*International Journal of Computational and Experimental Science and ENgineering (IJCESEN)*

*Vol. 11-No.1 (2025) pp. 475-483*
*http://www.ijcesen.com*

***ISSN: 2149-9144***



**Research Article**

# Prediction Of Postpartum Depression With Dataset UsingHybrid Data Mining Classification Technique

**Arya P. Pillai[1]*, N.V. Chinnasamy[2]**

[1]Research Scholar, Karpagam Academy of Higher Education Coimbatore
* **Corresponding Author Email:** aryappillai26@gmail.com- **ORCID:** 0000-0002-5247-78XX

[2]Assistant Professor, Karpagam Academy of Higher Education Coimbatore
**Email:** chinnasamy.nvaiyapuri@kahedu.edu.in - **ORCID:** 0000-0002-5247-78XY

**Article Info:**

**DOI:** 10.22399/ijcesen.750
**Received :** 11 October 2024
**Accepted :** 19 December 2024

**Keywords :**

PPD,
DM,
SVM,
ANN,
Hybrid Classifier.

**Abstract:**

Postpartum Depression is a condition or a state which usually affects the woman immediately after child birth. The birth of a baby not only brings delighted emotions such as excitement, but also fear and anxiety which may sometimes lead to depression. It is a period of physical, emotional and behavioral changes that happen in some woman immediately after the delivery. Apart from the chemical changes, there are many factors which affect a woman during and after pregnancy period. If PPD is not identified and treated at the earlier stages, it may lead to serious issues for mother and child. It is therefore of vital importance to sift through the woman at any early stage to prevent any consequences. The objective of this study is to find out the presence of PPD without getting worse. Data mining plays an important role in the health care industry with successful outcome. It helps to find out hidden patterns, trends and anomalies from large dataset to make the predictions. The proposed system is a combined classification technique for the prediction of postpartum depression that uses Support vector machine, Artificial Neural Network and Hybrid classifier algorithm to produce the best result.

## 1. Introduction

Data Mining is the investigation of substantial information that turn up intomeaningful facts. The aim of data mining in health care and personalized care is to identify useful and comprehensible patterns by analyzing large set of data. Data mining algorithms play an important role in diagnostics, treatment and healthcare. It helps the healthcare professionals to identify patterns from alarge amount of data. These data patterns help to predict meaningful information which improve patient quality of life and can bring the disease under control. Several approaches are available in data mining for the forecast of diseases. Recent investigation states that healthcare industry is reliable for producing about 30 % of all global data and by 2025,it will reach 36 %.

Postpartum depression is a complex situation which is often diagnosed inwoman after the birth of a child. It is a serious mental health issue which unfavorably affects the mother with suicide attempts which may leads to approximately 20 % of postpartum deaths. [1]. A woman goes through lots of physical and hormonal changes during pregnancy. The massive changes immediately after delivery can bring serious emotional changes in woman.

This emotional change which is usuallycalled baby blues in woman tends to decrease over the first 2 weeks after delivery. But PPD prone to be longer and severely affects the mother child relationship as well as the mental stability [2]. Symptoms include insomnia, loss of appetite, irritable or angry with their husband or partner, difficulty in bonding with the baby. Undiagnosed and untreated PPD may lead the mother into a depressionstate which included problems such as violent behavior, psychiatric and medical disorder, behavioral reticence [3].Untreated PPD not only creates an unfavorable environment not only for the mental development of mother but also to the overall development of the child [4]. This paper is an attempt to focus on the prediction of PPD in women based on the algorithm such as SVM, ANN and Hybrid classifier algorithm.

In recent days, health care industry reposes on data mining for effective analysis and decision making. In this research article, we use Python which is a powerful data analysis and manipulation

programming language, which provides various libraries and functions for organizing and managing data.

Section 2 explains about the information and strategies used. In Section III, data mining techniques and evaluation are described. Section IV concentrates on Results and Discussions. Finally, Section V deals with the conclusion part of the research article.

**Related Works**

In 2019, Iram Fathima and BushanUd Din Abbasi introduced generalized approach which can extract linguistic features from user-textual post on social media and based on these women can be categorized as general, depressive. Support Vector Machine algorithm is used in this studyand achieved 86.9 % accuracy for PPDcontent prediction [5-9.]

In 2022 Trine, Munk-Oslen and Kathrine Bang Madsen suggested a model to evaluate the possibility of PPD. For the study they use three prediction models (Core or basic model, Extended and modelExtended+ model) in a huge data set.Core models uses predictor variables for identifying PPD risk factors. Extended andExtended + models were best for Discrimination and Calibrations. In the conclusion the study recommends extended+ model for continued development in time ahead work [10].

In 2020, EldarHochman, Becca Feldman published a paper "A postpartum depression model based on the development and validation of machine learning:A nationwide cohort study". In this a nationwide cohort including 214,359data is taken for model training. Gradient- boosted decision tree algorithm was applied to EHR data. The model's predictors included well-recognized and less-recognized PPD risk factors [11].

