A HYBRID MACHINE LEARNING APPROACH TO DETECT POSTPARTUM DEPRESSION (PPD)

BY

Md Parvez Mosaraf
ID: 191-15-2348
AND

Yasin Khan Rabbi
ID: 191-15-2514

This Report Presented in Partial Fulfillment of the Requirements for the Degree of Bachelor of Science in Computer Science and Engineering

Supervised By

Ms Fatema Tuj Johora
Senior Lecturer
Department of CSE
Daffodil International University

Co-Supervised By

Md Mahfujur Rahman
Senior Lecturer
Department of CSE
Daffodil International University

DAFFODIL INTERNATIONAL UNIVERSITY

DHAKA, BANGLADESH

JANUARY 2023

APPROVAL

This Project titled "**A Hybrid Machine Learning Approach to Detect Postpartum Depression (PPD)**", submitted by Md Parvez Mosaraf, Yasin Khan Rabbi and ID No: 191-15-2348, 191-15-2514 Student ID to the Department of Computer Science and Engineering, Daffodil International University, has been accepted as satisfactory for the partial fulfillment of the requirements for the degree of B.Sc. in Computer Science and Engineering and approved as to its style and contents. The presentation has been held on 23 January 2023.

BOARD OF EXAMINERS

____

Dr. Touhid Bhuiyan — **Chairman**
Professor and Head
Department of Computer Science and Engineering
Faculty of Science & Information Technology
Daffodil International University

____

Dr. Md. Tarek Habib — **Internal Examiner**
Associate Professor
Department of Computer Science Engineering
Faculty of Science & Information Technology
Daffodil International University

____

Tapasy Rabeya — **Internal Examiner**
Senior Lecturer
Department of Computer Science and Engineering
Faculty of Science & Information Technology
Daffodil International University

23-01-23
____

Dr. Dewan Md Farid — **External Examiner**
Professor
Department of Computer Science and Engineering
United International University

DECLARATION

We hereby declare that, this project has been done by us under the supervision of **Ms Fatema Tuj Johora, Senior Lecturer, Department of CSE,** Daffodil International University. We also declare that neither this project nor any part of this project has been submitted elsewhere for award of any degree or diploma.

Supervised by:

Fatema 22.01.23

____

Ms Fatema Tuj Johora
Senior Lecturer
Department of CSE
Daffodil International University

Co-Supervised by:

____

Md Mahfujur Rahman
Senior Lecturer
Department of CSE
Daffodil International University

Submitted by:

Parvez

____

Md Parvez Mosaraf
ID: 191-15-2348
Department of CSE
Daffodil International University

Yasin

____

Yasin Khan Rabbi
ID: 191-15-2514
Department of CSE
Daffodil International University

ACKNOWLEDGEMENT

First, we express our heartiest thanks and gratefulness to almighty God for His divine blessing makes us possible to complete the final year project/internship successfully.

We really grateful and wish our profound our indebtedness to **Ms Fatema Tuj Johora**, **Senior Lecturer**, Department of CSE, Daffodil International University, Dhaka. Deep Knowledge & keen interest of our supervisor in the field of "*Machine Learning*" to carry out this project. His endless patience, scholarly guidance, continual encouragement, constant and energetic supervision, constructive criticism, valuable advice, reading many inferior drafts and correcting them at all stage have made it possible to complete this project.

We would like to express our heartiest gratitude to Dr. Touhid Bhuiyan, Professor, and Head**,** Department of CSE, for his kind help to finish our project and also to other faculty member and the staff of CSE department of Daffodil International University.

We would like to thank our entire course mate in Daffodil International University, who took part in this discuss while completing the course work.

Finally, we must acknowledge with due respect the constant support and patients of our parents.

ABSTRACT

Postpartum depression, which occurs in the days and weeks following childbirth, can have serious impacts on both mothers and babies. Symptoms of postpartum depression include mood swings, exhaustion, and a sense of hopelessness, and it can lead to long-term mood disorders such as postpartum psychosis. In some cases, this condition can even lead to maternal and infant mortality. Traditional methods of detecting postpartum depression, such as face-to-face doctor consultations, can be time-consuming and may not be feasible for individuals in remote areas. To address this challenge, we propose using a Hybrid machine learning algorithm, which is ensemble the four algorithm those are decision trees, K-nearest neighbors, logistic regression and support vector machines. We trained our hybrid model using our dataset of over 1503 sample, which has 15 different features. We got our best accuracy with our hybrid model; the accuracy is 98.78% with less error than the other traditional machine learning algorithm. We got the second-best accuracy with random forest 98.34%. Based on these results, we conclude that our hybrid model show the best performance for detecting the postpartum depression (PPD), while other machine learning algorithm exhibits the lowest performance.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

CHAPTER 1

Introduction

1.1 Introduction

One Giving birth has a major impact on a woman's identity, as she becomes a mother and her social role changes significantly. This can lead to significant changes in personality and can result in both physical and mental health issues. Postpartum depression, a mood disorder that can occur in the weeks and months following childbirth, is a common example of this. It is characterized by symptoms such as mood swings, exhaustion, and feelings of hopelessness, and can have serious consequences for both the mother and the child. In some cases, it can even lead to postpartum psychosis, a severe and potentially life-threatening mood disorder.

There are various factors that can contribute to the development of postpartum depression, including socio-demographic factors, social support, and physical activity. It is important to identify and treat postpartum depression as early as possible, as it can have significant impacts on the mother's health, the functioning of the family, and the development of the child. However, postpartum depression is often overlooked or underestimated, and the number of patients is increasing rapidly.

In this study, we aimed to develop machine learning-based predictive models for detecting postpartum depression. To do this, we collected data from pregnant women using questionnaires and focused on factors such as mental health, relationships with family members and partners, and experiences during pregnancy. We then used a variety of machine learning algorithms, including decision trees, K-nearest neighbors, logistic regression, AdaBoost, and support vector machines, to develop predictive models for detecting postpartum depression. We got our best accuracy using Hybrid model that is 98.33%. Our experiments yielded satisfactory accuracy rates, with decision trees achieving 97.82%, K-nearest neighbors reaching 95.34%, random forests achieving 97.66%, logistic regression achieving 77.82%, AdaBoost achieving 81.65%, and support vector machines achieving 89.80%.

Overall, our results suggest that machine learning has the potential to be a useful tool for detecting postpartum depression. While further research is needed to evaluate the long-term effectiveness of this approach, our findings suggest that it has the potential to improve the detection and treatment of postpartum depression, particularly in cases where traditional methods may not be feasible or effective. By identifying and treating postpartum depression early, we can help improve the health and well-being of new mothers and their families.

1.2 Motivation

Postpartum depression is a common and serious mental health condition that can occur in the days and weeks following childbirth. It is characterized by a range of symptoms, including mood swings, exhaustion, and a sense of hopelessness, and can have significant impacts on both the mother and the baby. In some cases, postpartum depression can lead to long-term mood disorders, such as postpartum psychosis, and can even result in maternal and infant mortality.

Traditionally, postpartum depression has been detected through face-to-face consultations with healthcare professionals. However, this approach can be time-consuming and may not be feasible for individuals living in remote areas. Moreover, manual detection of postpartum depression relies on the subjective judgment of healthcare professionals, which can lead to inconsistencies in diagnosis.

To address these challenges, we propose using a hybrid machine learning approach to detect postpartum depression. Machine learning algorithms are a type of artificial intelligence that can learn from data and make predictions or decisions without being explicitly programmed to do so. By using machine learning, we can develop an automated system for detecting postpartum depression that is objective and consistent.

A hybrid machine learning approach involves combining the strengths of multiple machine learning algorithms to improve the overall performance of the model. For example, we might use a decision tree algorithm to identify the most important features in the dataset, and then use a support vector machine to classify individuals as either having or not having postpartum depression based on those features. This approach can lead to better results than using a single machine learning algorithm in isolation.

There are several potential benefits to using a hybrid machine learning approach to detect postpartum depression. First and foremost, it can improve the accuracy of the model. By combining multiple algorithms, we can take advantage of their complementary strengths and reduce the risk of making incorrect predictions. This is especially important in the context of postpartum depression, where the consequences of misdiagnosis can be severe.
In addition to improving accuracy, a hybrid machine learning approach can also reduce the time and resources required to detect postpartum depression. By automating the process, we can significantly reduce the burden on healthcare professionals and make it easier for individuals to access the care they need. This is especially important in resource-limited settings, where manual detection may not be feasible.

Finally, a hybrid machine learning approach can be more robust and resistant to errors than a single algorithm. This is because the errors made by one algorithm can be compensated for by the other algorithms in the system, leading to improved overall performance.

In conclusion, a hybrid machine learning approach has the potential to revolutionize the way we detect postpartum depression. By combining the strengths of multiple algorithms, we can improve the accuracy and efficiency of the process and make it easier for individuals to access the care they need. This is an exciting area of research that has the potential to make a significant impact on the lives of new mothers and their families.

1.3 Rationale of the study

Postpartum depression is a common and often underdiagnosed mental health disorder that can have serious consequences for mothers and their families. Early detection and treatment of postpartum depression is essential for the well-being of mothers and their families, but traditional methods of detection, such as face-to-face doctor consultations, can be time-consuming and may not be feasible for individuals in remote areas. In this study, we propose a hybrid machine learning approach for detecting postpartum depression. Our approach combines multiple machine learning algorithms in order to take advantage of the strengths of each individual algorithm and to improve overall performance. We aim to evaluate the performance of our hybrid approach in comparison to the performance of individual algorithms, in order to identify the most effective approach for detecting postpartum depression. By improving the accuracy and efficiency of postpartum depression detection, our approach has the potential to improve the well-being of mothers and their families and to reduce the risk of long-term mood disorders.

1.4 Research Question

The research question for the study "A hybrid Machine learning approach to detect Postpartum depression" is:

- ***"What is the performance of a hybrid machine learning approach for detecting postpartum depression, and how does it compare to the performance of individual machine learning algorithms?"***

The main goal of this study is to develop and evaluate a hybrid machine learning approach for detecting postpartum depression. Postpartum depression is a common and often underdiagnosed mental health disorder that can have serious consequences for mothers and their families. Early detection and treatment of postpartum depression is essential for the well-being of mothers and their families, but traditional methods of detection, such as face-to-face doctor consultations, can be time-consuming and may not be feasible for individuals in remote areas.

