Ensemble Machine Learning Model for Predicting Postpartum Depression Disorder

Anand
Artificial Intelligence-Data Science
Vivekananda Institute of Professional Studies-Technical Campus
Delhi
anand18581@gmail.com

Yash Sharma
Artificial Intelligence-Data Science
Vivekananda Institute of Professional Studies-Technical Campus
Delhi
yashksarma@gmail.com

Vansh Jain
Artificial Intelligence-Data Science
Vivekananda Institute of Professional Studies-Technical Campus
Delhi
vjain6981@gmail.com

Sandhya Tarwani
Artificial Intelligence-Data Science
Vivekananda Institute of Professional Studies-Technical Campus
Delhi
sandhya.tarwani@gmail.com

Abstract— **This study explores the development of an ensemble machine learning (ML) model to predict Postpartum Depression (PPD) disorder, leveraging chi-square test driven feature selection techniques and a diverse array of ML algorithms. Initially, chi-square test is employed to select the most influential features for PPD prediction. From a pool of candidate features, nine key attributes demonstrating the highest association with PPD are identified for model input. Subsequently, eight distinct ML techniques, including K Nearest Neighbors (KNN), Support Vector Machine (SVM), Random Forest (RF), Logistic Regression (LR), Decision Tree (DT), XG Boost (XGB), and Artificial Neural Network (ANN), are applied to develop individual predictive models. Following the training and evaluation of individual models, an ensemble approach is adopted to combine the strengths of multiple algorithms, enhancing prediction accuracy and robustness. The ensemble model aggregates predictions from diverse ML techniques, leveraging their complementary strengths to yield more accurate and reliable predictions of PPD risk. Performance evaluation metrics, such as accuracy, precision, recall, and F1-score, are employed to assess the efficacy of the ensemble model in comparison to individual ML algorithms. The results demonstrate the effectiveness of the ensemble approach in improving prediction accuracy and generalization capability, thereby offering a valuable tool for early identification and intervention in PPD cases. This research contributes to the advancement of predictive modelling in mental health by showcasing the utility of ensemble ML techniques in PPD prediction. The findings underscore the potential of feature selection and ensemble modelling in enhancing the accuracy and effectiveness of PPD risk assessment, thereby facilitating proactive interventions to support maternal mental health.**

Keywords—Postpartum Depression, Machine Learning, Boosting, Prediction, Mental Health

I. Introduction

Postpartum depression (PPD) is a prevalent healthcare issue that affects mothers after childbirth and can have serious consequences for both mother and child. Maternal depression during and after pregnancy affects 10–15 percent of women globally, according to the National Institute of Mental Health in the United States; however, in low- and middle-income nations, the number may go up as high as 18–25 percent [1]. PPD is a serious public health concern that not only has an impact women's health but also affects the physical, mental, and cognitive development of children [2]. The babies of mothers suffering from PPD are associated with inadequate nutrition and poor general health, and an increased frequency of diarrheal episodes; In addition, maternal suicides and infanticides under severe circumstances were documented [3].

Hence, it is important to predict PPD because its symptoms are similar to other mental disorders. These symptoms encompass a depressed mood, feeling sad, tearful, anxious and guilty, loss of concern, feelings of guilt and worthlessness, and even suicidal thoughts. Even after having a substantially negative influence on a mother's mental health, PPD is still not well understood and often remains undiagnosed until serious symptoms develop.

Early detection of PPD can be greatly enhanced by machine learning (ML) algorithms, which are capable of analysing large data sets and carrying out more complex computations [4]. ML techniques have been used to predict the severity, duration, chronicity, and response to treatment of major depressive disorders [5]. The Support Vector Machine and Random Forest algorithms are the two primary ML algorithms commonly documented in depression prediction studies [6].

In this paper, we have used various ML techniques to predict PPD. Chi-square test was used to select 9 features and using 8 techniques (K Neighbours [KNN], Support Vector Machine [SVM], Random Forest [RF], Logistic Regression [LR], Decision Tree [DT], XG Boost [XGB], Artificial Neural Network [ANN]), we develop an Ensemble machine learning model to aid predicting PPD. We have used data from PPD dataset from Kaggle having 11 columns and 1503 rows and divided into 80-20 train-test ratio for developing and testing ML models.

