Self-Sturdy Deep Neural Network for Improving Postpartum Depression Detection

D. Suganthi[1], A. Geetha[2]

[a] Assistant Professor, Department of Computer Science, Chikkanna Government Arts College, Tiruppur - 641602, Tamil Nadu, India
Corresponding Author Email: *suga07bca[at]gmail.com*

[2]Associate Professor, Department of Computer Science, Government Arts and Science College, Avinashi - 641654, Tamil Nadu, India

Abstract: ***Postpartum depression (PPD) is a prevalent maternal health issue which adversely affects both mothers and their newborns. It negatively impacts mother-baby bonding, infant cognitive development, language, behaviors, sleep quality, and physical-mental health. In extreme cases, depressed mothers may resort to suicide or infanticide. PPD needs to be overlooked, requiring timely and early-stage treatment to prevent serious complications. For this reason, Osprey Parameter Optimized MLP (OPOMLP) is developed which utilizes the Multi-Layer Perception (MLP) for classification and Osprey optimization algorithm (OOA) for the feature selection and parameter optimization. But, model's performance was lower due to its inefficiency to perform on larger datasets. In this paper, Self-Sturdy Deep PPD network (SSPPDnet) is developed to tackle the for-mentioned challenges for the effective PPD prediction. In this method, Convolutional Neural Network (CNN) and Long Short-Term Memory (LSTM) are used to learn all feature types autonomously. The temporal aspects of data are captured using LSTM, which derives time-lapse attributes individually for all data format. CNN converts the extracted time-series features into the feature vector to capture the relationship between divergent. The retrieved features are fed into Self-Sturdy Deep Neural Network (SSDNN). SSDNN is developed for faster convergence with fixed global learning, resulting in lower initial learning rate training. Fully Connected (FC) layers are adopted to compute the postpartum accomplishment levels by combining the women's demographic information, associative characteristic and time-lapse attributes. The softmax layer is employed for PPD detection and classification. Finally, test results show that the SSPPDnet model obtains an accuracy of 95.03% and 97.63% on datasets respectively outperforming other models.***

Keywords: Postpartum Depression, Stochastic Gradient Descent, Per-Layer Stabilizer, Deep Neural Network, Long Short-Term Memory

1. Introduction

The primary cause of mother perinatal mortality is postpartum depression (PPD), a potentially severe mental health condition that may manifest up to a year after delivery [1]. PPD negatively affects a mother's physical and emotional well-being as well as the growth and development of her offspring [2]. Poor mother-baby bonding, perinatal physiological and mental growth, language development, infant behaviour, and sleep quality have all been linked to PPD symptoms [3]. Women who give birth are susceptible to a range of mental health conditions which is common in PPD [4]. PPD's clinical indications include difficulties falling or residing drowsy sleeping extended periods of time, mood swings, decreased appetite, severe concern about the baby, dread of hurting someone, sorrow or exorbitant sobbing, desperation and culpability sensations, trouble paying attention and recalling things, lack of interest in normal tasks and desires, suicidal thoughts, and persistent thoughts of suicide [5, 6]. Earlier PPD avoidance lowers the morality rate, enhances the physical and mental health of the mother, and increases the wellbeing of the infants.

It is widely acknowledged that participation in pharmacological treatment or psychological counselling can reduce the risk of PPD and improve the health of both mother and child; however, the efficacy of antidepressant medication contact during pregnancy and breastfeeding remains unclear [7, 8]. While the risk factors for PPD are well acknowledged, there are no measurable risk assessment tools to help in the diagnosis and treatment of women during the postpartum period. Electronic Health Records (EHR) from the primary care providers' include precise information about their patient's medical state such as diagnosis, medication, treatments, and test findings. This information, together with the patient's socio-demographic background, is an important resource for assessing risk and diagnosing illness [9, 10]. On the other hand, the EPDS evaluation lacks demographic statistics or social support information, potentially causing concerns about its inability to detect a wide range of pre- and postpartum symptoms and illnesses, and lower positive predictive value in normal populations.