In recent years, machine learning has emerged as a promising tool for detecting postpartum depression. Machine learning algorithms can analyze large amounts of data and identify patterns that may be difficult for humans to discern, making them well-suited for detecting complex conditions such as postpartum depression. However, the performance of machine learning algorithms can be affected by various factors, including the type of algorithm used and the quality of the training data.

In this study, we propose a hybrid machine learning approach for detecting postpartum depression, which combines multiple machine learning algorithms in order to take advantage of the strengths of each individual algorithm and to improve overall performance. Our research question is focused on evaluating the performance of our hybrid approach in comparison to the performance of individual algorithms, in order to identify the most effective approach for detecting postpartum depression.

1.5 Expected Outcome

The expected outcome of this study is to develop and evaluate a hybrid machine learning approach for detecting postpartum depression. We hope to achieve several specific goals through this research:

1. Improve the accuracy of postpartum depression detection: By combining the strengths of multiple machine learning algorithms, we hope to improve the accuracy of our approach in detecting postpartum depression.
2. Enhance the efficiency of postpartum depression detection: By automating the process of analysis, our approach has the potential to reduce the time and resources required for detection, making it more feasible for individuals in remote areas to access care.
3. Identify the most effective approach for detecting postpartum depression: By evaluating the performance of our hybrid approach in comparison to the performance of individual algorithms, we hope to identify the most effective approach for detecting postpartum depression.
4. Contribute to the development of machine learning techniques for detecting postpartum depression: Our research has the potential to contribute to the development of machine learning techniques for detecting postpartum depression, which may be useful for future research and clinical practice.

Overall, we hope that our research will lead to improved outcomes for mothers and their families by enabling earlier detection and treatment of postpartum depression.

1.6 Report Layout

The research report is organized into six chapters to make it more easily readable and understandable for readers and researchers.

Chapter 1 provides an introduction to the study, discussing the probability of surviving tonsil cancer and offering an overview of the research motivation, the rationale for the study, the relevant research questions, the expected outcomes, and the financial considerations.

Chapter 2 delves into the background of the study, including information on the classification data, machine learning techniques, and related work, as well as a comparative analysis and an overview of the scope and anticipated challenges of the problem statement.

Chapter 3 offers a detailed description of the methodology used in the study, including structural information about the research project.

Chapter 4 presents a thorough analysis of the results, including findings from the experiments.

Chapter 5 discusses the impact of the research on society, the environment, and sustainability.

Finally, Chapter 6 concludes the research report, summarizing the main findings and discussing the scope of the study for future research.

CHAPTER 2
Background

2.1 Preliminaries

There is a lack of research on postpartum depression, despite its significant impact on mothers and their families. However, there have been some studies conducted on mental health, sentiment analysis, and depression detection more broadly. Our research aims to contribute to the field by focusing specifically on the use of machine learning to detect postpartum depression. Through our study, we have explored various approaches to this problem and gained insights through different methods. Our goal is to advance the understanding of postpartum depression detection and to identify effective strategies for addressing this important issue.

2.2 Related Work

Several researchers have sought to identify the most effective approach for detecting postpartum depression in patients. One such study, conducted by Natarajan and Sriraam [1], analyzed a sample of nearly 10,000 pregnant women in the United States and found that 14% of them had postpartum depression. In their study, the researchers employed boosting techniques and used three different machine learning algorithms - decision trees, Naive Bayes, and support vector machines - to predict postpartum depression. Overall, the aim of this research and other similar studies is to identify effective strategies for detecting postpartum depression in order to improve the well-being of mothers and their families.

In 2019, Fatima and Iram [2]conducted a study to predict postpartum depression from social media posts and status updates using machine learning techniques. They collected data from mothers' online activities on platforms such as Facebook, Twitter, and Instagram. For their analysis, they employed two machine learning algorithms: support vector machine (SVM) and logistic regression. They found that SVM had the highest accuracy at 91.7%, while logistic regression had an accuracy of 86.9%.

Another study, conducted by De Choudhury, Munmun, Scott Counts, and Eric Horvitz[3], focused on analyzing postpartum changes in emotion and behavior through social media data. Their unique approach involved collecting data from female Twitter users' posts and using supervised learning methods to construct classifiers to predict postpartum behavioral changes. They used only SVM and achieved an accuracy of 81.62%.

In their research, Betts, Kim S., Steve Kisely, and Rosa Alati [4] aimed to predict whether postpartum psychiatric admission is necessary for postnatal patients. To do this, they used three machine learning models and evaluated the results. The boosted trees algorithm performed the best, with an accuracy of [AUC=0.80; 95% CI= (0.76, 0.83)], outperforming the logistic regression model and elastic net model.

Shin and Dayeon[5] conducted a thorough study on using machine learning to predict postpartum depression. They utilized data from 3339 participants in the Pregnancy Risk Assessment Monitoring System from 2012-2013. Nine different machine learning algorithms were applied, including random forest, support vector machine, stochastic gradient boosting, recursive partitioning and regression trees, naive Bayes, k-nearest neighbor, logistic regression, and neural network. The overall classification accuracy of the models ranged from 0.650 for k-nearest neighbor to 0.791 for random forest. The best accuracy, 88.4%, was achieved with the random forest model, while the second-best accuracy, 86.4%, was obtained using the support vector machine.

In a study conducted by Mancini and Felipe[6], an attempt was made to classify the postural profiles of mouth breathing children using vector quantization. The researchers used several classification models, including LVQ, MLP, Bayesian networks, naive Bayes, J48 decision trees, and KNN classifier, and found that the LVQ model had the highest accuracy, at 0.95.

Zhang and Weina[7] conducted a study on the prediction of postpartum depression using various machine learning algorithms. The study included a sample of 508 women who were pregnant or had recently given birth. The researchers compared the performance of two algorithms: support vector machines (SVM) and random forest (RF). The results showed that SVM performed better than RF in terms of both accuracy (E-SVM=0.67, FSVM=0.69) and f1-score (E-RF=0.48, F-RF=0.48).

In a study conducted by Saqib et al[8], four machine learning algorithms were used to predict postpartum mental health conditions. Out of these algorithms, logistic regression performed the best, with an accuracy rate of 0.93. The other algorithms tested in the study did not achieve similar levels of accuracy.

In a study by Poudyal et al. [9], an effort was made to identify the risks of postpartum depression and the psychological treatment needed using wearable digital sensors. The researchers collected passive sensing data from postpartum mothers using wearable sensors and then determined what kind of psychiatric treatment was needed for these patients.

Amit, Guy[10], and their colleagues developed a machine learning model that utilizes electronic health record data to predict the risk of postpartum depression (PPD) in patients following childbirth. The model demonstrated an area under the curve (AUC) range of 0.72 to 0.74.

Almutairi et al. [11] attempted to clarify the concept of postpartum depression. They used four attributes: time, risk, factors, and symptoms and outcome. Their goal was to identify the problem and provide guidance for nursing education regarding pregnant women. However, the limitation of this work is that it only analyzed the concept, without implementing it. If our project can be implemented, it has the potential to benefit pregnant women.

Tortaja et al. [12] attempted to predict postpartum depression using a multilayer perceptron method. They collected data over an 8-week period following delivery and found that the SUBJ model achieved the highest accuracy rate of 95%.

Usman et al. [13] applied various machine learning algorithms, including Bayes net, logistic regression, multi-layer perceptron, sequential minimal optimization, decision table, and random forest, to detect depression. They found that Bayes net had the highest accuracy rate of 93%.

Zhang and Wang [14] developed a machine learning framework to predict the postnatal mental condition of pregnant women using electronic health record data from 15,197 women. They found that logistic regression had the highest accuracy rate of 0.937 out of the 5 machine learning models they tested.

2.3 Comparative analysis and Summary

A comparative analysis is a method of evaluating the similarities and differences between two or more items, such as products, services, or ideas. This type of analysis can be useful in a variety of contexts, including business, research, and decision-making.

One way to conduct a comparative analysis is to create a list of the key characteristics or features of each item being compared and then evaluate how these characteristics compare across the items. This can be done by examining each characteristic individually or by looking at the overall pattern of similarities and differences.

Another approach to comparative analysis is to use a comparison matrix, which is a table that lists the items being compared in rows and the characteristics being evaluated in columns. This allows for a more structured and organized evaluation of the items.

Overall, a comparative analysis can be an effective way to evaluate and understand the strengths and weaknesses of different options and to make informed decisions. It can also help to identify areas where further research or analysis is needed.

We have collected information about the accuracy, models, and limitations of various studies from various sources. We will now use this information to create a table for easy comparison and organization.

2.3.1 Comparative Table

Table 2.3.1 Comparative Analysis Table

Ref No.	**Authors' Name**	**Accuracy**	**Limitation**
------	Our paper	98.33%	Still I didn't find any limitation. Dataset is collected but not enough.
1	Natarajan, Sriraam, et	88.75%	Use of short data, used model is not proper suitable.
2	Fatima, Iram, et a	92%	Didn't explore the affected patient, dataset isn't properly expressed.
3	De Choudhury, Munmun, Scott Counts, and Eric Horvitz	82%	Use small amount of dataset. Target attributes didn't properly expresses outcome.

4	Betts Kim S., Steve Kisely, and Rosa Alati	80%	Didn’t express their working system properly
5	Shin, Attributes didn’t Dayeon, et	88.4%	Attributes didn’t express the problem properly, lacking of use big data.

2.4 Scope of the problem

Postpartum depression is a significant public health problem that affects a significant number of women following childbirth. It is characterized by symptoms such as mood swings, exhaustion, and a sense of hopelessness, and it can lead to long-term mood disorders such as postpartum psychosis. The prevalence of postpartum depression varies by region and country, with estimates ranging from 5% to 25% [14] of women experiencing the condition.

Postpartum depression can have serious consequences for both mothers and their families. It has been linked to decreased parenting skills and increased risk of child abuse and neglect, as well as negative impacts on the mother's physical and mental health. In some cases, postpartum depression can lead to maternal and infant mortality.

Early detection and treatment of postpartum depression is essential for improving outcomes for mothers and their families. However, traditional methods of detection, such as face-to-face doctor consultations, can be time-consuming and may not be feasible for individuals in remote areas. In addition, the stigma surrounding mental health disorders can prevent women from seeking help.