In September 2020, Dayeon Shin and Kyung Ju Lee published a paper "Machine Learning-based Predictive Modeling of Postpartum Depression". This paper describes the impact of Life stress and history of depression in the occurrence of Postpartum depression [12]. Moreover, this paper uses Random Forest(RF) Gradient Boosting Machine(GBM).Based on the study the paper concludes that Random Forest (RF) attain the best execution for predicting postpartum depression with classification accuracy of value 0.91 and an AUC value of 0.894.

In January 2021, Yiye Zhang and Shuojia Wang published a paper "Development and Validation of a machine learning algorithm for predicting the risk of Postpartum depression among women". In this paper a machine learning framework issuggested for prediction of PPD using datafrom electronic health records. The frame was implemented using multilayer perceptron (MLP) and evaluated data collected from regular intervals during pregnancy. The study demonstrates a data- driven, scalable decision support systemfor the prediction of PPD [13].

In October 2020, Xiago M and Yan C published a paper "Risk Prediction for postpartum depression based on Random Forest." In this work, Random Forest algorithm is applied on the data set consisting of 406 participants provided an accuracy and sensitivity of 80.10 % and 61.40 % respectively. The study concludedthat Random Forest has a great advantage in predicting the PPD [14].

## 2. Material and Methods

### 2.1 Data Source

In order to make an accurate prediction of PPD in woman, a real data set has been used. The study was conducted and questions were prepared based on the suggestions by Gynecologists, Psychologist and Medical Counselors practicing in various hospitals in Ernakulam district, Kerala.

Two different methods are used for data collection:

a) By preparing Questionnaire
b) By preparing an e-form.

In the first method, a questionnaire is prepared by listing the questions and it is distributed in the nearby Health centers and hospitals in and around Perumbavoor municipality in Ernakulam district. Data is manually collected from the women who completed at least one delivery. In the second method an electronic-form was prepared to input the entries from the women who had completed at least one delivery. Using these methods data set is collected which consist of 233 instances and 10 attributes, including numerical values and non-numeric values such as YES/NO/MAY BE. The entries areconverted in to a range of values between 0 and 4 for the ease of assessing attribute value. Table 1 describes the set of attributes that is used for predicting thePostpartum depression.

### 2.2 Data Partitioning

From the 233 record of data, 70 % of PPD data in the data set are used for the trainingpurpose, and the remaining 30 % are selected for testing purpose as shown in the Table:2. Data set can be partitioned into training data and testing data based onthe model which provides various patterns and model accuracy can be evaluated effectively.

Attributes play an important role model in understanding and analyzing the data set as it acts as predictors in making decisions. Attribute representation along with the percentage of occurrence is shown in Figure: 1.

***Table 1.*** *Data attributes with description*

| **Attributes** | **Description** | **Information attribute** |
|---|---|---|
| Age | Age | Numeric |
| Feeling sad or Tearful | Whether the respondent feels sad during the period. | Non-numeric |
| Irritable towards baby & partner | Whether the respondent feels irritation towards baby and partner during the period. | Non-numeric |
| Trouble sleeping at night | Respondent was not able to sleep at night (in days or weeks). | Numeric |
| Problems concentrating or making decision | Respondent may not able to take decision in any matter. | Non-numeric |
| Overeating or loss of appetite | Respondent may feel exhausted or loosing energy. | Non-numeric |
| Feeling anxious | Whether the respondent may feel anxious in most of the situations. | Non-numeric |
| Feeling of guilt | Whether the respondent may feel guilty in looking after the baby, family etc. | Non-numeric |
| Problems of bonding with baby | Respondent may not feel mental attachment with baby. | Non-numeric |
| Suicide attempt | Whether respondent may try to suicide in any situation. | Non-numeric |