To overcome the aforementioned difficulties, Machine Learning (ML) models have been created, which are essential in predicting PPD at an early stage [11]. ML models efficiently classify vast amounts of EHR data, minimizing computational difficulties, lengthy process durations, and complexity. ML are also utilized in psychiatry to rapidly assess data, make precise judgments, and give proper therapy based on PPD development and stage differentiation [12, 13]. Various ML models have been developed for the prediction of PPD. For example, ML model [14] was created to predict the risk of PPD in pregnant women. This strategy included collecting, pre-processing, and normalizing questionnaire EHR data on the needs of expecting mother's demographics, diagnoses, pharmaceutical medication, treatments, laboratory measures and socioeconomic health determinants. Structured query and natural language processing (NLP) were used to identify the smallest collection of characteristics from the datasets. Finally, several ML models such as RF, DT, LR, XGBoost, and Multi-Layer Perceptron (MLP) [14] were used to find collective patterns across features that most effectively differentiate outcome classes for PPD detection.

However, this model constitutes the common characteristics of PPD patients such as mental health and behavioral changes, women's demographic positions, health service application, mental health history, newly identified psychological disorders during pregnancy, other obstetric and/or medical conditions during pregnancy and vital signs factors. Moreover, other physiological characteristics of PPD patients (mental health and behavioral changes) must be analyzed in order to predict the mental state of women with PPD risk. Furthermore, there is a need for cost-effective and innovative methods for identifying PPD patients and identifying future developing tendencies.

To solve this, Osprey Parameter Optimized MLP (OPOMLP) [15] was developed to improve the classification of mental health and behavioral changes in women. Initially, the Application Programming Interface (API) function of online social network like twitter (tweets) and instagram (comments) are used to collect the data posted by health care professionals. Then, the collected data was pre-processed using Natural Language Processing (NLP). The psychological attributes like mental health and behavioral changes attributes of women are extracted using Linguistic Inquiry Word Count (LIWC) and Latent Semantic Analysis (LSA). Next, Multi-Layer Percepton (MLP) network was trained using the extracted attributes along with the attributes of demographic status, mental health history and vital signs. Since, the parameters of MLP were not optimized properly which leads to computational complexities in the PPD prediction. So, the hyper-parameters of MLP is optimized using an Osprey optimization algorithm (OOA) for optimal classification with less time and computational resources. However, this model model does not provide efficient results on larger datasets, culminating in lower accuracy performances.

In this paper, SSPPDnet is proposed which ensembles CNN, LSTM and SSDNN model to process on larger data samples and improve models performances for efficient PPD prediction. This model utilizes the CNN and LSTM to automatically learn all feature types. For any kind of data, LSTM extracts its time-sequence attributes by capturing its temporal properties. 2DCNN receives the feature tensor derived from the temporal features in order to learn the connection among different characteristics. After that, combining the temporal, correlation and demographic features of women yields a single feature vector. These extracted features from 2DCNN and LSTM are given as input to SSDNN as it provides rapid convergence with constant global learning, yielding in a reduced initial learning rate. SSDNN model introduces an additional scalar parameter to stabilize SGD training to mutually trained with the actual network parameters. Number of hidden layers and nodes in each layer with training modules like learning rate and regularization methods are the characteristics presented. These additional parameters are snatched as a per-layer stabiliser with the data to decide whether to increase or lower the previously set values in order to decrease the target operation. FC layers are utilized to calculate postpartum accomplishment levels by combining time-series correlation and women's demographic information. The softmax layer is employed with FC layer for PPD detection and classification. This model effectively processes on the large number of samples and increases the classifier's performances.

The further extension of the paper is stated as follows: Section II covers many models intended to identify PPD using EHR. The suggested PPD model is provided in Section III. The performance assessment of past and proposed models is given in Section IV. The whole investigation is summed up in Section V and recommends for future extensions.

2. Literature Survey

Raisa and colleagues [16] developed a ML method for PPD early detection in Bangladesh. The survey based on socio-demographic questions and EPDS data was collected, pre-processed, and augmented using the synthetic minority oversampling approach (SMOTE), which provides synthetic members based on feature space distances. Finally, SVM, RF, LR and XGBoost were employed for the for PPD detection. But, this model has an inappropriate bias selection and trained with limited data samples.