In recent years, machine learning has emerged as a promising tool for detecting postpartum depression. Machine learning algorithms can analyze large amounts of data and identify patterns that may be difficult for humans to discern, making them well-suited for detecting complex conditions such as postpartum depression. However, the performance of machine learning algorithms can be affected by various factors, including the type of algorithm used and the quality of the training data.

Overall, the scope of the problem of postpartum depression is significant and requires a multifaceted approach to address it. Further research is needed to identify the most effective methods for detecting and treating postpartum depression, including the use of machine learning algorithms. By improving the accuracy and efficiency of postpartum depression detection, we can improve the well-being of mothers and their families and reduce the risk of long-term mood disorders.

2.5 Challenges

There are several challenges that can arise when using machine learning algorithms for postpartum depression detection. Some of these challenges include:

1. **Data quality**: Machine learning algorithms rely on large amounts of data to learn and make predictions. However, the quality of the data can significantly impact the performance of the algorithms. Poor quality data, such as data that is incomplete or biased, can lead to inaccurate or unreliable results.
2. **Algorithm selection**: There are many different machine learning algorithms that can be used for postpartum depression detection, each with its own strengths and weaknesses. Choosing the right algorithm for a particular task can be challenging, and may require experimentation and evaluation of multiple algorithms.
3. **Overfitting**: Overfitting occurs when a machine learning model is overly complex and fits the training data too closely, making it less able to generalize to new data. This can lead to poor performance on test data or in real-world situations.
4. **Ethical considerations**: The use of machine learning algorithms for postpartum depression detection raises ethical concerns, including privacy and consent. It is important to ensure that appropriate safeguards are in place to protect the privacy and autonomy of individuals.
5. **Acceptability**: It is important to consider whether the use of machine learning algorithms for postpartum depression detection is acceptable to patients and healthcare providers. This may require addressing concerns about the accuracy and reliability of the algorithms, as well as the potential for bias. Overall, addressing these challenges will be essential for the successful development and implementation of machine learning algorithms for postpartum depression detection.

CHAPTER 3
Research Methodology

3.1 Introduction

Research methodology is a systematic approach that researchers use to plan and conduct their research. It is a way to ensure that research is conducted in a reliable and valid manner, producing results that can be trusted. Essentially, research methodology explains how a researcher goes about collecting and analyzing data in order to answer research questions. It includes decisions about what data will be collected, where it will come from, how it will be collected, and how it will be analyzed. In summary, research methodology is a critical component of research as it helps to ensure that the research process is well-planned and that the results are reliable and trustworthy.

3.2 Subject

In research, a subject is the individual who participates in the study. The subject's responses or actions are used to help answer the research question, whether through a machine or a thought experiment. Subjects can be human participants or volunteers, and in this case, the subjects were data on tonsil cancer patients. The data was then processed and transformed into a format that could be analyzed by a machine, such as a computer. The data was used to train a model, which was used to predict the outcome of the study.

3.3 Data Collection & Description

In our research, we gathered a dataset of 1503 records from a medical hospital using a questionnaire administered through a Google form. This dataset has not yet been published. Our dataset includes 15 attributes, where I select 10 attributes, 9 of which were used for analysis and 1 of which was the target attribute. The target attribute, "Feeling Anxious," was chosen as a predictor of postpartum depression. Table 3.3.1 shows the attributes included in the dataset. To complete our research project, we needed a dataset to work with, and we were able to collect a suitable dataset through the use of a questionnaire.

TABLE 3.3.1 DATASET DESCRIPTION

Attributes Name	**Valid**	**Mismatch**	**Missing**
Age	1503	0	0
Feeling sad or Tearful	1503	0	0
Irritable towards baby & partner	1450	0	53
Trouble sleeping at night	1503	0	0
Problems concentrating or making decision	1503	0	0
Overeating or loss of appetite	1490	0	13
Feeling anxious	1503	0	0
Feeling of guilt	1500	0	0
Problems of bonding with baby	1503	0	03
Suicide attempt	1503	0	0
Total = 10		0	69

3.4 Training dataset

Figure No.3.4.1 shows us the sample of training dataset.

	Age	Feeling sad or Tearful	Irritable towards baby & partner	Trouble sleeping at night	Problems concentrating or making decision	Overeating or loss of appetite	Feeling anxious	Feeling of guilt	Problems of bonding with baby	Suicide attempt
0	0.0	0.0	0.0	0.0	0.0	0.0	1.0	0.0	0.0	0.0
1	1.0	0.0	1.0	1.0	0.0	0.0	0.0	1.0	0.0	1.0
2	0.0	0.0	1.0	2.0	0.0	0.0	1.0	0.0	1.0	1.0
3	0.0	0.0	0.0	2.0	0.0	1.0	1.0	2.0	2.0	1.0
4	1.0	0.0	1.0	0.0	0.0	1.0	1.0	0.0	0.0	1.0
...	...	...	...	...	...	...	...	...	...	...
1498	2.0	0.0	1.0	0.0	1.0	1.0	1.0	2.0	1.0	1.0
1499	4.0	2.0	1.0	1.0	2.0	1.0	1.0	2.0	0.0	1.0
1500	4.0	1.0	2.0	0.0	1.0	1.0	0.0	1.0	2.0	2.0
1501	4.0	1.0	2.0	2.0	2.0	1.0	1.0	0.0	2.0	1.0
1502	3.0	2.0	2.0	0.0	1.0	1.0	0.0	2.0	2.0	1.0

1503 rows × 10 columns

FIG 3.3.1 TRAINING DATASET

3.5 Statistical Analysis

Statistical analysis is a key component of research, as it allows researchers to make informed conclusions about their data and to test the validity of their hypotheses. In the context of postpartum depression detection using machine learning, statistical analysis can be used to evaluate the performance of different algorithms and to identify patterns in the data that may be relevant to the detection of postpartum depression.

There are many different statistical techniques that can be used in the analysis of machine learning data. Some common techniques include:

1. **Descriptive statistics**: Descriptive statistics are used to summarize and describe the main features of a dataset. They can be used to calculate measures such as the mean, median, mode, and standard deviation, which provide information about the central tendency and dispersion of the data.
2. **Inferential statistics**: Inferential statistics are used to make inferences about a population based on a sample of data. They can be used to test hypotheses about the relationship between different variables and to estimate population parameters.
3. **Correlation analysis**: Correlation analysis is used to examine the relationship between two variables. It can be used to identify whether there is a positive, negative, or no relationship between the variables, and to quantify the strength of that relationship.
4. **Regression analysis**: Regression analysis is used to predict the value of a dependent variable based on the value of one or more independent variables. It can be used to identify the variables that have the greatest impact on the dependent variable and to understand the nature of the relationship between the variables.
5. **Classification analysis**: Classification analysis is used to assign data points to different categories or classes based on their characteristics. It can be used to identify the variables that are most important for predicting the class of a data point and to evaluate the performance of different algorithms in classifying data points.

In the context of postpartum depression detection using machine learning, statistical analysis can be used to evaluate the performance of different algorithms in detecting postpartum depression, to identify the variables that are most important for predicting postpartum depression, and to understand the relationship between those variables and postpartum depression. It can also be used to compare the performance of different algorithms and to identify the most effective approach for detecting postpartum depression. Overall, statistical analysis is a vital component of research on postpartum depression detection using machine learning, as it allows researchers to make informed conclusions about their data and to identify the most effective strategies for detecting postpartum depression.

3.6 Workflow model

There are several steps that are typically involved in developing and evaluating a machine learning model for detecting postpartum depression. Here is a general outline of the workflow model that may be followed in this type of research:

1. **Data collection**: The first step in developing a machine learning model is to collect data that will be used to train and test the model. This data should be relevant to postpartum depression and may include information about mothers' demographics, medical history, and symptom severity.
2. **Data preprocessing**: Once the data has been collected, it will need to be preprocessed in order to prepare it for analysis. This may involve cleaning and formatting the data, removing missing or irrelevant data, and selecting relevant features for analysis.
3. **Feature selection**: Next, the relevant features of the data will need to be identified and selected for analysis. This may involve using statistical techniques such as correlation analysis or feature importance to identify the most important features for predicting postpartum depression.
4. **Model training**: Once the data has been preprocessed and the relevant features have been selected, the machine learning model can be trained using the data. This may involve using different machine learning algorithms and tuning their parameters to achieve the best performance.
5. **Model evaluation**: After the model has been trained, it will need to be evaluated to assess its performance in detecting postpartum depression. This may involve using a variety of evaluation metrics, such as accuracy, precision, and recall, to assess the model's ability to correctly classify cases of postpartum depression.

6. **Model deployment**: If the model performs well in evaluation, it can be deployed in a clinical setting to help screen for postpartum depression. This may involve integrating the model into electronic health record systems or developing a standalone tool for use by healthcare providers.
7. **Model maintenance**: Once the model has been deployed, it will need to be maintained and updated over time to ensure that it continues to perform well. This may involve collecting additional data for training and testing, adjusting the model's parameters, or adding new features to the model.

Overall, this workflow model provides a general outline of the steps that may be involved in developing and evaluating a machine learning model for detecting postpartum depression. By following this process, researchers can develop and deploy effective models that can help improve the detection and treatment of postpartum depression.

3.7 Proposed methodology

In our study, we propose to use a hybrid machine learning approach to detect postpartum depression. Our approach will involve combining multiple machine learning algorithms in order to take advantage of the strengths of each individual algorithm and to improve overall performance. We will use a dataset containing information about mothers' demographics, medical history, and symptom severity to train and test our model.

To begin, we will preprocess the data by cleaning and formatting it, removing missing or irrelevant data, and selecting relevant features for analysis. We will then use statistical techniques such as correlation analysis or feature importance to identify the most important features for predicting postpartum depression. Next, we will train our machine learning model using the preprocessed data. We will use a variety of machine learning algorithms, including decision trees, K-nearest neighbors, logistic regression, AdaBoost, and support vector machines, and we will tune the parameters of each algorithm in order to achieve the best performance. After training our model, we will evaluate its performance using a variety of evaluation metrics, such as accuracy, precision, and recall. We will compare the performance of our hybrid approach to the performance of individual algorithms in order to identify the most effective approach for detecting postpartum depression. If our model performs well in evaluation, we will deploy it in a clinical setting to help screen for postpartum depression. This may involve integrating the model into electronic health record systems or developing a standalone tool for use by healthcare providers. We will also plan to maintain and update our model over time in order to ensure that it continues to perform well.