***Table 2.*** *Percentage of Training and testing datarepresentation*

| 70 % of Training data | 30 % of Testdata |
|---|---|

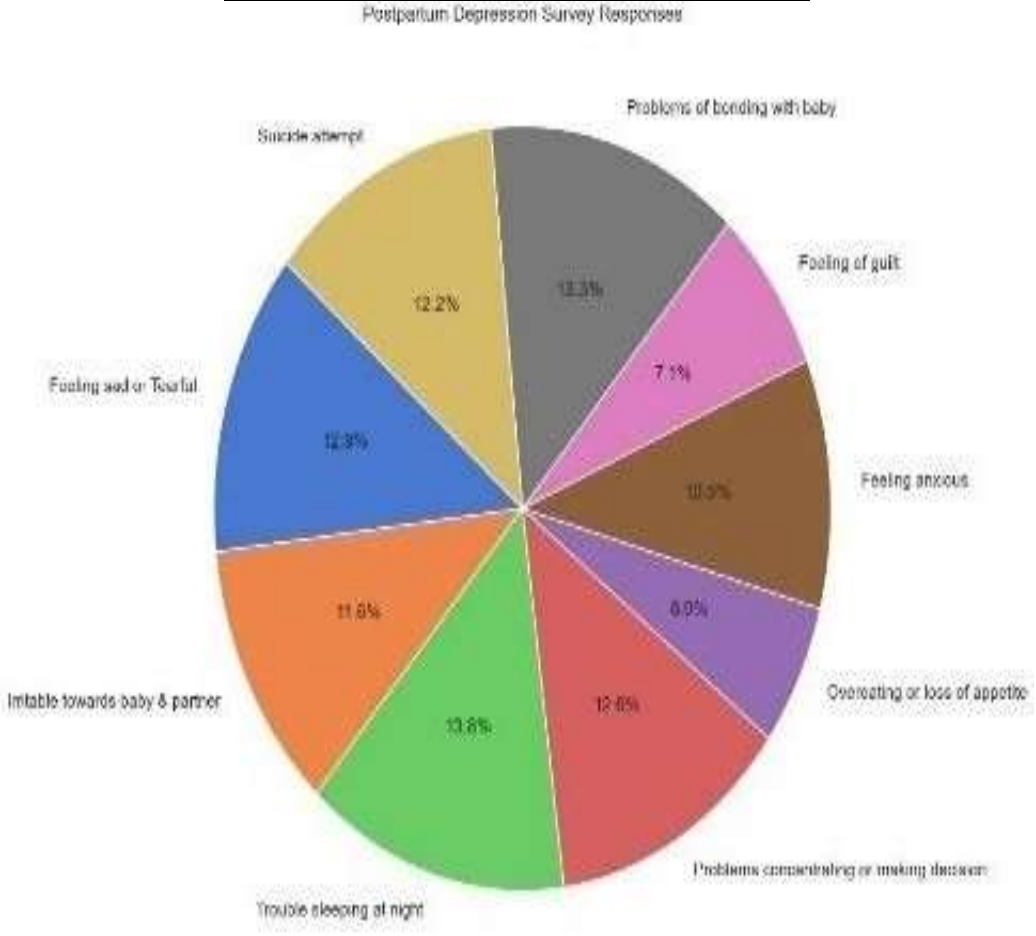


***Figure 1.*** *Data attributes with percentage of occurrence*

### 2.3 Hybrid Classifier and Performance Parameters

Several researches are already crossed through data mining techniques to envisionthe chances of PPD in women.

Dayeon Shin [5] conducted a prediction on PPD using a data set of 3339 and concluded thatSVM achieved highest performance in predicting PPD.Aris Supriyanto [6]proposed c4.5 algorithm for the prediction of PPD. Weina Zhang proposed that SVM algorithm in suitable for Predicting PPD [7].

**Support Vector Machine (SVM)**

Support Vector Machine (SVM) is a leading technology that works on both linear and nonlinear data [8]. SVM converts the actual data in to other dimension using mapping.

Within these high dimensions creates a decision boundary which is usually called separating hyperplane that separates the data from two classes. SVM always find out the points for composing hyperplane, where these points are calledvectors or support vectors.

SVM is used for numeric predictions in various areas including health care and clinical data. A partitioning hyperplane can be written as

$$A.xi + c = 0 \quad (1)$$

In equation 1, w is a weight vector, namely, A={a1,a2,a3,……an};n is the number of attributes ,say 10 and c is scalar which is the difference between the model's expected output and the true valueand x is the input vector. Class label is denoted in equation 2.

$$yi£\{-1,+1\} \quad (2)$$

Data xi belongs to negative class (-1) whenit satisfies the equation 3

$$A.xi + c \leq 1 \quad (3)$$

and xi belongs to positive class(+1) whenit satisfies the equation 4.

$$A.xi + c \geq 1 \quad (4)$$

The optimal boundary is obtained byconsidering the range between hyperplane and observation nearest to the hyperplane, d.

This distance is substituted in the equation 5.

$$A.( x1 - x2) = 2$$

$$||A|| * d = 2$$

$$\therefore d = \frac{2}{||W||} \quad (5)$$

Using the python programming language, while it provides various powerful data analysis and manipulation tool such as pandas, matplotlib, seaborn. The classifier uses a linear kernel, which can be modified using the 'svc_kernel' parameter. Enabled with the probability Estimation,' probability=True' parameter, allowing the SVC to provide class probabilities.

**Artificial Neural Network (ANN)**

Artificial Neural Network (ANN) ,which isa non linear predictive model consist of a large set of historical data and these data is trained or analyzed in order to predict the output of a future situation.ANN whichconsists of three layers: input layer, hiddenlayer and output layer resembles the hierarchical architecture of various units orneurons in human brain.

Using the python library for constructing ANN classifier, three layers are constructed; where input layer is the number of input features in the data set implicitly determined by the training data. Through the input layer, PPD related attributes are passed in to the network and multiplied by their respective weights. Hidden layer is the one layer with 20 neurons, using the Rectified Linear Unit activation function and it add a bias (bi) with the weighted sum and output layer is the one neuron with the sigmoid activationfunction suitably for binary classification tasks. Figure 2 represents the three layers;input,hidden and output layer in the ANN algorithm.

**Hybrid Classifier**

In PPD prediction, hybrid classifier provides a reasonable accuracy compared to other existing techniques. The hybrid classifier incorporates the two classifiers SVM and ANN by determining the predicts values; and these values is related with the original value in the table and produces the best hybrid classification. It combines the predictions from the SVM and ANN using a voting classifier. Voting classifier algorithm is used for making predictions based on the highest likelihood of selected class as output. The PPD prediction system architecture is shown in figure 3.