Prabhashwareem et al. [17] developed a ML model for predicting mothers with PPD. Min-Max normalisation was used to pre-process and normalise the mother's family, social background, and other data-related statuses that were gathered. For the prediction of PPD risk levels, the acquired data was input into Feed-Forward Neural Network (FFANN), Adaptive Neuro-Fuzzy Inference System with Genetic Algorithm (ANFIS - GA), RF, and SVM algorithms. But, insufficient data samples lead to degrade the accuracy performance.

Gopalakrishnan et al. [18] constructed a ML model for the prediction of PPD. Initially, the data was collected and then it was divided into EPDS and Background Information (BI) of pregnancy women. The RF model was employed to select the relevant features. Extremely Randomised Trees (XRT) model was fed the chosen characteristics to anticipate the risk variables and the frequency of PPD indications. However, high class disparity makes it difficult to train the model.

Rousseau et al. [19] devised the possibility of pregnant women experiencing persistent Post-Traumatic Stress Following Childbirth (PTS-FC). This model implicit category analysis was used to classify women's patterns from the gathered information into Stable-High-PTS-FC and Stable-Low-PTS-FC. After that, the DT was used to find women who were at concern of Stable-High PTS-FC for early prediction. Meanwhile, when the data was increased, the complexities would also increase.

Huang et al. [20] developed a prenatal depression and assessing model bias using ML model. The EHR data from a large urban hospital was pre-processed and normalized, then analyzed using Shapley Additive Explanations (SHAP), Disparate Impact (DI), and Equal Opportunity Difference (EOD). The interpreted data was then fed into an elastic net to classify and predict different stages of PPD. But, this model failed to extract the optimal features subset to enhance the accuracy and reduce misclassified instances.

Wakefield and Frasch [21] presented a method for detecting PPD using EHR data. In this method, four ML models were developed using distributed RF and Logistic Regression (LR) models. Model 1 used easily accessible socio-demographic data, Model 2 included mother mental health data before pregnancy, Model 3 used recursive feature removal, Model 4 simplified pre-pregnancy mental health factors. Finally, the LR model was used for the final PPD prediction. But, the model's performance was notably lower due to its training on a limited dataset.

Liu et al. [22] developed an optimization ML model for PPD risk assessment and preventive intervention strategy. This model collected and pre-processed EHR data from cesarean delivery patients, used SHAP for data interpretation, developed Propensity Score Matching (PSM) to compare PPD incidence between training and testing groups and finally XGBoost model was employed for early intervention. However, this model takes long time to train the model.

Gopalakrishnan et al. [23] developed an Attribute Selection Hybrid Network (ASHN) for risk factors analysis of PPD. A post-level attention layer and coupled neural networks was employed to choose features based on the depression hypothesis. Moreover, the state of everyday life and PPD phases were compared using title and word-by-word word clouds visualisation. At last, the several mood disorders associated with PPD were predicted using ASHN. But the little amount of training data input made the performance worse and less comprehensible.

Srivatsav et al., [24] developed a DL based detection of postpartum depression stages on social media through sentiment analysis on data. The feature extraction was performed with the CNN component and the temporal and sequential patterns in textual data were obtained through the LSTM layer. The hybrid LSTM-CNN model was found to be more effective in the prediction of postpartum depression as compared to the RF model, thereby showing the feasibility of DL methods in detecting postpartum depression. However, it need the high volumes of labeled data and greater computing resources which might be limiting in its application to smaller datasets or low resource settings.

Lilhore et al. [25] presented Improved Bi-directional Long Short-Term Memory (IBi-LSTM) and CNN for PPD identification. The preprocessing of the gathered data guarantees data integrity, and feature selection offers the risk variables that have a major influence on PDD prediction. Then, IBi-LSTM integrates written material and sound data to produce complex temporal interpersonal linkages for PPD prediction, while CNN was employed along with TL to extract the essential information from text and audio features. But the outcomes of this strategy include significant information loss and inadequate data filtering.

Lin and Zhou [26] devised a postpartum depression prediction model is an ensemble neural network model that predicts postpartum depression based on clinical questionnaire data. This hypothesized a Fully Connected Neural Network (FCNN) and Deep Neural Network (DNN) with a Dropout methodology to make the model more interpretable and increase the capacity of the model to generalize. Nevertheless, questionnaire-based data used in the study were based on only one single source of data that might restrain extensive clinical generalization.