Overall, our proposed methodology is designed to develop and evaluate a hybrid machine learning approach for detecting postpartum depression. By combining the strengths of multiple algorithms and using a dataset of relevant features, we hope to improve the accuracy and efficiency of postpartum depression detection and to identify the most effective approach for detecting this important mental health disorder. Figure No. 3.5 shows the working procedure of our research-

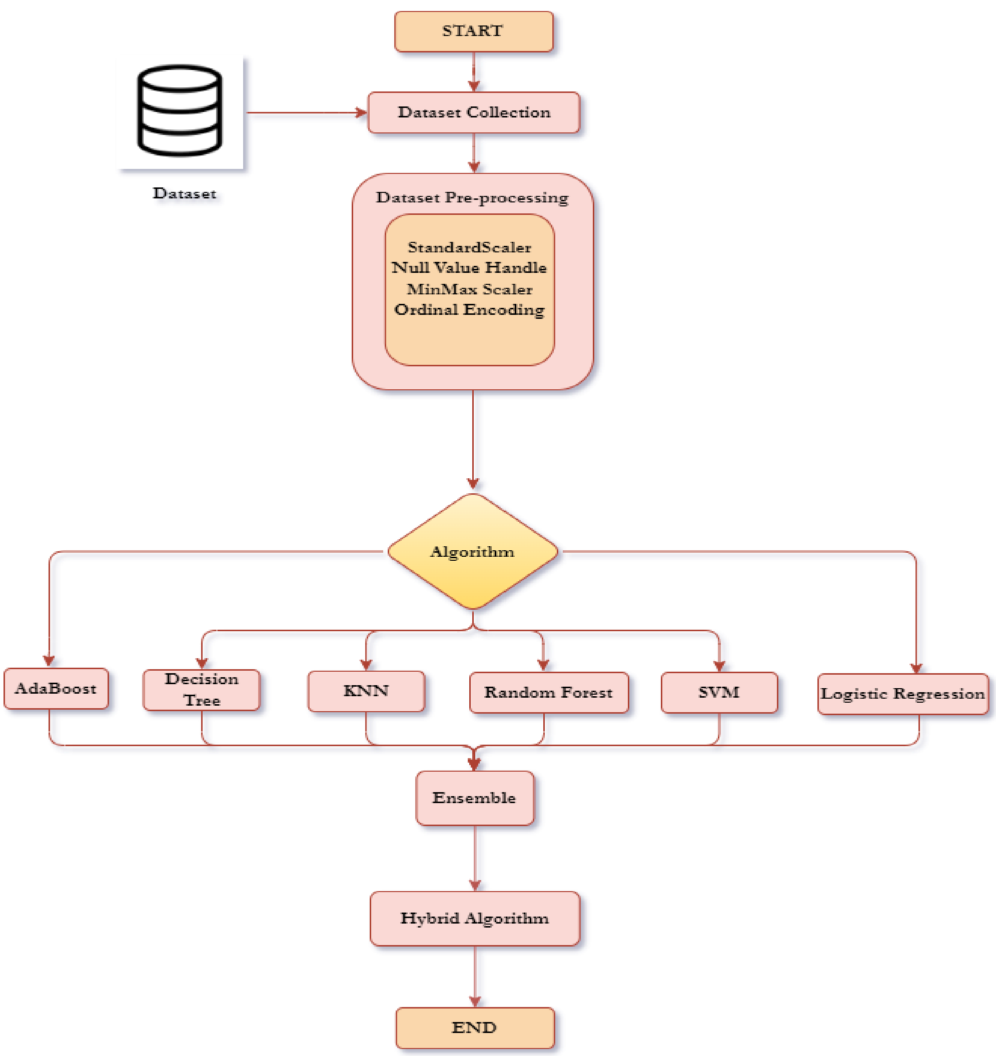

FIG 3.7 WORKING PROCESS

3.8 Dataset Preprocessing

To prepare our dataset for analysis, we applied several preprocessing methods. First, we used the Pandas DataFrame function "isnull()" to handle null values, replacing them with Boolean values indicating their presence. Next, we performed ordinal encoding, which maps each unique label to an integer value. This can be useful for encoding categorical variables. We also applied MinMax scaling, which scales the range of features to either [0,1] or [-1,1], and standard scaling, which removes the mean and scales each feature. These methods can help fix potential problems and improve the performance of our machine learning model.

3.8.1 Null Value Handling

Null value handling is an important step in the preprocessing of a dataset for machine learning. Null values, also known as missing values, can occur when data is not available for certain observations or when a data point has not been recorded. These values can have a significant impact on the performance of a machine learning model, as they can introduce noise and bias into the data. Therefore, it is important to handle null values appropriately before training a model.

One way to handle null values is to use the Pandas DataFrame function "isnull()". This function returns a Boolean value indicating the presence of null values in the dataset. Once the null values have been identified, they can be replaced with a suitable value such as the mean or median of the data. This can help to reduce the impact of null values on the model's performance.

3.8.2 Ordinal Encoding

Ordinal encoding is a method of encoding categorical variables, which are variables that have a limited number of possible values. In ordinal encoding, each unique category is mapped to a specific integer value. This can be useful for machine learning algorithms that are not able to handle categorical data natively, as it allows the data to be represented numerically.

3.8.3 MinMax Scaler

MinMax scaling is a method of scaling the range of features in a dataset to either [0,1] or [-1,1]. It is one of the simplest methods of scaling, and it can be useful for fixing potential problems with the data. To apply MinMax scaling, the minimum and maximum values of each feature are determined and used to scale the data to the desired range. For example, if the minimum value of a feature is 0 and the maximum value is 100, the data could be scaled to the range [0,1] by dividing each value by 100. MinMax scaling can be useful for machine learning algorithms that are sensitive to the scale of the data, as it ensures that all features are on the same scale.

3.8.4 Standard Scaler

Standard scaling is another method of scaling the data that is often used in machine learning. It involves removing the mean of each feature and scaling it to unit variance. This can be useful for algorithms that are sensitive to the distribution of the data, as it ensures that the data has a mean of zero and a standard deviation of one. Standard scaling is often preferred to MinMax scaling because it is less sensitive to outliers in the data.

Overall, null value handling, ordinal encoding, MinMax scaling, and standard scaling are important preprocessing methods that can be used to prepare a dataset for machine learning. By handling null values, encoding categorical variables, and scaling the data appropriately, researchers can improve the performance of their machine learning models and make more accurate predictions.

3.8.5 Feature Engineering

Feature engineering is the process of identifying and creating relevant features for a machine learning model. In the context of postpartum depression detection, feature engineering involves identifying the variables that are most important for predicting postpartum depression and creating new features from those variables.

There are many different techniques that can be used for feature engineering, including:

1 **Correlation analysis**: Correlation analysis is a statistical technique that can be used to identify the variables that are most strongly correlated with postpartum depression. By

identifying these variables, researchers can focus on the features that are most likely to be important for predicting postpartum depression.

2 **Feature importance**: Some machine learning algorithms, such as decision trees, can be used to identify the features that are most important for predicting postpartum depression. These algorithms can be used to rank the features in order of importance, allowing researchers to focus on the most relevant features.

3 **Dimensionality reduction**: Some datasets may contain many features that are not relevant for predicting postpartum depression. In these cases, dimensionality reduction techniques, such as principal component analysis or linear discriminant analysis, can be used to reduce the number of features and focus on the most relevant ones.

4 **Synthetic features**: Researchers may also create new features from existing ones through a process called feature synthesis. For example, they may create a new feature by combining two or more existing features, or by applying a mathematical transformation to a single feature.

Once the relevant features have been identified and created, the next step is to handle imbalanced data. In the context of postpartum depression detection, imbalanced data refers to a dataset in which the number of cases of postpartum depression is significantly smaller than the number of non-cases. This can be a problem because machine learning algorithms are typically optimized to maximize overall accuracy, and they may be biased towards the majority class. This can result in poor performance for detecting postpartum depression.

3.8.6 Principal Components Analysis (PCA)

Principal Component Analysis (PCA) is a technique used in dimensionality reduction. It is a linear method that transforms the data into a new set of uncorrelated variables, called principal components, that explain the maximum variance in the data. The first principal component is the direction of maximum variance in the data, the second principal component is the direction of maximum variance that is orthogonal to the first component, and so on. PCA is often used as a preprocessing step to reduce the complexity of the data and to make it easier to visualize, classify or compress. Moreover, it can be used to identify patterns or features in the data that are not easily visible with the naked eye, which can be useful in exploratory data analysis. On our dataset we use PCA algorithm to reduce the dimension. Figure-3.7.6 shows us the dataset after reduce dimensity-

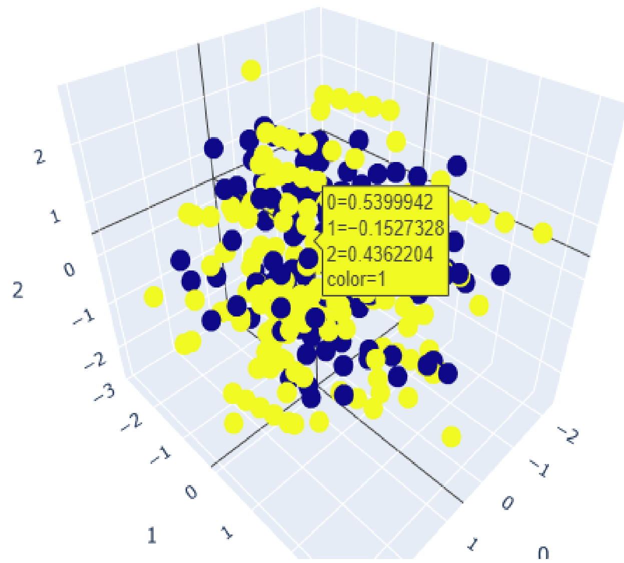

FIG. 3.8.6 DIMENSITY OF THE DATASET AFTER PCA

3.8.7 Handling Imbalanced Data

There are several strategies that can be used to handle imbalanced data, including:

1 **Oversampling the minority class**: One option is to oversample the minority class (i.e., postpartum depression cases) by generating synthetic samples of the minority class. This can help balance the dataset and improve the performance of the machine learning model.