**Performance Evaluation Metrics**

For Performing the PPD classification evaluation, metrics such as Accuracy, Precision, Recall, Specificity and f1-score is calculated by using the formulas given in the Table: 3.

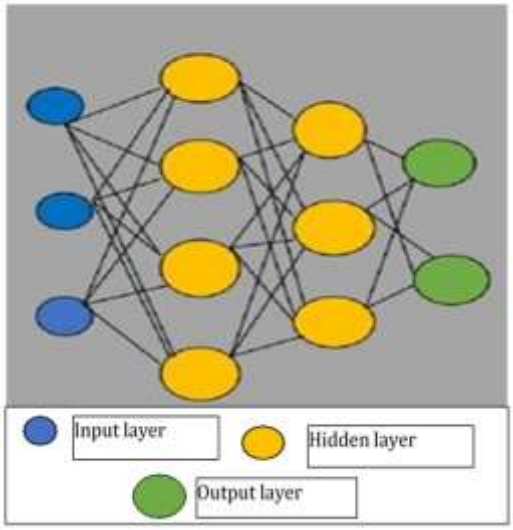


***Figure 2.*** *ANN representation diagram*

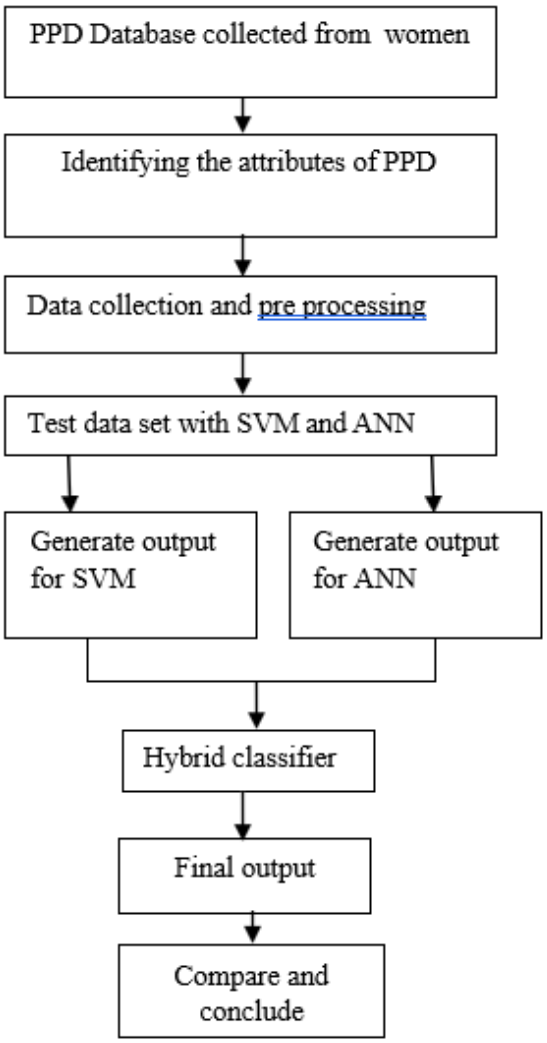


***Figure 3.*** *PPD prediction architecture*

A confusion matrix with 4 variables, True Positives (tpv), True Negatives (tnv), FalsePositives (fpv) and False Negatives (fnv) describes the performance of classificationalgorithms.

***Table 3.*** *Performance metrics with formula*

| | |
|---|---|
| Accuracy | $\frac{tpv + tnv}{tpv + fnv + tnv + fpv}$ |
| Precision | $\frac{tpv}{tpv + fpv}$ |
| Recall | $\frac{tpv}{tpv + fnv}$ |
| Specificity | $\frac{tnv}{fpv + tnv}$ |
| f1-score | $\frac{2tpv}{2tpv + fpv + fnv}$ |

## 3. Results and Discussions

The study was conducted on the data set consisting of 233 cases which was collected from various hospitals and health centers at perumbavoor. For evaluating the performance, the training data and testing data is evaluated.

The performance metrics such as specificity,accuracy,precision, recall and f1-score of SVM,ANN,Hybrid classifier models was evaluated and the confusion matrix is also shown in Figure :4(a),4(b) and 4(c).

```
Accuracy: 0.7238095238095238
Specificity: 0.7358490566037735
Classification Report:
              precision    recall  f1-score   support

           0       0.72      0.74      0.73        53
           1       0.73      0.71      0.72        52

    accuracy                           0.72       105
   macro avg       0.72      0.72      0.72       105
weighted avg       0.72      0.72      0.72       105

Confusion Matrix:
 [[39 14]
 [15 37]]
```

***Figure. 4(a)*** *SVM classifier*

```
ANN Classifier Accuracy: 0.8095238095238095
ANN Classifier Specificity: 0.8867924528301887
Classification Report:
              precision    recall  f1-score   support

           0       0.77      0.89      0.82        53
           1       0.86      0.73      0.79        52

    accuracy                           0.81       105
   macro avg       0.82      0.81      0.81       105
weighted avg       0.82      0.81      0.81       105

Confusion Matrix:
 [[47  6]
 [14 38]]
```