The main objectives of our proposed system are stated below:

- Comparing the suitability of various machine learning algorithms in predicting PPD.
- Address the need for early detection and intervention to prevent adverse outcomes for both mothers and infants.

II. Literature Review

Acharya et al. [7] discussed the overview of the current methods for identifying and treating postpartum depression (PPD). Their methodology found out that Support Vector Machine (SVM) has been the most accurate classifier for PPD detection in the majority of cases. SVM predicts PPD risk levels with a higher precision rate than other classifiers. Due to its superior ability to estimate cognitive abilities in real-

979-8-3503-6486-6/24/$31.00 ©2024 IEEE

time, SVM is becoming a popular substitute for PPD detection. Prabhashwaree and Wagarachchi [8] identified causing factors which include mother's family, social background, and other pertinent information in order to diagnose postpartum depression (PPD) in Sri Lankan mothers six months after giving birth. The Edinburgh Postpartum Depression Scale (EPDS) was used to categorize the risk rates of PPD into four groups: mild, moderate, severe, and profound. The Feed Forward Neural Network (FFANN) model performed the best with multi-classification out of all the models, with an accuracy of 95 percent in predicting PPD risk levels. Wakefield and Frasch [9] identified pregnant patients who are at risk of developing depression in the postpartum period (PPD) by utilizing distributed random forest (DRF) models. A total of 10,038 pregnant women who had never given birth before were included in the study. The risk of PPD in these women was estimated using the DRF models. Using DRF models made it easier to identify potential risk factors and develop a predictive framework for PPD in pregnant nulliparous women.

Jiménez-Serrano et al. [10] focused on generating classification models to identify females at risk of postpartum depression (PPD) in the first week following childbirth. Predictive models were trained utilizing machine learning approaches to estimate the risk of PPD. In order to predict the danger of PPD, the research takes into consideration a variety of variables, comprising socioeconomic status, traumatic life events, emotional changes during pregnancy, and personal and family histories of mental illness. Saqib et al. [4] found that PPD can be foreseen using a range of machine learning (ML) techniques, including regression, XGBoost, decision trees, Naive Bayes, support vector machines, random forests, and artificial neural networks. Supervised learning techniques were adopted to predict PPD in the majority of the investigation that was found. Machine learning algorithms may be advantageous in the early discovery of post-traumatic depression (PPD) based on the extent of values noticed in the area under the receiver operating characteristic curve (0.77-8.93). Qasrawi et al. [11] has used machine learning models such as Random Forest and Gradient Boosting, mother anxiety and depression were foreseen during the COVID-19 pandemic. 3569 women from five Arab nations volunteered in the study, which discovered that anxiety and depression were significantly foreseen by variables like income, stress during pregnancy, family support, social support, and financial difficulties. During lockdown, anxiety symptoms were more predominant in pregnant females than in depression, which may have consequences for the health of the mother and child. Kimwomi et al. [12] show that in prior analyses it have been noticed that popular algorithms such as the Support Vector Machine and Bayes Net classifier are advantageous in the prediction of postpartum depression. The review aimed to inform scientists about the ongoing status of the field, encourage the role of machine learning models in postpartum depression prediction, and direct future research. Gopalakrishnan et. al [13] has examined information from the Edinburgh Postnatal Depression Scale (EPDS), the Patient Health Questionnaire-9 (PHQ-9), and the Postpartum Depression Screening Scale (PDSS) surveys to foresee postpartum depression. Machine learning algorithms, particularly the Extremely Randomized Trees (XRT) algorithm, were aggregated with background data, PHQ-9, and PDSS data in order to accurately predict postpartum depression. The research addresses the question of imbalances in datasets for machine learning classifiers and encompasses a cost in the objective function to weaken false negatives in the prediction of postpartum depression. Amit et al. [14] produced a machine learning model that uses information from electronic health records to forecast a patient's possibility of developing postpartum depression (PPD) after giving birth. The model's area under the curve (AUC) ranged from 0.72 to 0.74 points. García-Gómez et al. [15] used the multilayer perceptron approach in an attempt to predict postpartum depression. Data was gathered for eight weeks after delivery, and it was discovered that the model had the highest accuracy rate of 95 percent. As shown in the table 1 we have compared our paper with other recent papers.