3. Proposed Methodology

In this part, the complete illustration of SSPPDnet model for early PPD prediction. In this method, the input data, pre-processing and feature extraction strategies are considered from [15]. The extracted features will be fed into CNN-LSTM-SSDNN model for extracting time series features, correlation features and classification. The figure 1 depicts the overview of the proposed model.

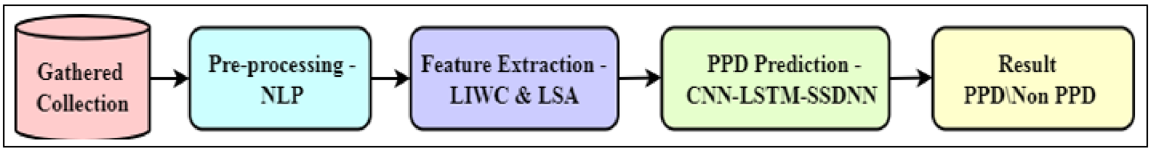

Figure 1: Pipeline of the proposed model

3.1 Data Pre-processing

The attained data is pre-accessed using several NLP approaches in which the collected tweets and comments are combined to provide a comprehensive set of features. The data from Twitter and comments is pre-processed using Natural Language Processing (NLP) techniques to provide a comprehensive set of features. This includes removing emojis, punctuation, stopwords, typescripts, and special characters. The tweet content is tokenized, stemmed, and lemmatized to remove redundant or undesired terms. After pre-processing, online behaviors like emotion, event, activity, user-specific features, and consumption of depression-related texts are extracted to estimate the PPD phases. By removing duplicate data, these methods may reduce the amount of information about PPD instances. Additionally, attributes from online social networks are acquired to train the proposed network model.

3.2 Feature Extraction

Following the collection of the dataset, the feature extraction techniques like LIWC and LSA [15] are used to extract physiological, behavioral and demographic variables related to women's mental health and changes from online Twitter tweets and Instagram comments, as well as medical record aspects of the women's PPD phases.

LIWC is a feature extraction software technique that gathers and classifies language features based on psychological factors. Based on the LIWC lexical dictionary, the proportion of all mental health classes for every woman afflicted with PPD conditions. LIWC attributes are physiological words such as a woman's mental health behavior that varies considerably from mood swings in women the previous day. It identifies phrases that depressed patients often use and that are associated with their psychological actions as a result of their psychological backgrounds (e.g., sentiments, happiness, regrets and sorrow).

LSA is an unsupervised indexing approach used in natural language processing that retrieves semantically linked terms from text sources. This approach incorporates Singular Value Decomposition (SVD) to semantically categorize words and texts, preserving noiseless data for optimal retrieval. The singular value, representing a vector depiction, emphasizes the relevance of pattern in the content. The LSA employs a database with an arbitrary mix of latent themes, where every topic is described by a distribution over terms such as physiological terminology frequently utilized by PPD patients.

3.3 Proposed PPD prediction Model

This method constitutes to four major phases like input (extracted features), (ii) Temporal Feature Mining (LSTM), Correlation Feature Mining (CNN) and output (SSDNN).

3.3.1 Input Data

This framework constitutes different PPD data forms like women mental health, behavioural, demographic and physiological data. Every women data will be deliberated as the time period sequences where all the records initiate to provide a timestamp, but different women have different features. In this case, $A_x = \left(A_{x1}, \dots, A_{xy}, \dots, A_{xn}\right)$ represents the n classes of multi-labelled features expressed for PPD womens x, $A_{xy} = \left[a_{xy}^{1}, \dots, a_{xy}^{T}, \dots, a_{xy}^{t_{xy}}\right]$ is the y^{th} characteristic variable of women with x, a_{xy}^{T} $\left(1 \leq t \leq t_{xy}\right)$ is a matrix significant with initial incident scored data values with the limited time segment t considered as single route registration data in which t_{xy} defines the extension of the y^{th} feature in PPD women x.

3.3.2 LSTM based Temporal Feature Mining

LSTM is used for temporal feature mining, which involves modeling each attribute category individually and learning their temporal features. LSTM employs to process the series of values. Extraction of long-term dependencies from the lengthier sequences may be made more simple with its help. LSTM assists to processes a succession value and effectively reduces the difficulty of extracting long-term connections from lengthier series. LSTM is used to automatically retrieve features from women's numerous qualities gathered from demographic, physiological and behavioural attribute data that are provided in a time-series way. These attribute data are fed to the LSTM to mine temporal characteristics after the normalization of attributes.