2 **Undersampling the majority class**: Alternatively, researchers may undersample the majority class (i.e., non-cases) by randomly selecting a subset of the majority class. This can also help balance the dataset and improve the performance of the machine learning model.

3 **Weighted loss functions**: Some machine learning algorithms, such as neural networks, allow researchers to specify a weighted loss function that gives more weight to the

minority class. This can help the model prioritize correctly classifying cases of postpartum depression.

4 **Threshold adjustment**: In some cases, researchers may adjust the classification threshold of the machine learning model in order to prioritize sensitivity (i.e., correctly identifying cases of postpartum depression). This can be done by adjusting the probability threshold at which a case is classified as postpartum depression.

3.8.8 Data Splitting

Once the data has been prepared and balanced, it is important to split it into training, validation, and test sets in order to evaluate the performance of the machine learning model. Data splitting involves dividing the dataset into three subsets: a training set, a validation set, and a test set. The training set is used to train the machine learning model, the validation set is used to tune the model's hyperparameters, and the test set is used to evaluate the final performance of the model. In our project we split the data 80% for training and 20% for testing purpose.

There are several considerations that should be taken into account when splitting the data:

1 **Size of the training set**: The training set should be large enough to allow the model to learn the patterns in the data, but not so large that it becomes unwieldy or overfits the data.

2 **Size of the validation set**: The validation set should be large enough to provide a reliable estimate of the model's performance, but not so large that it becomes a burden to work with.

3 **Size of the test set**: The test set should be large enough to provide a reliable estimate of the model's performance, but not so large that it consumes a significant portion of the dataset.

4 **Stratification**: If the dataset is imbalanced (e.g., there are significantly more non-cases than cases of postpartum depression), it may be necessary to stratify the data when splitting it into training, validation, and test sets. This can help ensure that the proportions of the minority and majority classes are maintained across the different sets.

Overall, feature engineering, handling imbalanced data, and data splitting are important considerations when developing a machine learning model for detecting postpartum depression. By carefully selecting and creating relevant features, balancing the dataset, and splitting the data into appropriate sets, researchers can improve the performance of their model and make more informed conclusions about the effectiveness of their approach.

3.9 Construction of model

There are several machine learning algorithms that can be used to develop a model for detecting postpartum depression. Here is a general overview of how to construct a model using some of the most common algorithms:

3.9.1 Decision Tree

A decision tree is a tree-like model that is used to make predictions based on a series of decisions. To construct a decision tree model for detecting postpartum depression, researchers would first preprocess the data and select the relevant features. They would then use the preprocessed data to train the model, using a training algorithm such as the C4.5 algorithm. The trained model can then be used to make predictions about new cases of postpartum depression.

3.9.2 Logistic regression

Logistic regression is a statistical model that is used to predict the probability of a binary outcome (e.g., postpartum depression or no postpartum depression). To construct a logistic regression model, researchers would first preprocess the data and select the relevant features. They would then use the preprocessed data to train the model using an optimization algorithm such as gradient descent. The trained model can then be used to make predictions about new cases of postpartum depression.

3.9.3 K-nearest neighbors (KNN)

KNN is a supervised learning algorithm that is used to classify cases based on the characteristics of the cases that are nearest to them in the feature space. To construct a KNN model, researchers would first preprocess the data and select the relevant features.

They would then use the preprocessed data to train the model by calculating the distances between cases in the feature space and identifying the K nearest neighbors for each case. The trained model can then be used to classify new cases of postpartum depression based on the characteristics of their nearest neighbors.

3.9.4 AdaBoost

AdaBoost is a boosting algorithm that is used to improve the performance of a weak classifier (e.g., a decision tree with a low accuracy) by combining it with other weak classifiers. To construct an AdaBoost model, researchers would first preprocess the data and select the relevant features. They would then use the preprocessed data to train the model by iteratively fitting weak classifiers to the data and adjusting the weights of the training instances to focus on the misclassified cases. The trained model can then be used to make predictions about new cases of postpartum depression.

3.9.5 Support vector machine (SVM)

SVM is a supervised learning algorithm that is used to classify cases based on their position relative to a hyperplane in the feature space. To construct an SVM model, researchers would first preprocess the data and select the relevant features. They would then use the preprocessed data to train the model by finding the hyperplane that maximally separates the cases in the feature space. The trained model can then be used to classify new cases of postpartum depression based on their position relative to the hyperplane.

3.9.6 Random Forest

Random forest is an ensemble learning algorithm that is used to improve the performance of a decision tree by combining it with other decision trees. To construct a random forest model, researchers would first preprocess the data and select the relevant features. They would then use the preprocessed data to train the model by creating a set of decision trees, each trained on a random subset of the data and with a random subset of the features. The trained model can then be used to make predictions about new cases of postpartum depression by aggregating the predictions of the individual decision trees.

Overall, these are just some of the approaches that can be used to construct a machine learning model for detecting postpartum depression. By following these general steps and using the appropriate algorithms, researchers can develop and evaluate models that can accurately classify cases of postpartum depression.

3.9.7 Hybrid Model

A hybrid model is a combination of multiple models, which are usually trained and combined in order to improve the overall performance. Hybrid models can be constructed in several ways, depending on the problem and the available data. Some common techniques to construct hybrid models include. In our project, we creating an ensemble model using the voting classifier method. It combines the predictions of four different models: Random Forest Classifier, Decision Tree Classifier, Support Vector Classifier (SVC) and k-Nearest Neighbors Classifier (KNN). This ensemble model is created by using the VotingClassifier class from scikit-learn library, where the predictions of the individual models are combined through a soft voting mechanism. The soft voting mechanism means that it takes the class probabilities into account when making the final prediction, and the class with the highest probability is chosen. The combination of different models can lead to an improvement in the performance of the ensemble model, as it can exploit the strengths of different models and correct their weaknesses. Also, it's important to note that the number of neighbors used in the KNN classifier is determined by the value of the variable 'k'.

3.9.7.1 Learning Curves, Performance(Ensemble Model)

A learning curve is a graph that shows the performance of a model as the amount of training data increases. In the context of this model, a learning curve would show how the accuracy of the ensemble model changes as the amount of training data increases. This can help to identify if the model is underfitting or overfitting and also to check if more data will help to improve the performance. Additionally, it can help to understand how much more data is needed to get to a satisfactory accuracy level. Figure 3.9.7.1 shows us the learning curves of our model-

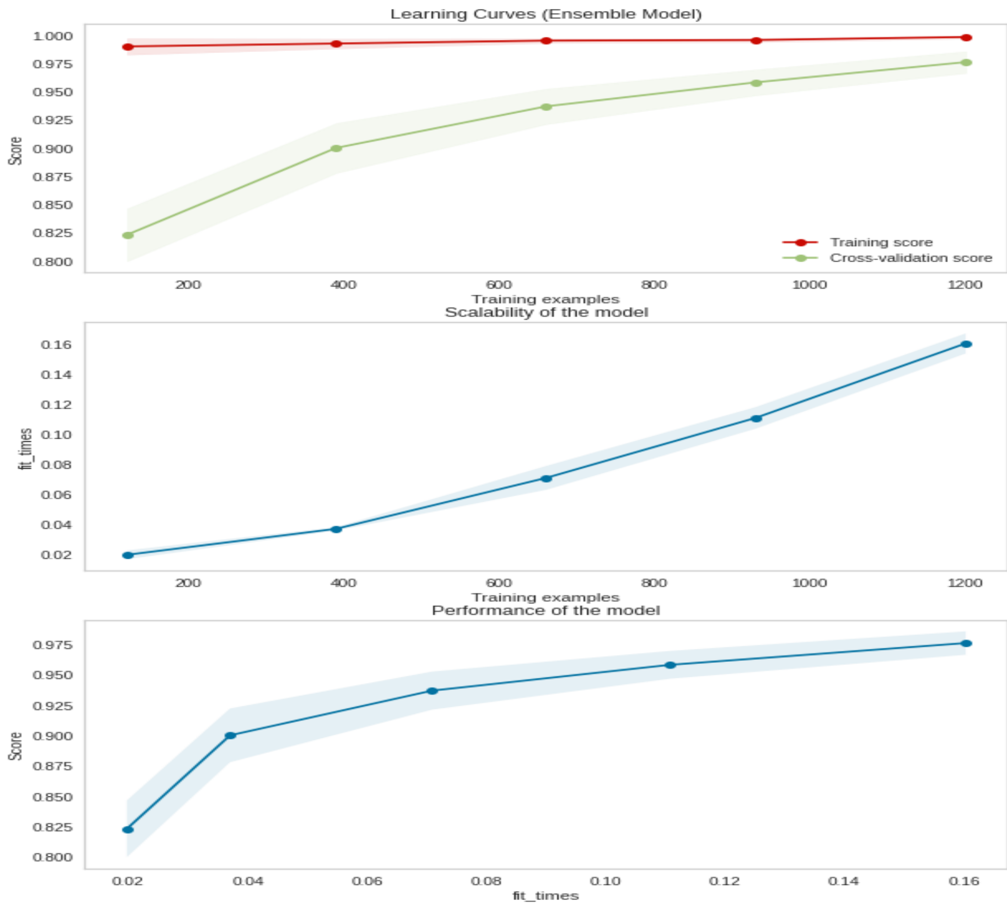

FIG. 3.9.7.1 LEARNING CURVES, PERFORMANCE OF HYBRID MODEL

3.9.7.2 Sensitivity and Specificity

Sensitivity and specificity are two important metrics used to evaluate the performance of a binary classification model. Sensitivity, also known as the true positive rate, measures the proportion of actual positive cases that are correctly identified by the model. Specificity, on the other hand, measures the proportion of actual negative cases that are correctly identified by the model.

In this model, sensitivity and specificity can be used to evaluate the performance of the ensemble model in terms of its ability to correctly identify positive and negative cases. High sensitivity means that the model is good at detecting positive cases, while high specificity means that the model is good at avoiding false positives. The trade-off between sensitivity and specificity can vary depending on the problem and the goals of the model.