TABLE I COMPARISON OF OUR RESEARCH WORK

Ref no.	***Authors name***	***Model***	***Accuracy***
	Our paper	Multi-layered Stacking	94.31%
[8]	Prabhashwaree and Wagarachchi (2022)	FFAN	97.08%
[9]	Wakefield and Frasch (2022)	DRF	95(±0.02)%
[10]	Jimenez-Serrano et. al	ANN	79%
[11]	Qasrawi R et al. (2022)	Gradient Boosting	83.3%
[13]	Gopalakrishnan et al. (2022)	XRT	73%
[15]	García-Gómez et al. (2009)	Neural Network	95%

While the accuracy of our proposed model (94.31%) is slightly lower than that achieved by Prabhashwaree and Wagarachchi [8] (97.08%), Wakefield and Frasch [9] (95 %) and García-Gómez et al. [15] (95%), our model still gives reliable and consistent results and ranks amongst top performing models.

III. DATA COLLECTION & METHODS

A. *Data Selection and Description*

To conduct this study, we selected PPD dataset from Kaggle which was gathered using a questionnaire from new mothers using google forms. It has 11 columns and 1503 rows. It contains various attributes like Timestamp, Age, Problems of bonding with baby, Irritable towards baby & partner, Trouble sleeping at night, Problems concentrating, feeling sad or Tearful, Suicide attempt, Feeling anxious and etc. as shown in Table 2. Here 'Count' is total number of response collected from different new mothers, 'Unique' is the no. of unique responses for a given feature like 'yes', 'no' & etc. and 'Top' is the response with highest frequency.

The data is properly documented with descriptions, sources, and file formats. It is licensed under CC0: Public Domain and has a Kaggle usability score of 9.4 [16].

TABLE II DATASET DESCRIPTION

Feature	***Count***	***Unique***	***Top***
timestamp	1503	90	6/15/2022 22:24
Age	1503	5	25-30

Feeling sad or tearful	1503	3	Yes
Problem concentrating or making decision	1491	3	No
Irritable towards baby & partner	1497	3	Yes
Suicide attempt	1503	3	No
Overeating loss of appetite	1503	3	No
Feeling anxious	1503	2	Yes
feeling of guilt	1494	3	No
Trouble sleeping in night	1503	3	Two or more days a week
Problem of bonding with baby	1491	3	No

B. Machine learning algorithms

i. Decision Tree

Decision tree was first introduced by Quinlan in 1993 [17]. A decision tree model resembles a tree, it is used to forecast outcomes based on a sequence of choices. To create a decision tree model, we first select the important features and pre-process the data. Next, we train the data using algorithms such as the C4.5 algorithm. The trained model is evaluated and used for prediction.

ii. Random Forest

Random Forest is a type of Bagging ensemble machine learning technique [18]. It is formed by combining several decision trees to improve their accuracy and performance. A random forest model can be constructed by training different decision trees on a random subset of pre-processed dataset. Random forest then makes prediction by aggregating the prediction from many single decision trees [18].

iii. Support Vector Machine

Support Vector Machine seeks to identify the best hyperplane to divide data points into distinct groups such that it maximizes the margin from support vectors. This improves performance on unseen data and improves generalization. The optimal hyperplane divides the data point into 2 different groups, and the maximum margin is taken from the support vector of both groups.

iv. K-Nearest Neighbours

K in KNN stands for the number of closest neighbours taken into account when a data point is being predicted. By determining which class is most prevalent among its K closest neighbours, KNN predicts the class of the data point. Equation 1 illustrates how the Euclidean function calculates K-nearest neighbours. The two data points that need to be compared, i and j, each have n attributes [20].

$$distance = \sqrt{(i_1 - j_1)^2 + (i_2 - j_2)^2 + \ldots + (i_n - j_n)^2} \quad (1)$$

v. Logistic regression

Logistic regression is a classification algorithm that predicts the likelihood of an instance belonging to a given class. The sigmoid function is used to transform a linear regression into a logistic regression [20]. The formula for Logistic regression is given by,

$$y = \frac{1}{1+e^{-(b_0+b_1x_1+b_2x_2+\cdots+b_nx_n)}} \quad (2)$$

Here x_n is the input value, y is output value, b_0 is bias and b_n is coefficient for x_n.

