. Predicting suicidal and self-injurious events in a correctional setting using ai algorithms on unstructured medical notes and structured data. *Journal of psychiatric research*, 160:19–27, 2023.

[149] Shaoxiong Ji, Xue Li, Zi Huang, and Erik Cambria. Suicidal ideation and mental disorder detection with attentive relation networks. *Neural Computing and Applications*, 34(13):10309–10319, 2022.

[150] Syed Tanzeel Rabani, Akib Mohi Ud Din Khanday, Qamar Rayees Khan, Umar Ayoub Hajam, Ali Shariq Imran, and Zenun Kastrati. Detecting suicidality on social media: Machine learning at rescue. *Egyptian Informatics Journal*, 24(2):291–302, 2023.

[151] Dheeraj Kodati and Ramakrishnudu Tene. Identifying suicidal emotions on social media through transformer-based deep learning. *Applied Intelligence*, 53(10):11885–11917, 2023.

[152] En-Liang Wu, Chia-Yi Wu, Ming-Been Lee, Kuo-Chung Chu, and Ming-Shih Huang. Development of internet suicide message identification and the monitoring-tracking-rescuing model in taiwan. *Journal of affective disorders*, 320:37–41, 2023.

[153] Hannah Yao, Sina Rashidian, Xinyu Dong, Hongyi Duanmu, Richard N Rosenthal, and Fusheng Wang. Detection of suicidality among opioid users on reddit: machine learning–based approach. *Journal of medical internet research*, 22(11):e15293, 2020.

[154] Jihad S Obeid, Jennifer Dahne, Sean Christensen, Samuel Howard, Tami Crawford, Lewis J Frey, Tracy Stecker, and Brian E Bunnell. Identifying and predicting intentional self-harm in electronic health record clinical notes: deep learning approach. *JMIR medical informatics*, 8(7):e17784, 2020.

[155] Hayoung Kim Donnelly, Yoonsun Han, Suna Kim, and Dong Hun Lee. Predictors of suicide ideation among south korean adolescents: A machine learning approach. *Journal of affective disorders*, 329:557–565, 2023.

[156] Brian Stasak, Julien Epps, Heather T Schatten, Ivan W Miller, Emily Mower Provost, and Michael F Armey. Read speech voice quality and disfluency in individuals with recent suicidal ideation or suicide attempt. *Speech Communication*, 132:10–20, 2021.

[157] Marika Cusick, Prakash Adekkanattu, Thomas R Campion Jr, Evan T Sholle, Annie Myers, Samprit Banerjee, George Alexopoulos, Yanshan Wang, and Jyotishman Pathak. Using weak supervision and deep learning to classify clinical notes for identification of current suicidal ideation. *Journal of psychiatric research*, 136:95–102, 2021.

[158] Juliet Beni Edgcomb, Trevor Shaddox, Gerhard Hellemann, and John O Brooks III. Predicting suicidal behavior and self-harm after general hospitalization of adults with serious mental illness. *Journal of psychiatric research*, 136:515–521, 2021.

[159] Wesllei Felipe Heckler, Luan Paris Feijó, Juliano Varella de Carvalho, and Jorge Luis Victória Barbosa. Thoth: An intelligent model for assisting individuals with suicidal ideation. *Expert Systems with Applications*, 233:120918, 2023.

[160] MD Parvez Mosaraf. Postpartum depression, 2023.

[161] Alberto Fernández, Salvador Garcia, Francisco Herrera, and Nitesh V Chawla. Smote for learning from imbalanced data: progress and challenges, marking the 15-year anniversary. *Journal of artificial intelligence research*, 61:863–905, 2018.

[162] John Hancock, Taghi M Khoshgoftaar, and Justin M Johnson. The effects of random undersampling for big data medicare fraud detection. In *2022 IEEE International Conference on Service-Oriented System Engineering (SOSE)*, pages 141–146. IEEE, 2022.

[163] Gustavo Sosa-Cabrera, Santiago Gómez-Guerrero, Miguel García-Torres, and Christian E Schaerer. Feature selection: A perspective on inter-attribute cooperation. *International Journal of Data Science and Analytics*, 17(2):139–151, 2024.

[164] Giorgos Borboudakis and Ioannis Tsamardinos. Forward-backward selection with early dropping. *Journal of Machine Learning Research*, 20(8):1–39, 2019.

[165] Jialu Lin, Yili Zhang, Mei Xiong, and Hanyan Huang. Adaptive lasso based on regularized weight and its application in variable selection for regression. In *International Conference on Statistics, Data Science, and Computational Intelligence (CSDSCI 2022)*, volume 12510, pages 172–178. SPIE, 2023.

[166] Songshan Yang, Jiawei Wen, Xiang Zhan, and Daniel Kifer. Et-lasso: a new efficient tuning of lasso-type regularization for high-dimensional data. In *Proceedings of the 25th ACM SIGKDD international conference on knowledge discovery & data mining*, pages 607–616, 2019.

[167] Goedele Dierckx. Logistic regression model. *Encyclopedia of actuarial science*, 2, 2006.

[168] Paul Mather and Brandt Tso. *Classification methods for remotely sensed data*. CRC press, 2016.

[169] Vili Podgorelec, Peter Kokol, Bruno Stiglic, and Ivan Rozman. Decision trees: an overview and their use in medicine. *Journal of medical systems*, 26:445–463, 2002.

[170] Leo Breiman. Random forests. *Machine learning*, 45:5–32, 2001.

[171] Marina Sokolova and Guy Lapalme. A systematic analysis of performance measures for classification tasks. *Information processing & management*, 45(4):427–437, 2009.

[172] Shish Kumar Dubey, Sunil Gupta, Shri Krishna Pandey, and Sonu Mittal. Ethical considerations in data mining and database research. In *2024 11th International Conference on Reliability, Infocom Technologies and Optimization (Trends and Future Directions)(ICRITO)*, pages 1–4. IEEE, 2024.

[173] Thanaporn Phattanaviroj, Massoud Moslehpour, and Ankita Manohar Walawalkar. Data ethics and privacy. In *Challenges in Large Language Model Development and AI Ethics*, pages 321–353. IGI Global, 2024.