***Figure. 4(b)*** *ANN classifier*

```
Hybrid Classifier Accuracy: 0.8571428571428571
Hybrid Classifier Specificity: 0.8679245283018868
Classification Report:
              precision    recall  f1-score   support

           0       0.85      0.87      0.86        53
           1       0.86      0.85      0.85        52

    accuracy                           0.86       105
   macro avg       0.86      0.86      0.86       105
weighted avg       0.86      0.86      0.86       105

Confusion Matrix:
 [[46  7]
 [ 8 44]]
```

***Figure. 4(c)*** *Hybrid classifier*

The Table 4 describes the comparison of percentage of performance metrics and the Figure: 5(a), 5(b) and 5(c) show the performance of SVM, ANN and Hybrid classifier respectively. Figure 6 represents the heatmap of variables or attributes which is used to define the relationshipbetween data values. Figure 7 represent the correlation heatmap with target variable to find out the features that are strongly correlated or not correlated with target variable or with each other.

***Table 4.*** *Comparison of Classifies with Performance Metrics*

| Classifiers | Class | Accuracy | Specificity | Recall | Precision | F1 Score |
|---|---|---|---|---|---|---|
| SVM | 0 | 0.72 | 0.73 | 0.74 | 0.72 | 0.73 |
| | 1 | | | 0.71 | 0.73 | 0.72 |
| ANN | 0 | 0.80 | 0.88 | 0.89 | 0.77 | 0.82 |
| | 1 | | | 0.73 | 0.86 | 0.79 |
| Hybrid | 0 | 0.85 | 0.86 | 0.87 | 0.85 | 0.86 |
| | 1 | | | 0.85 | 0.86 | 0.85 |