vi. XGBoost

XGBoost (Extreme Gradient Boosting) was first proposed by Chen and Guestrin [21]. It is a scalable tree-boosting system [21]. Fundamentally, XGBoost works on the boosting principle, in which weak learners (usually decision trees) are trained successively to correct mistakes made by their predecessors. It uses gradient decent optimization to iteratively fit the model to the residuals of prior predictions in order to minimize loss function.

vii. Neural network

Neural Network are essentially made up of nested nodes that are connected and arranged into layers, with a weight assigned to each connection between nodes to indicate its significance and An activation function is applied to the weighted sum of inputs to determine each node's output as shown in Fig. 1. The earliest research in neural modelling can be found in 1943 by McCulloch & Pitts [22]. Later Perceptron, a type of neural network which can learn to divide input patterns into two categories, was developed by Rosenblatt [24].

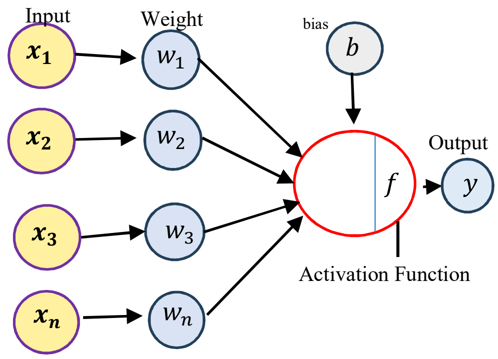


Figure 1. Perceptron

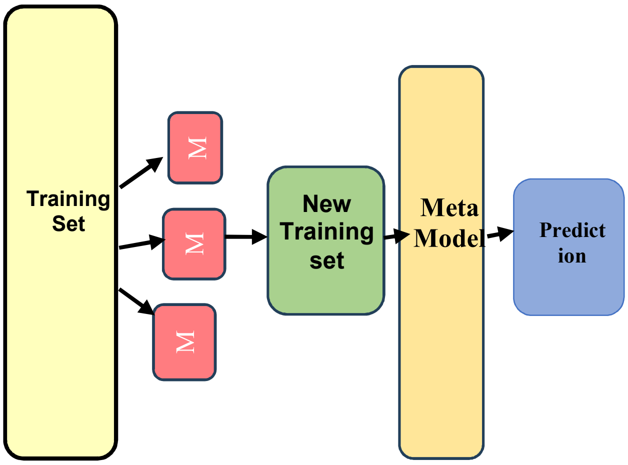
viii. Voting Classifier

The basic idea behind a voting classifier is to train multiple classifiers individually on the same dataset and then aggregate their results to generate a final prediction [23]. In voting classifier, we have used DT, RF, KNN and SVC as classifier and performed hard voting in which the model selects the label that is most predicted by multiple classifiers as final prediction.

ix. Stacking Ensemble Method

The stacked Ensemble method is a machine-learning method in which multiple classifiers are stacked into multiple layers to make a more accurate and robust model and improve overall performance. In 1992, Wolpert introduced the term “stacked generalization”, wherein he devised a method by combining multiple predictive algorithms into a single model to increase prediction accuracy [24]. It comprises of 2 levels, a base classifier and a meta classifier as shown in Figure 2.

For training the model, the original dataset is divided in training and validation sets and Base classifiers are trained individually with training set and later fed through validation set. For Meta classifiers training, the dataset includes the original features from training data and predictions by base classifiers. Finally, the unseen data is fed through both the



trained layers and final predictions are made

Figure 2. Mechanism of Stacking Ensemble Method

C. Evaluation Metrics

We evaluate the proposed model's performance by comparing it to all other ML models using various performance metrics, including F1-score, Accuracy, Precision, and Recall and calculated it using confusion matrix as shown in Table 3.