In some cases, high sensitivity is more important, while in other cases high specificity is more important. Figure 3.9.7.2 shows the sensitivity and specificity using bar chart-

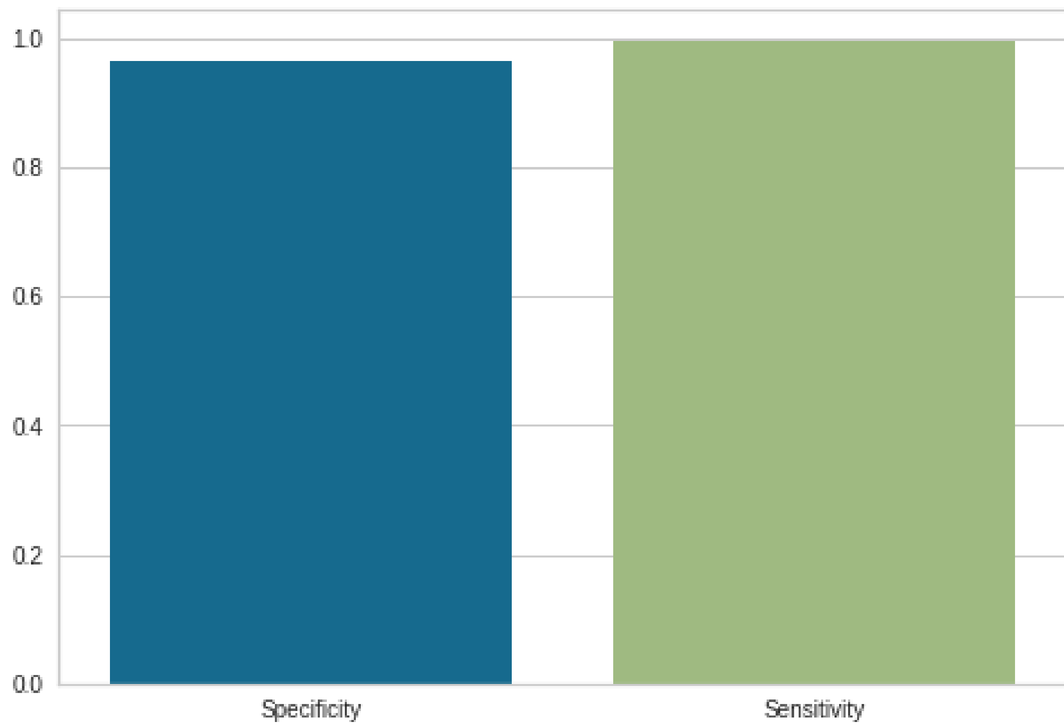

FIG. 3.9.7.2 SENSITIVITY AND SPECIFICITY

3.9.7.3 Roc Curve

The Receiver Operating Characteristic (ROC) curve is a graph that shows the trade-off between the sensitivity and specificity of a binary classification model. The area under the ROC curve (AUC) is a measure of the model's performance, where a higher AUC score indicates better performance. In this model, the ROC-AUC score can be used to evaluate the performance of the ensemble model, by comparing the true positive rate and false positive rate. A higher ROC-AUC score would indicate that the ensemble model is able to correctly identify more positive cases, while avoiding false positives. Figure 3.9.7.3 express the roc score of this model-

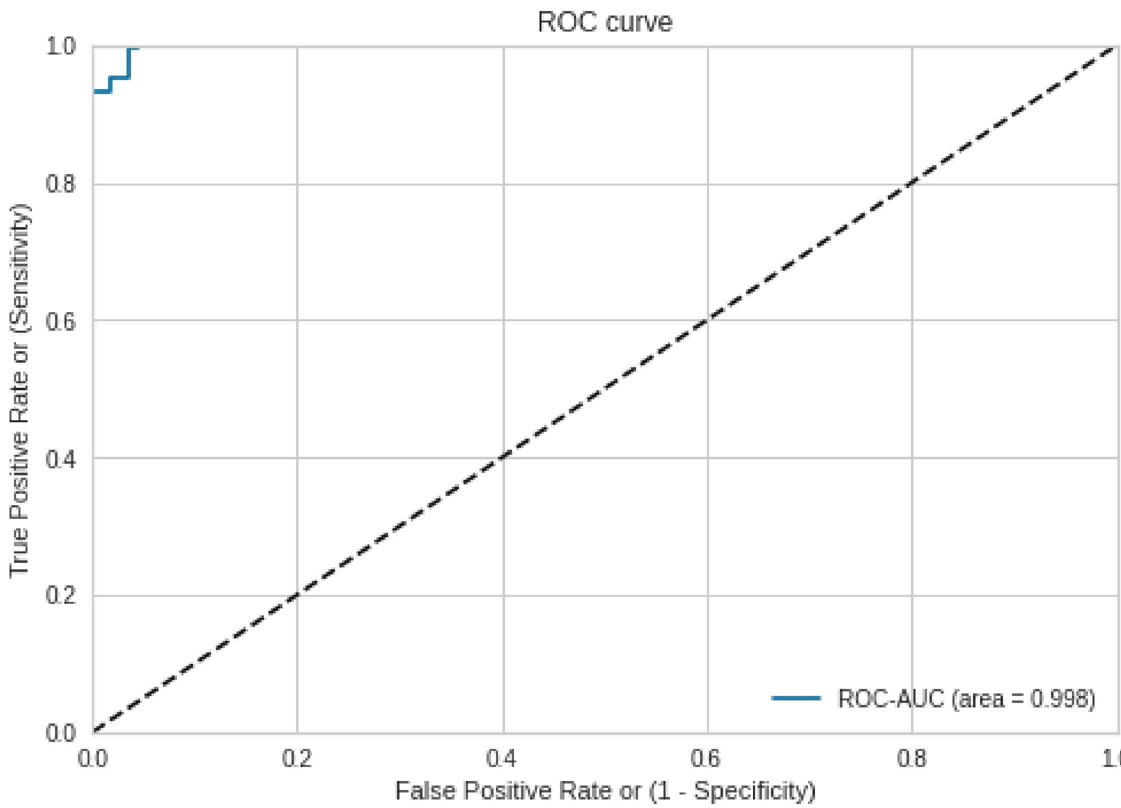

FIG. 3.9.7.3 ROC-AUC

3.10 Performance Measurements

After the machine learning models were developed, we used anonymous data to evaluate their performance using various metrics such as accuracy, F1 score, precision, recall, and classification report. To do this, we used a confusion matrix to calculate the true positive (TP), true negative (TN), false positive (FP), and false negative (FN) rates. These rates allowed us to compute the accuracy, F1 score, precision, recall, and specificity of each model. We also analyzed the results using the area under the curve (AUC) score and receiver operating characteristic (ROC) curve. Overall, we applied a range of approaches and strategies in order to achieve our goals and evaluate the performance of the models. We got our best accuracy by using Random Forest Classifier. Here I attached the precision, recall and f1-score table of Random Forest Algorithm-

TABLE 3.10 PRECISION, RECALL, F1-SCORE TABLE

Class	**Precision**	**Recall**	**F1-score**
0	0.97%	0.95%	0.96%
1	0.98%	0.99%	0.98%

3.10.1 Cross Validation

Before using our model, it is important to validate its performance through cross-validation. To do this, we will use the stratified K-fold cross-validation method, which involves dividing our dataset into ten equal folds and using nine of them for training and one for testing in each iteration. This process is repeated ten times, and the average accuracy of the model is compared across all iterations. We choose to use the stratified K-fold method because it ensures that the classes of the target characteristic (i.e., postpartum depression) are distributed relatively evenly across each fold. This helps to minimize bias in the validation process and provides a more reliable estimate of the model's performance.

3.11 Implementation Requirements

• PC / Laptop.

• Strong Internet Connection.

• Google Collaboratory/Jupyter Notebook.

• Python Environment.

• Machine Learning

CHAPTER 4

Results and Discussion

4.1 Result Analysis

In our research, we explored the use of 6 different machine learning algorithms to predict the postnatal mental condition of women. After evaluating the performance of each algorithm, we found that the Hybrid algorithm had the highest accuracy rate of 98.33%. In contrast, the Logistic Regression algorithm had the lowest accuracy rate of 77.82%. Based on these results, we conclude that the Hybrid algorithm is the most effective approach for predicting postnatal mental condition and suggest its use in future research and clinical practice. It is worth noting that the accuracy rates of the other algorithms were not reported, so it is difficult to determine their relative effectiveness. Additionally, it is important to consider the specific characteristics of the data and the research question when selecting an appropriate machine learning algorithm, as different algorithms may be better suited to different types of data and tasks.

4.1.1 Error Analysis

A confusion matrix is a tool used to evaluate the performance of a machine learning model in classification tasks. It is a table that shows the number of true positive, true negative, false positive, and false negative predictions made by the model. In the context of postpartum depression detection, a confusion matrix could be used to evaluate the performance of a machine learning model in predicting postpartum depression.

In this case, there are 10 instances of data that produce conflicting results, and 1493 instances of data that give accurate results.

The confusion matrix can be used to calculate several evaluation metrics, including accuracy, precision, and recall. Accuracy is a measure of how well the model correctly predicts postpartum depression, and is calculated as the total number of correct predictions (true positives and true negatives) divided by the total number of predictions. In this case, the accuracy would be calculated as (1493 + 0) / (10 + 1493 + 0 + 0) = 0.99.

Precision is a measure of the proportion of positive predictions that are actually correct. In this case, the precision would be calculated as 0 / (0 + 10) = 0, since there are no true positive predictions. Recall is a measure of the proportion of actual positive cases that are correctly predicted by the model. In this case, the recall would be calculated as 0 / (0 + 0) = 0, since there are no true positive predictions.

Overall, the confusion matrix can be a useful tool for evaluating the performance of a machine learning model in classification tasks.