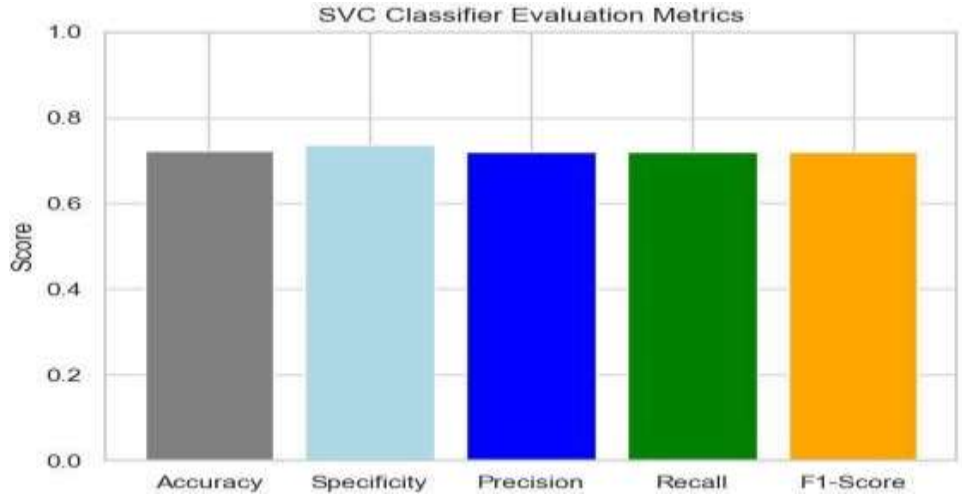


***Figure. 5(a)*** *SVM Classifier*

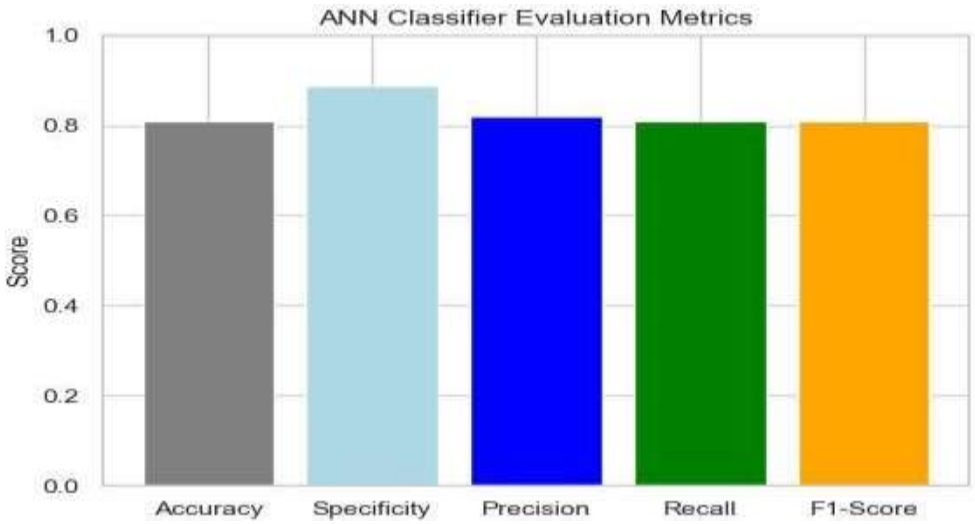


***Figure. 5(b)*** *ANN Classifier*

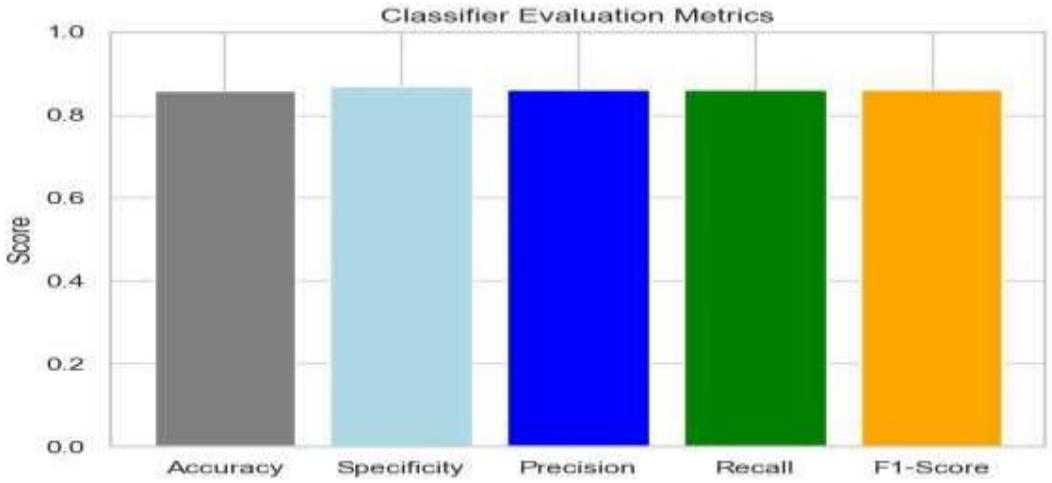


***Figure. 5(c)*** *Hybrid Classifier*

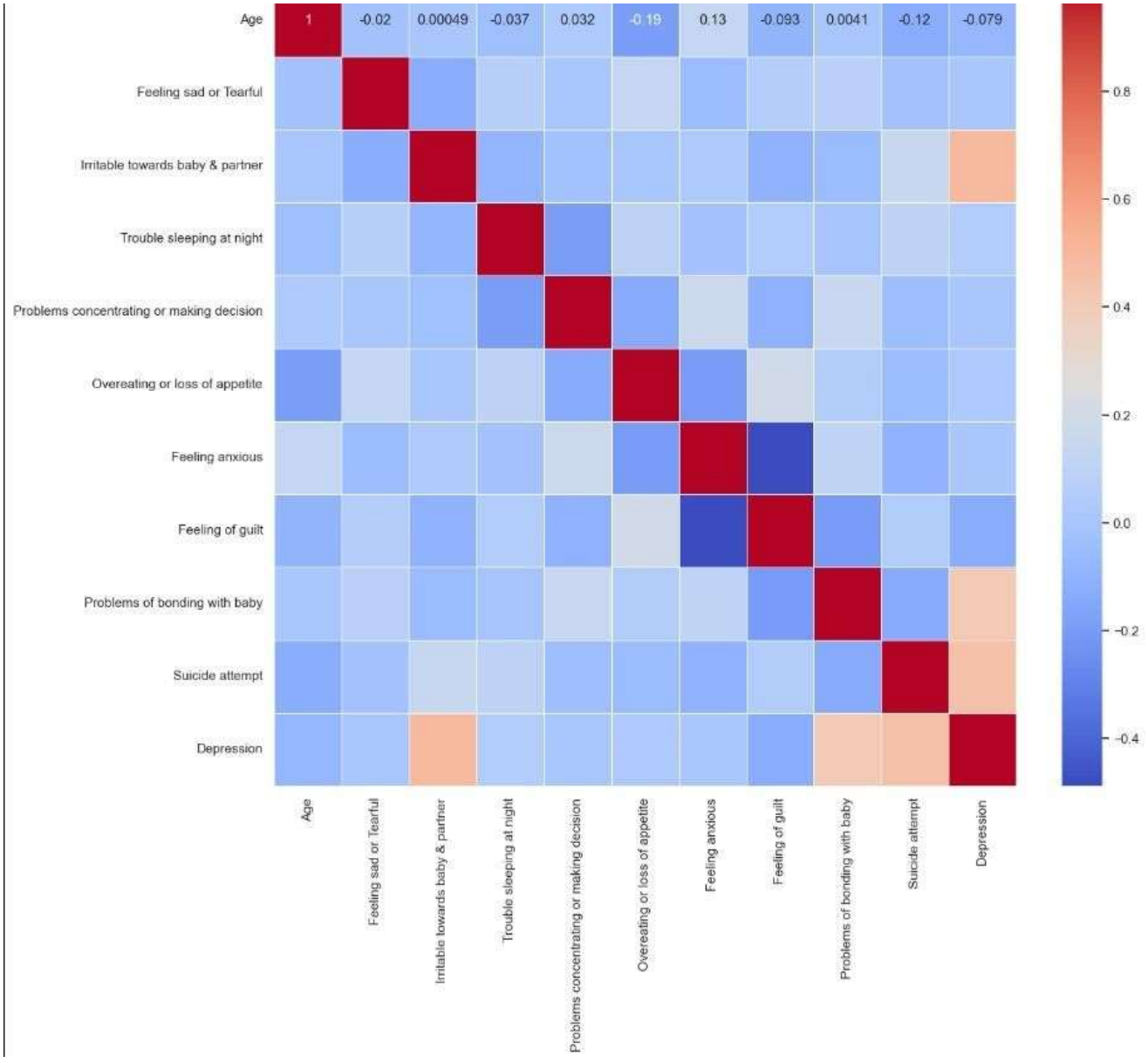


***Figure 6.*** *Heatmap of variables*

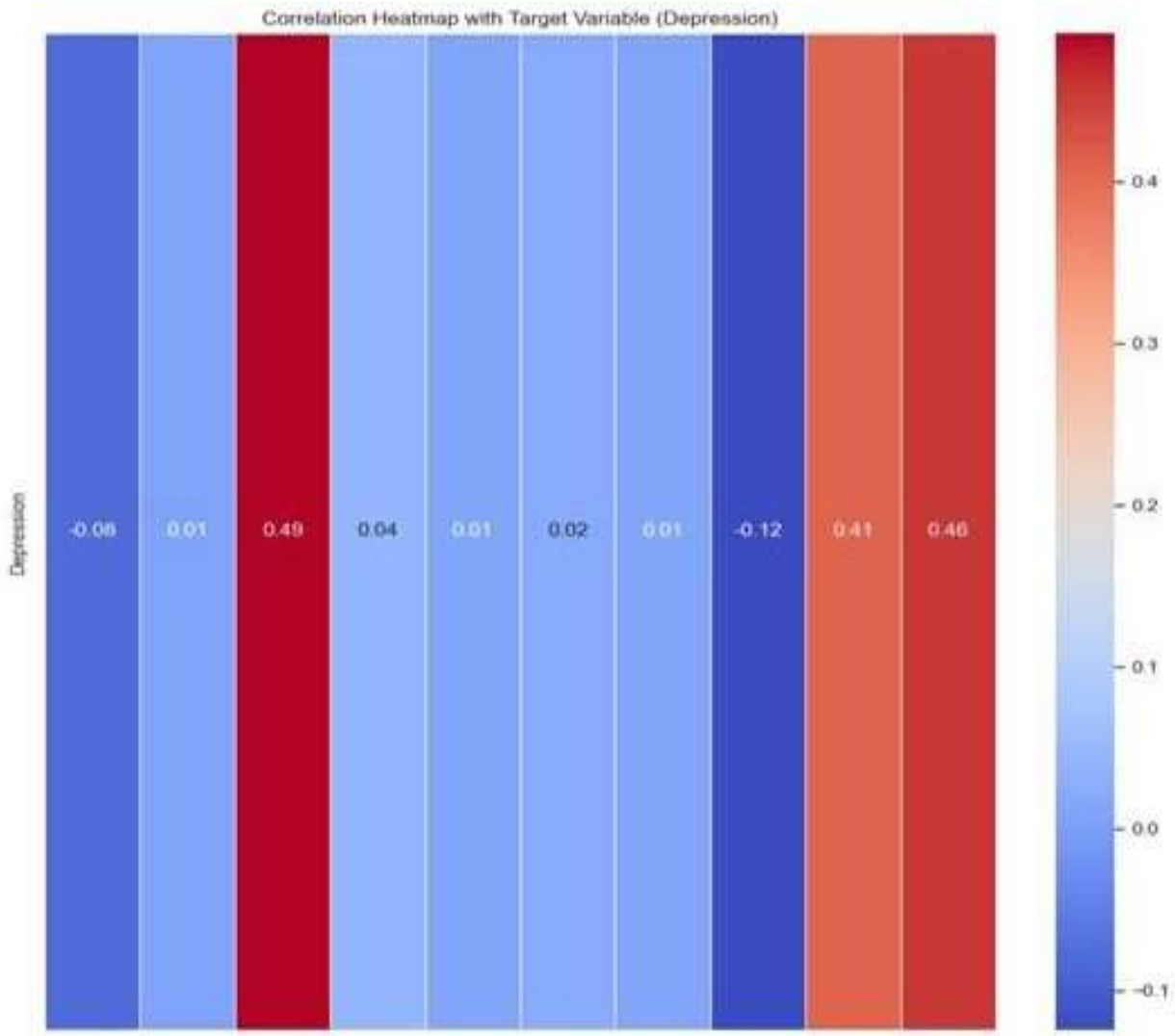


***Figure 7.*** *Correlation Heatmap with target variables*

## 4. Conclusions

PPD is becoming a serious issue among women in these days, as it is increasing 5- 10 % every year. Earlier prediction along with proper medication can reduce the seriousness of this disease. In the study, wehave incorporated the both classifier such as Support Vector Machine (SVM) classifier and Artificial Neural Network (ANN) classifier to achieve the desired output. By considering the PPD, the Support Vector Machine gives reasonable accuracy of 72 %, subsequently by Artificial Neural Network (ANN)algorithm with an accuracy of 80 % later the Artificial Neural Network is the Hybridclassifier with an optimal precision of 85 %.From these results, we can interpret that Hybridization of classifiers produce the best result when comparing to others.

Data mining algorithms can analyze large data set and perform more advanced data processing techniques which improve the detection of PPD at an earlier stage. Furthermore clinical information can be added to calibrate data mining algorithms for prediction and treatment. Similar works have been done and reported in the literature [15-19].