TABLE III CONFUSION MATRIX

	Predicted 0	***Predicted 1***
Actual 0	True Negative (TN)	False Positive (FP)
Actual 1	False Negative (FN)	True Positive (TP)

i. Accuracy: The measurement of accuracy is percentage of correctly predicted instances, both positive and negative out of all predicted instances [25].

$$Accuracy(A) = \frac{TP+TN}{TP+TN+FP+FN} \tag{3}$$

ii. Precision: Precision is expressed as the ratio of correctly predicted positive instances to all instances identified as positive by the model [25].

$$Precision(P) = \frac{TP}{TP+FP} \tag{4}$$

iii. Recall: Recall, also know as sensitivity, gauges the accuracy of identifying positive predicted instances by the model among all instances that are genuinely positive [25].

$$Recall(R) = \frac{TP}{TP+FN} \tag{5}$$

iv. F1-Score: As shown in Formula 6, F_1score defines the harmonic mean of both recall and precision

$$F_1score = 2 * \frac{P*R}{P+R} \tag{6}$$

IV. RESEARCH METHODOLOGY

Here we have used data set containing 1503 row and 11 columns. And divided it in the train-test ratio of 80:20 (training data =1192 samples, testing data = 299 samples).

Out of 11 features (columns) we dropped timestamp as it was irrelevant. Further we selected 'Feeling anxious' as our target variable. And to provide more interpretability to our model, we used chi-square testing to select features out of the 9 features.

Chi-Square Test is a statistical method which ascertains the relationship between categorical variables and used as test for independence of variable [26]. It provides us with an understanding of the strength of the association between the two variables [26]. In this we first assume Null Hypothesis(H_0) which states that there is no association between categorical variables. We also take an Alternate Hypothesis (H_1) which states that, there is an association between given variable. Using chi- square statistics (χ^2) and Degree of freedom (df) we calculate p-value. If p-value is less than significance level (α) then we reject Null Hypothesis. If p-value is greater than significance level α, we fail to reject Null Hypothesis. For our study we have taken significance level α to be 0.05.

TABLE IV P VALUES OF FEATURES

Feature	***p-values***
Age	1.63e-02
Feeling sad or tearful	2.73e-02
Irritable towards baby and partner	5.23e-39
Trouble sleeping at night	4.67e-11
Problems concentrating or making decision	4.62e-32
Overeating or loss of appetite	7.85e-14
Problems of bonding with baby	8.65e-24
Suicide attempt	7.21e-29
Feeling of guilt	1.36e-97

In table 4 above , since p-value of all the features are less than 0.05, we reject H_0 and there is an association between the features and target variable 'Feeling anxious'. Thus we take all 9 features for training the model.

Of 1503 participants 80% (1192) were randomly selected and 8 machine learning algorithms including Decision Tree, Random Forest Classifier, Logistic Regression, Support Vector Classifier, K Neighbors Classifier, XG Boost Classifier, Neural Network, and Voting Classifier (DT, RF, KNN, SVC) were trained on training data. Then we compared different evaluation metrics like accuracy, precision, recall and F1score of the trained model on testing data.

Further in our study, we built an ensemble machine-learning model involving a multilayered stacking approach. We aim to improve the accuracy and efficiency of postpartum depression detection and determine the most effective method for detecting this significant mental health disorder through the integration of various algorithms and the utilization of pertinent feature datasets. To Identify the best method for PPD detection, we contrasted the performance of our ensemble approach with that of individual algorithms.

As illustrated in figure 3, we trained each classifier separately in the base classifier having DT, KNN, SVM, and RF using training data. Following training, the layer 1 classifier produces a prediction for each data point of validation set, which is then combined by original features of training dataset used as an input by the layer 2 classifier comprising of Neural Network (NN), SVM, and XGB. Likewise, the Meta classifier is trained using prediction of layer 2 classifier.

After training the ML model we tested it on 299 unseen data samples and created a confusion matrix. Using confusion matrix, we evaluate several performance metrics of model and checked for and over fitting or under fitting of the model. Finally we compared it with previously trained 8 models on accuracy, precision, recall and F_1 score.