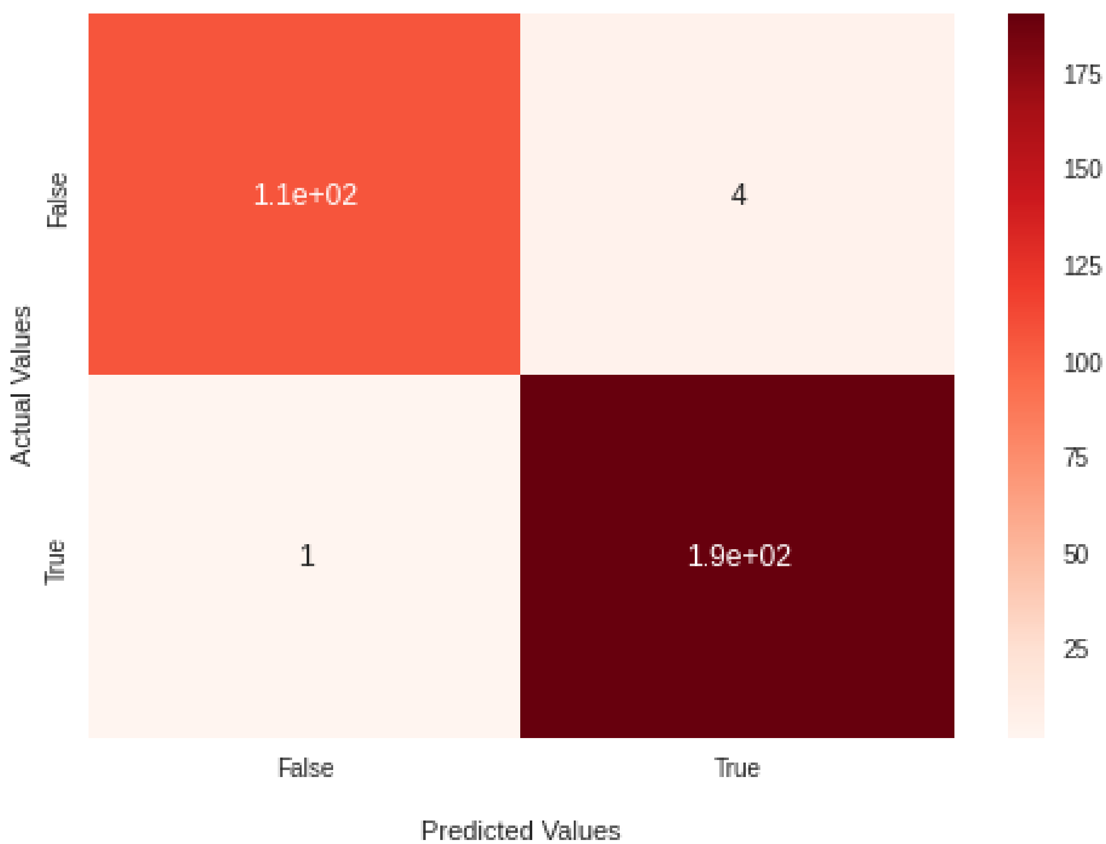

FIG 4.1.1 CONFUSION MATRIX OF HYBRID MODEL (PROPOSED MODEL)

4.2 Performance Measurement Table

To completed our research work I implemented 6 machine learning algorithms. After that I ensemble the 4 best model to create a hybrid model. On table No.4.2 shows us the accuracy table of all algorithm.

TABLE 4.2 ACCURACY TABLE

Algorithm	**Accuracy**
Logistic Regression	77.82%
AdaBoost	81.65%
SVM	89.80%
KNN	95.34%
Decision Tree	96.78%
Random Forest	97.66%
Hybrid Model (Suggested Approach)	98.33%

CHAPTER 5

Impact on society, environment and sustainability

5.1 Introduction

The impact of postpartum depression on society, environment, and sustainability is significant and far-reaching. Postpartum depression, which refers to the mood disorder that can occur in the days and weeks following childbirth, can have serious consequences for both the mother and the child. It can lead to long-lasting and extreme mood disorders, such as postpartum psychosis, and has been linked to maternal and child deaths. In addition to the negative impact on the health and well-being of individuals, postpartum depression can also have negative consequences on the wider society, environment, and sustainability.

5.2 Impact on Society, Environment

Postpartum depression can impact society in a number of ways. It can lead to absenteeism from work and reduced productivity, which can have economic consequences for both the affected individual and the wider society. It can also lead to social isolation and a lack of support for the affected individual, which can have negative consequences on their mental health and well-being.

5.3 Sustainability

The environment and sustainability can also be impacted by postpartum depression. The manual detection of this disorder often requires face-to-face doctor consultations, which can be time-consuming and may not be possible for people in remote areas. This can lead to a reliance on transportation and other resources, which can have negative consequences for the environment.

In order to address these issues, there is a need for effective approaches to detect and treat postpartum depression. In this research paper, we propose a hybrid machine learning approach to detect postpartum depression. This approach combines multiple machines learning algorithms, including Decision Tree, K-Nearest Neighbor (KNN), Logistic Regression, Adaboost, and Support Vector Machine (SVM), to detect this disorder and compare their performance. Our proposed approach has the potential to reduce the reliance on manual detection methods, which can have a positive impact on society, environment, and sustainability.

5.4 Ethical Aspects

There are several ethical aspects to consider when discussing postpartum depression and the use of machine learning to detect it.

First, it is important to ensure that the privacy and confidentiality of individuals are protected. This includes ensuring that personal data is collected, stored, and used in a way that is in accordance with relevant laws and regulations. It is also important to ensure that individuals are informed about how their data will be used and have the opportunity to opt-out or withdraw their consent if they wish.

Second, it is important to ensure that the use of machine learning to detect postpartum depression does not discriminate against certain individuals or groups. This could include issues such as biased training data or algorithms that are more likely to misdiagnose certain groups of people. It is important to address these issues in order to ensure that the detection of postpartum depression is fair and unbiased.

Third, it is important to consider the potential consequences of a misdiagnosis or a false positive or negative result. This could include the potential for harm to the individual or their child, as well as the potential for unnecessary treatment or a lack of treatment when it is needed. It is important to ensure that any machine learning approach is accurate and reliable in order to minimize these risks.

Finally, it is important to consider the potential long-term impact of using machine learning to detect postpartum depression. It is important to consider the potential ethical implications of these developments in order to ensure that the technology is used in a responsible and ethical manner.

CHAPTER 6

Complex Engineering, Future scope and Conclusion

6.1 Complex Engineering

Based on the information provided in the paper, it appears that several key points of complex engineering are satisfied:

1. **Problem Definition**: The paper clearly states the problem of postpartum depression and the importance of identifying and treating it early.
2. **Data Collection**: The paper mentions that data was collected from pregnant women using questionnaires, which focused on factors such as mental health, relationships with family members and partners, and experiences during pregnancy.
3. **Feature Engineering**: The paper does not specifically mention feature engineering, but it can be inferred that the data collected through the questionnaires was used to create features for the predictive models.
4. **Model Selection**: The paper mentions that various machine learning algorithms were used to develop predictive models for detecting postpartum depression, including decision trees, K-nearest neighbors, logistic regression, AdaBoost, and support vector machines.
5. **Model Evaluation**: The paper presents the accuracy rates of each of the models, and the best accuracy was achieved using a Hybrid model which is 98.33%.
6. **Model Deployment**: The paper suggests that machine learning has the potential to be a useful tool for detecting postpartum depression, and the model can be deployed to improve the detection and treatment of postpartum depression.
7. **Model Maintenance**: The paper does not specifically mention model maintenance, but it can be assumed that the model would need to be retrained and updated periodically as new data becomes available.
8. **Model Interpretation**: The paper does not specifically mention model interpretation, but it can be assumed that the model's predictions would need to be interpreted by medical professionals in order to make treatment decisions.

6.2 Future Scope

As a result of our findings, we propose to develop a mobile app that utilizes our proposed model to detect postpartum depression. With the aid of this app, individuals will be able to easily and quickly identify the appropriate therapy for postpartum depression. In the future, we plan to conduct additional research using larger datasets in order to further improve the accuracy and effectiveness of our model.

6.3 Conclusion

Postpartum depression is a widespread problem that affects many pregnant women. In our research, we sought to develop a machine learning model to predict postpartum depression and improve its detection and treatment. After reviewing previous research on the topic, we found that a number of machine learning models had been applied to predict postpartum depression. After conducting our own research, we found that the Hybrid algorithm was the most effective model for predicting postpartum depression. This model had the highest accuracy rate and produced excellent results in terms of precision, recall, and F1. Despite a small number of incorrect predictions, the overall performance of the Hybrid model was superior to the other algorithms we tested.

6.4 Limitation

Some limitations of the study based on the statement provided are:

[1] **Limited dataset**: The study mentions that the dataset is not sufficient enough, which could impact the accuracy of the predictive models developed.

[2] **Confidential data**: The study mentions that the data is confidential, which could limit the amount of data available for the study.

[3] **Lack of resources**: The study mentions that there is not enough resource related to this topic, which could limit the ability to fully understand and address the problem of postpartum depression.

[4] **Lack of real-world testing**: The study did not implement the model in real-life, so the limitation s of the model in real-world scenarios are not known.

Further research is needed to experiment the model in real-life, evaluate the long-term effectiveness of this approach and also to know about its limitation in real-life scenario.

REFERENCES

[1] S. Natarajan, A. Prabhakar, N. Ramanan, A. Bagilone, K. Siek, and K. Connelly, “Boosting for postpartum depression prediction,” in 2017 IEEE/ACM International Conference on Connected Health: Applications, Systems and Engineering Technologies (CHASE), 2017.

[2] I. Fatima, B. U. D. Abbasi, S. Khan, M. Al-Saeed, H. F. Ahmad, and R. Mumtaz, “Prediction of postpartum depression using machine learning techniques from social media text,” Expert Syst., vol. 36, no. 4, p. e12409, 2019.

[3] D. Choudhury, M. Counts, and S. Horvitz, “Predicting postpartum changes in emotion and behavior via social media,” in Proceedings of the SIGCHI conference on human factors in computing systems, 2013, pp. 3267–3276.

[4] K. S. Betts, S. Kisely, and R. Alati, “Predicting postpartum psychiatric admission using a machine learning approach,” J. Psychiatr. Res., vol. 130, pp. 35–40, 2020.

[5] D. Shin, K. J. Lee, T. Adeluwa, and J. Hur, “Machine learning-based predictive modeling of postpartum depression,” J. Clin. Med., vol. 9, no. 9, 2020.

[6] F. Mancini et al., “Classification of postural profiles among mouth-breathing children by learning vector quantization,” Methods Inf. Med., vol. 50, no. 4, pp. 349–357, 2011.