## Author Statements:

- **Ethical approval:** The conducted research is not related to either human or animal use.
- **Conflict of interest:** The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper
- **Acknowledgement:** The authors declare that they have nobody or no-company to acknowledge.
- **Author contributions:** The authors declare that they have equal right on this paper.
- **Funding information:** The authors declare that there is no funding to be acknowledged.

- **Data availability statement:** The data that support the findings of this study are available on request from the corresponding author. The data are not publicly available due to privacy or ethical restrictions.

## References


[1] Jennifer L Payne, Jamine Magurie. (2019). Pathophysiological mechanisms implied in Postpartum depression. Frontier in Neuro endocrinology. 52:165-180. DOI: 10.1016/j.yfrne.2018.12.001

[2] Carlson K, Mughal S, Azhar Y, Siddiqui. (2023). Postpartum depression. *In:stat Pearls, Treasure Island (FL).* PMID:30085612. Bookshelf ID: NBK519070

[3] Slomian J, Honvo G,Emonts P,Reginster J,Bruyere O. (2019). Consequences of maternal postpartum depression .A systematic review of maternal and infants outcomes. Womens Health (Lond). 15:1745506519844044. DOI: 10.1177/1745506519844044

[4] Bridget F.Hutchmens MSN,CNM,Joan Kearney. (2020). Risk factors for postpartum depression:An Umbrella Review. 65(1):96-108. DOI: 10.1111/jmwh.13067

[5] Dayeon Shin, KyungJu Lee. (2020). Machine Learning based Predictive modeling of Postpartum depression. *Journal of Clinical Medicine.* 9(9):2899. DOI:10.3390/jcm9092899

[6] Aris Supriyano. (2018). Implementation Data mining and Decision tree methodalgorithm c4.5 for PPD diagnosis. *E3S Webof conferences.* 73:12012. DOI:10.1051/e3sconf/20187312012

[7] Weina Zhang, et. al. (2020). Machine Learning models for the prediction of postpartumdepression. Application and Comparison based on Cohort study. *JMIR Med Inform*. 30;8(4):e15516. DOİ: 10.2196/15516

[8] Milan Kumari, Sunila Godara. (2011). Comparative study of Data mining classification methods in Cardiovascular Disease Prediction. *International Journal of computer science and Technology*. 2(2).

[9] Iran Fathima, Bushan Ud Din Abbasi, et. al. (2019). Prediction of PPD using ML urietechniques from social media text. *Expert Systems*. 36(4):e12409. DOI:10.1111/exsy.12409

[10] Trine Munk- Oslen,Xiaoqin Liu, Kathrine BangMadsen, et al. (2022). Postpartum depression :a developed and validated model predicting individual risk in new mothers. *Translational Psychiatry*. 12(419). DOI:10.1038/s41398-022-02190-8.

[11] Eldar Hochman, Becca Feldman, et. al. (2021). Development and validation of amachine learning based postpartumdepression model:A nationwide cohort study. *Depress Anxiety*. 38(4):400-411. DOI: 10.1002/da.23123

[12] Dayeon Shin, Kyung Ju Lee. (2020). Machine Learning based Predictive modeling of postpartum depression. *Clin Med*. 9(9):2899. DOI:10.3390/jcm9092899.

[13] Yiye Zhang, Shuojia Wang. (2021). Development and Validation of a machine learning algorithm for predicting the risk of postpartum depression among women. *J Affect Disord*. 279:1- 8. DOI:10.1016/j.jad.2020.09.113.

[14] Xiao M and Yan C. (2020). Risk Prediction for postpartum depression based on Random forest. *Journal of Central South University. Medical Sciences*. 45(10):1215-1222. DOI: 10.11817/j.issn.1672-7347.2020.190655.

[15]ÇOŞGUN, A. (2024). Estimation Of Turkey's Carbon Dioxide Emission with Machine Learning. *International Journal of Computational and Experimental Science and Engineering*, 10(1). https://doi.org/10.22399/ijcesen.302

[16]DAYIOĞLU, M., & ÜNAL, R. (2024). Comparison of Different Forecasting Techniques for Microgrid Load Based on Historical Load and Meteorological Data. *International Journal of Computational and Experimental Science and Engineering*, 10(4). https://doi.org/10.22399/ijcesen.238

[17]Polatoglu, A. (2024). Observation of the Long-Term Relationship Between Cosmic Rays and Solar Activity Parameters and Analysis of Cosmic Ray Data with Machine Learning. *International Journal of Computational and Experimental Science and Engineering*, 10(2). https://doi.org/10.22399/ijcesen.324

[18]Vijayadeep GUMMADI, & Naga Malleswara Rao NALLAMOTHU. (2025). Optimizing 3D Brain Tumor Detection with Hybrid Mean Clustering and Ensemble Classifiers. *International Journal of Computational and Experimental Science and Engineering*, 11(1). https://doi.org/10.22399/ijcesen.719

[19]Johnsymol Joy, & Mercy Paul Selvan. (2025). An efficient hybrid Deep Learning-Machine Learning method for diagnosing neurodegenerative disorders. *International Journal of Computational and Experimental Science and Engineering*, 11(1). https://doi.org/10.22399/ijcesen.701