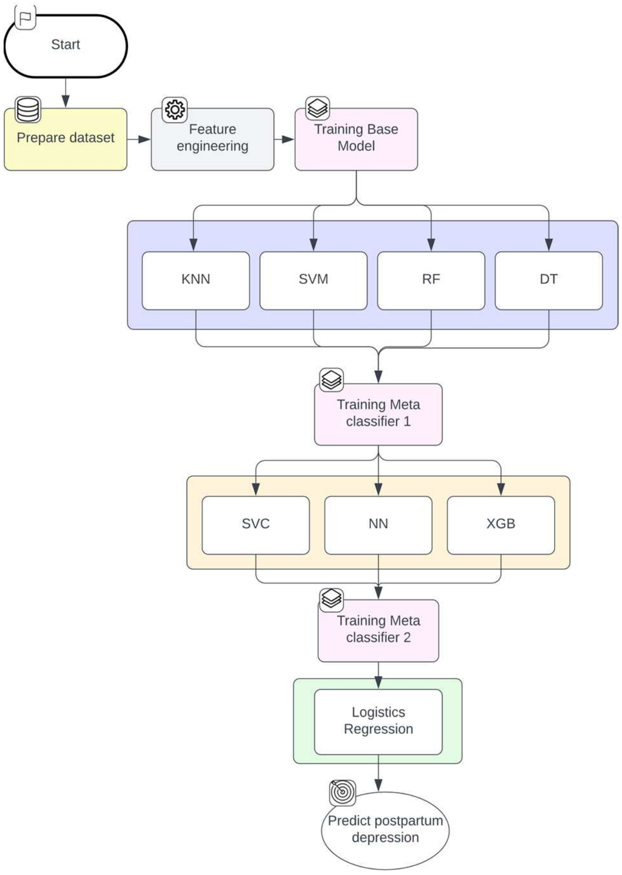


Figure 3. Research Methodology

V. RESULT AND DISCUSSION

A. Result Analysis

In our research, we investigated how well 9 different machine learning models can predict PPD. To evaluate the performance of our model we predicted the outcome of the model on 299 unseen instances and made confusion matrix.

As shown in confusion matrix in Figure 4, out of total 299 instances, the model correctly predicted 196 positive cases and 88 negative cases. However it made 9 false positive errors and 8 false negative errors. Further we calculated performance metrics based of confusion matrix for multi-layered stacking model and found accuracy to be 94.31%.

Additionally, we made sure the model was neither overfitting nor underfitting.

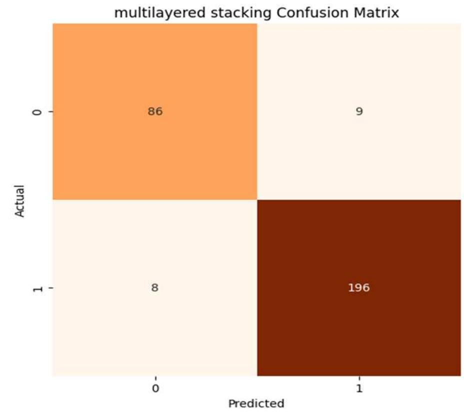


Figure 4. Confusion Matrix of Proposed Approach

B. Performance Measurement Table

The results of the evaluation of several models on different metrics such as accuracy precision and recall are given below in Table 5.

TABLE V PERFORMANCE OF VARIOUS CLASSIFIERS

Model	*Accuracy*	*Precision*	*Recall*	*F1 Score*
Logistic Regression	0.7726	0.7742	0.9412	0.8495
Decision Tree	0.9064	0.9489	0.9118	0.9300
Random Forest	0.9130	0.9320	0.9412	0.9366
SVM	0.9152	0.9388	0.9324	0.9356
KNN	0.9264	0.9505	0.9412	0.9458
XGBoost	0.9308	0.9461	0.9493	0.9477
Voting Classifier (dt, knn, svc, rf)	0.9365	0.9695	0.9362	0.9526
Neural Network	0.9375	0.9621	0.9426	0.9522
Multi-layered Stacking (proposed model)	0.9431	0.9561	0.9608	0.9584

Our proposed model achieved the highest accuracy of 94.31% followed by neural network at 93.75%. Other models which performed better were VC and XGB with 93.65% and 93.08% accuracy respectively. We also compared Recall, Precision and F1 score of different models and plotted it on line chart as shown in Figure 5.