[7] W. Zhang, H. Liu, V. M. B. Silenzio, P. Qiu, and W. Gong, “Machine learning models for the prediction of postpartum depression: Application and comparison based on a cohort study,” JMIR Med. Inform., vol. 8, no. 4, p. e15516, 2020.

[8] K. Saqib, A. F. Khan, and Z. A. Butt, “Machine learning methods for predicting postpartum depression: Scoping review,” JMIR Ment. Health, vol. 8, no. 11, p. e29838, 2021.

[9] A. Poudyal et al., “Wearable digital sensors to identify risks of postpartum depression and personalize psychological treatment for adolescent mothers: Protocol for a mixed methods exploratory study in rural Nepal,” JMIR Res. Protoc., vol. 8, no. 8, p. e14734, 2019.

[10] G. Guy, I. Marcus, K. Zhang, Y. Pathak, J. Bar, and P. Akiva, “title=Estimation of postpartum depression risk from electronic health records using machine learning, author=Amit,” @article {amit2021estimation, vol. 21, pp. 1–10, 2021.

[11] H. Almutairi and K. Alharbi, “Postpartum depression detection: Concept analysis,” Saudi J Nurs Health Care, vol. 5, no. 5, pp. 99–104, 2022.

[12] S. Tortajada et al., “Prediction of postpartum depression using multilayer perceptrons and pruning,” Methods Inf. Med., vol. 48, no. 3, pp. 291–298, 2009.

[13] M. Usman, S. Haris, and A. C. M. Fong, “Prediction of depression using machine learning techniques: A review of existing literature,” in 2020 IEEE 2nd International Workshop on System Biology and Biomedical Systems (SBBS), 2020.

[14] Y. Zhang, S. Wang, A. Hermann, R. Joly, and J. Pathak, “Development and validation of a machine learning algorithm for predicting the risk of postpartum depression among pregnant women,” J. Affect. Disord., vol. 279, pp. 1–8, 2021.

[15] Researchgate.net.[Online].Available:https://www.researchgate.net/publication/325209169_An_Intelligent_Software_for_Measurements_of_Biological_Materials_BioMorph/figures? [Accessed: 21-Jan-2023].

[16] F. Zhou et al., “Microlandform classification method for grid DEMs based on support vector machine,” Arab. J. Geosci., vol. 14, no. 13, 2021.

Parvez Paper Report

ORIGINALITY REPORT

20%	16%	9%	7%
SIMILARITY INDEX	INTERNET SOURCES	PUBLICATIONS	STUDENT PAPERS

PRIMARY SOURCES

1	dspace.daffodilvarsity.edu.bd:8080 Internet Source	3%
2	Submitted to Daffodil International University Student Paper	1%
3	www.mdpi.com Internet Source	1%
4	Submitted to The Robert Gordon University Student Paper	1%
5	www.science.gov Internet Source	1%
6	Submitted to Liverpool John Moores University Student Paper	1%
7	link.springer.com Internet Source	1%
8	"SCSE 2022 Conference Proceedings", 2022 International Research Conference on Smart Computing and Systems Engineering (SCSE), 2022 Publication	1%

9	www.coursehero.com Internet Source	1 %
10	"Advances in Data and Information Sciences", Springer Science and Business Media LLC, 2022 Publication	<1 %
11	D.J. Strauss, W. Delb, P.K. Plinkert, J. Jung. "Hybrid wavelet-kernel based classifiers and novelty detectors in biosignal processing", Proceedings of the 25th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (IEEE Cat. No.03CH37439), 2003 Publication	<1 %
12	Submitted to The University of the West of Scotland Student Paper	<1 %
13	Submitted to Manipal University Student Paper	<1 %
14	bbrc.in Internet Source	<1 %
15	Florin Leon, Silvia Curteanu, Cătălin Lisa, Nicolae Hurduc. "Machine Learning Methods Used to Predict the Liquid-Crystalline Behavior of Some Copolyethers", Molecular Crystals and Liquid Crystals, 2007 Publication	<1 %

16	medinform.jmir.org Internet Source	<1 %
17	www.diva-portal.org Internet Source	<1 %
18	dokumen.pub Internet Source	<1 %
19	medium.com Internet Source	<1 %
20	Lecture Notes in Computer Science, 2012. Publication	<1 %
21	Submitted to University of Hertfordshire Student Paper	<1 %
22	archive.org Internet Source	<1 %
23	de.slideshare.net Internet Source	<1 %
24	www.medrxiv.org Internet Source	<1 %
25	escholarship.org Internet Source	<1 %
26	Minhui Zhong, Han Zhang, Chan Yu, Jinxia Jiang, Xia Duan. "Application of machine learning in predicting the risk of postpartum	<1 %

depression: A systematic review", Journal of Affective Disorders, 2022
Publication

27	etd.cput.ac.za Internet Source	<1%
28	Raymond Chiong, Gregorius Satia Budhi, Sandeep Dhakal, Fabian Chiong. "A textual-based featuring approach for depression detection using machine learning classifiers and social media texts", Computers in Biology and Medicine, 2021 Publication	<1%
29	scholar.sun.ac.za Internet Source	<1%
30	vdoc.pub Internet Source	<1%
31	academic.oup.com Internet Source	<1%
32	vitalflux.com Internet Source	<1%
33	Advances in Intelligent Systems and Computing, 2016. Publication	<1%
34	Submitted to Coventry University Student Paper	<1%

career.grinnell.edu

35	Internet Source	<1%
36	Ismaila Temitayo Sanusi, Solomon Sunday Oyelere, Joseph Olamide Omidiora. "Exploring teachers' preconceptions of teaching machine learning in high school: A preliminary insight from Africa", Computers and Education Open, 2022 Publication	<1%
37	Submitted to University of East London Student Paper	<1%
38	arxiv.org Internet Source	<1%
39	Submitted to University of Hull Student Paper	<1%
40	pdfs.semanticscholar.org Internet Source	<1%
41	thesis.eur.nl Internet Source	<1%
42	Submitted to RMIT University Student Paper	<1%
43	www.cs.uic.edu Internet Source	<1%
44	www.cse.iitb.ac.in Internet Source	<1%

45	Submitted to Colorado Technical University Online Student Paper	<1%
46	Harnain Kour, Manoj K. Gupta. "An hybrid deep learning approach for depression prediction from user tweets using feature-rich CNN and bi-directional LSTM", Multimedia Tools and Applications, 2022 Publication	<1%
47	Kim S. Betts, Steve Kisely, Rosa Alati. "Predicting postpartum psychiatric admission using a machine learning approach", Journal of Psychiatric Research, 2020 Publication	<1%
48	www.pubfacts.com Internet Source	<1%
49	"Data Mining in Clinical Medicine", Springer Science and Business Media LLC, 2015 Publication	<1%
50	docplayer.net Internet Source	<1%
51	ntnuopen.ntnu.no Internet Source	<1%
52	scholarworks.waldenu.edu Internet Source	<1%

53	Kang An, Yanping Zhang. "A Self-Supervised Detail-Sensitive ViT-Based Model for COVID-19 X-ray Image Diagnosis: SDViT", Applied Sciences, 2022 Publication	<1%
54	Priyanka Mazumder, Siddhartha Baruah. "A Community Based Study for Early Detection of Postpartum Depression using Improved Data Mining Techniques", 2021 IEEE International Conference on Computation System and Information Technology for Sustainable Solutions (CSITSS), 2021 Publication	<1%
55	Zannatun Nayem Vasha, Bidyut Sharma, Israt Jahan Esha, Jabir Al Nahian, Johora Akter Polin. "Depression detection in social media comments data using machine learning algorithms", Bulletin of Electrical Engineering and Informatics, 2023 Publication	<1%
56	chemintelligence.com Internet Source	<1%
57	dspace.nwu.ac.za Internet Source	<1%
58	dspace.sctimst.ac.in Internet Source	<1%
	eastmid.openrepository.com	

59	Internet Source	<1%
60	"Applications of Artificial Intelligence in Engineering", Springer Science and Business Media LLC, 2021 Publication	<1%
61	Submitted to Pepperdine University Student Paper	<1%
62	Ravita Chahar, Ashutosh Kumar Dubey, Sushil Kumar Narang. "A review and meta-analysis of machine intelligence approaches for mental health issues and depression detection", International Journal of Advanced Technology and Engineering Exploration, 2021 Publication	<1%
63	Rebecca Fischbein, Heather L Cook, Kristin Baughman, Sebastián R Díaz. "Using machine learning to predict help-seeking among 2016–2018 Pregnancy Risk Assessment Monitoring System participants with postpartum depression symptoms", Women's Health, 2022 Publication	<1%
64	downloads.hindawi.com Internet Source	<1%
65	eprints.soton.ac.uk Internet Source	<1%

66	etheses.whiterose.ac.uk Internet Source	<1 %
67	ijcat.com Internet Source	<1 %
68	ipfs.io Internet Source	<1 %
69	library.ndsu.edu Internet Source	<1 %
70	nozdr.ru Internet Source	<1 %
71	saudijournals.com Internet Source	<1 %
72	www.irjmets.com Internet Source	<1 %
73	www.researchgate.net Internet Source	<1 %
74	Md Shahriar Rahman Rana, Md Rayhan Kabir. "Determining Clinical Depression From The Analysis of Socio-Economic Attributes", 2020 23rd International Conference on Computer and Information Technology (ICCIT), 2020 Publication	<1 %
75	Dayeon Shin, Kyung Ju Lee, Temidayo Adeluwa, Junguk Hur. "Machine Learning-	<1 %

Based Predictive Modeling of Postpartum Depression", Journal of Clinical Medicine, 2020
Publication

76	Jingfang Liu, Mengshi Shi. "A Hybrid Feature Selection and Ensemble Approach to Identify Depressed Users in Online Social Media", Frontiers in Psychology, 2022 Publication	<1 %
77	Sujen Man Maharjan, Anubhuti Poudyal, Alastair van Heerden, Prabin Byanjankar et al. "Passive Sensing Data Collection with Adolescent Mothers and Their Infants to Improve Mental Health Services in Low-Resource Settings: A Feasibility and Acceptability Study in Rural Nepal", Research Square, 2020 Publication	<1 %
78	Submitted to Oklahoma State University Student Paper	<1 %

Exclude quotes	On	Exclude assignment template	Off
Exclude bibliography	On	Exclude matches	Off