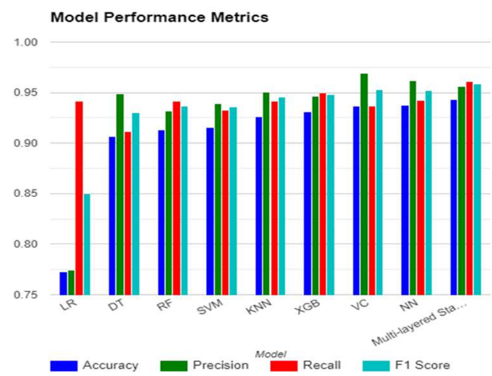


Figure 5. Performance comparison for ML model for other performance metrics

Multilayered Stacking algorithm also achieved the highest recall value (96.08%) indicating that it is highly effective in identifying true positive cases, minimizing the risk of false negative cases, and thus reducing the risk of missed diagnoses which is essential in medical diagnostic field for early identification and intervention.

Our proposed model outperformed all the ML algorithms used in our study i.e. LR, DT, RF, SVM, KNN, XGB, VC, NN in accuracy, recall and F_1 score. Based on these results we suggest that Multilayer Stacking algorithm is the one of the most effective approach for predicting postpartum mental health and recommend its utilization, in future research and clinical environments.

VI. FUTURE WORK

While the present study provides a robust foundation for predicting Postpartum Depression (PPD) using ensemble machine learning (ML) techniques, several avenues for future research could further enhance the accuracy and applicability of the predictive model by training on large number of dataset. Incorporation of additional data like demographic, lifestyle, and psychosocial factors such as pre-existing depression, economic status of family, gender of baby and etc. could enrich the feature set and potentially improve the model's predictive power. Exploring advanced feature engineering techniques and incorporating domain-specific knowledge may also yield more informative features for PPD prediction.

VII. CONCLUSION

In conclusion, this study presents a comprehensive approach to predicting PPD using ensemble machine learning models, incorporating chi-square test driven feature selection and a diverse range of ML algorithms. By leveraging data from a PPD dataset and employing an 80-20 train-test split, the ensemble model demonstrates promising results in accurately identifying individuals at risk of PPD. Ultimately, the application of machine learning techniques holds significant promise in improving mental health outcomes for postpartum individuals and their families.

REFERENCES

[1] J. Fisher et al., "Prevalence and determinants of common perinatal mental disorders in women in low- and lower-middle-income countries: a systematic review," Bulletin of the World Health Organization, vol. 90, no. 2, pp. 139-149H, Nov. 2011, doi: 10.2471/blt.11.091850.

[2] L. MURRAY and P. J. COOPER, "Effects of postnatal depression on infant development," Archives of Disease in Childhood, vol. 77, no. 2, pp. 99–101, Aug. 1997, doi: 10.1136/adc.77.2.99.

[3] Muzik, M., & Borovska, S. (2010). Perinatal depression: implications for child mental health. Mental health in family medicine, 7(4), 239–247.

[4] K. Saqib, A. F. Khan, and Z. A. Butt, "Machine Learning Methods for Predicting Postpartum Depression: Scoping Review," JMIR Mental Health, vol. 8, no. 11, p. e29838, Nov. 2021, doi: 10.2196/29838.

[5] R. C. Kessler et al., "Testing a machine-learning algorithm to predict the persistence and severity of major depressive disorder from baseline self-reports," Molecular Psychiatry, vol. 21, no. 10, pp. 1366–1371, Jan. 2016, doi: 10.1038/mp.2015.198.

[6] A. B. R. Shatte, D. M. Hutchinson, and S. J. Teague, "Machine learning in mental health: a scoping review of methods and applications," Psychological Medicine, vol. 49, no. 09, pp. 1426–1448, Feb. 2019, doi: 10.1017/s0033291719000151.

[7] A. Acharya, R. Ramesh, T. Fathima, T. Lakhani, and S. K. S, "Clinical tools to detect Postpartum Depression based on Machine learning and EEG: A Review," 2023 2nd International Conference on Computational Systems and Communication (ICCSC), Mar. 2023, doi: 10.1109/iccsc56913.2023.10142970.

[8] T. H. K. R. Prabhashwaree and N. M. Wagarachchi, "Towards Machine Learning Approaches for Predicting Risk Level of Postpartum Depression," 2022 6th SLAAI International Conference on Artificial Intelligence (SLAAI-ICAI), Dec. 2022, doi: 10.1109/slaai-icai56923.2022.10002477.

[9] C. Wakefield and M. G. Frasch, "Identifying patients requiring treatment for depression in the postpartum period from common electronic medical record data available antepartum using machine learning," Dec. 2022, doi: 10.1101/2022.11.30.22282941.

[10] S. Jiménez-Serrano, S. Tortajada, and J. M. García-Gómez, "A Mobile Health Application to Predict Postpartum Depression Based on Machine Learning," Telemedicine and e-Health, vol. 21, no. 7, pp. 567–574, Jul. 2015, doi: 10.1089/tmj.2014.0113.

[11] R. Qasrawi et al., "Machine learning techniques for predicting depression and anxiety in pregnant and postpartum women during the COVID-19 pandemic: a cross-sectional regional study," F1000Research, vol. 11, p. 390, Apr. 2022, doi: 10.12688/f1000research.110090.1.

[12] G. Kimwomi, M. Mgala, F. Mwakondo, and P. Kimeto, "Machine Learning Prediction Models for Postpartum Depression, a Review of Literature," International Journal of Computer Applications Technology and Research, vol. 11, no. 06, pp. 205–212, Jun. 2022, doi: 10.7753/ijcatr1106.1005.

[13] A. Gopalakrishnan, R. Venkataraman, R. Gururajan, X. Zhou, and G. Zhu, "Predicting Women with Postpartum Depression Symptoms Using Machine Learning Techniques," Mathematics, vol. 10, no. 23, p. 4570, Dec. 2022, doi: 10.3390/math10234570.

[14] G. Amit et al., "Estimation of postpartum depression risk from electronic health records using machine learning," BMC Pregnancy and Childbirth, vol. 21, no. 1, Sep. 2021, doi: 10.1186/s12884-021-04087-8.

[15] J. M. García-Gómez et al., "Prediction of Postpartum Depression Using Multilayer Perceptrons and Pruning," Methods of Information in Medicine, vol. 48, no. 03, pp. 291–298, 2009, doi: 10.3414/me0562.

[16] Kaggle "PostPartum Depression" Kaggle Dataset, Online [Available]:https://www.kaggle.com/datasets/parvezalmuqtadir2348/postpartum-depression

[17] S. L. Salzberg, "C4.5: Programs for Machine Learning by J. Ross Quinlan. Morgan Kaufmann Publishers, Inc., 1993," Machine Learning, vol. 16, no. 3, pp. 235–240, Sep. 1994, doi: 10.1007/bf00993309.

[18] M. Pal, "Random forest classifier for remote sensing classification," International Journal of Remote Sensing, vol. 26, no. 1, pp. 217–222, Jan. 2005, doi: 10.1080/01431160412331269698.

[19] V. P. Vishwakarma, S. Dalal, and V. Sisaudia, "Efficient Feature Extraction using DWT-DCT for Robust Face Recognition under varying Illuminations," 2018 2nd IEEE International Conference on Power Electronics, Intelligent Control and Energy Systems (ICPEICES), Oct. 2018, doi: 10.1109/icpeices.2018.8897464.

[20] R. O. Sinnott, H. Duan, and Y. Sun, "A Case Study in Big Data Analytics," Big Data, pp. 357–388, 2016, doi: 10.1016/b978-0-12-805394-2.00015-5.

[21] T. Chen and C. Guestrin, "XGBoost," Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, Aug. 2016, doi: 10.1145/2939672.2939785.

[22] McCulloch, W.S., Pitts, W., A logical calculus of the ideas immanent in nervous activity.Bul l. Math. Phys. 5 (1943) 115-133

[23] T. G. Dietterich and G. Bakiri, "Solving Multiclass Learning Problems via Error-Correcting Output Codes," Journal of Artificial Intelligence Research, vol. 2, pp. 263–286, Jan. 1995, doi: 10.1613/jair.105.

[24] D. H. Wolpert, "Stacked generalization," Neural Networks, vol. 5, no. 2, pp. 241–259, Jan. 1992, doi: 10.1016/s0893-6080(05)80023-1.

[25] H. Dalianis, "Evaluation Metrics and Evaluation," Clinical Text Mining, pp. 45–53, 2018, doi: 10.1007/978-3-319-78503-5_6.

[26] M. L. McHugh, "The Chi-square test of independence," Biochemia Medica, pp. 143–149, 2013, doi: 10.11613/bm.2013.018.