3.3.3 2DCNN based Correlation Feature Mining

In order to learn the correlation properties across many attributes, 2DCNN functions as associative feature mining, applying on the tensor converted from the temporal features of every data class. Multiple qualities should correlate as the women in the multi-source performance data are identical. A tensor method is used to transform the temporal feature vector of each attribute into a 3D tensor and an n-dimensional CNN is used to retrieve the association between many features in order to discover the relationship values among several data. This 2DCNN is used to extract image features which is described by a tensor (W, H, C), where W, H and C represent the width, height and channels, accordingly. The temporal feature vectors of Z types are transformed into a 3D tensor, with $W \times H = Z$ and $C = Q$, where Q is the dimension of the temporal feature vector. The associative characteristics are extracted using an n-dimensional CNN based on the tensor matrix.

3.4 Self-Sturdy Deep Neural Network (SSDNN) for classification

The extracted women's demographic attributes, temporal attributes and correlation attributes from the LSTM and 2DCNN are fused and fed into the SSDNN classifier for PPD prediction. In SSDNN, an additional scalar parameter is integrated to each layer of the DNN model. SSDNN incorporates an extra scalar parameter into each layer of the DNN model. This model is intended to stabilize SGC training and is trained simultaneously with the initial network configurations (CNN and LSTM). The new parameters are intercepted as the per-layer stabiliser, with data determining whether to decrease or raise their integers to reduce the desired operations ℓ. In their fundamental layout, the parameter of DNN layers is enhanced using a constant δ.

$$b = \omega(\delta \times \mathcal{W}a + y) \quad (1)$$

In the preceding Eq. (1), an input matrix is denoted by a, a linear translation by $\mathcal{W}$ and y, while an output column is represented as b. Initialising $\delta = 1$, the model is comparable with and without the additional parameter. The additive parameters δ are just another factor that SGD has to be trained throughout the iteration. It is fairly simple to satisfy the current criteria δ. The shift in layer u with regard to input array a is computed in process with the backward pass in Eq. (2),

$$\frac{\partial \ell}{\partial a} = \delta\, \mathcal{W}^{T} \frac{\partial \ell}{\partial b} \quad (2)$$

Eq. 3 determines the diagonal slope with regard to the parameter δ.

$$\frac{\partial \ell}{\partial \delta} = \frac{\partial \ell}{\partial b} \times \frac{\partial b}{\partial \delta} = \frac{\partial \ell^{T}}{\partial b}\, \mathcal{W}a \quad (3)$$

While the inclination for parameter δ may be represented in equation (4),

$$\frac{\partial \ell}{\partial \delta} = \frac{1}{\delta}\left(\frac{\partial \ell}{\partial a}\right)^{T} a = \frac{1}{\delta} < \frac{\partial \ell}{\partial a}, a > \quad (4)$$

So δ is provided in Eq. (5),

$$\delta_{t+1} = \delta_t - \frac{\eta}{\delta} < \frac{\partial \ell}{\partial a}, a > \quad (5)$$

The alteration in δ indicates the connection among the layer input a and the slope of the gradient desired operation with respect to that input values. Primarily, the desired operation will be improvised by enhancing a, in which δ will increase gradually. If the goal function improves with a lower a, δ will reduce. If the input's relative direction and gradient are arbitrary, the value of δ will stabilize. It is marginally anticipated to occur around the convergence rate. As in Eq. (6) the current integer δ directly regulates the stride dimensions for parameter $w_{u,v}$.

$$\mathcal{G}_{u,v} = \frac{\partial \ell}{\partial w_{u,v}} = \delta \frac{\partial \ell}{\partial b_u}\, a_v \quad (6)$$

Practically, the $exp\,(\delta)$ is utilized instead of δ and it is initialized as $\delta = 0$. The effective stabiliser was thus constrained to be affirmative and to diminish more slowly as it got closer to the zero.

$$b = \omega(exp(\delta) \times Wa + y) \qquad (7)$$

According to Eq. (7), the substitution produces similar models without the possibility of the stabilizser parameter modifying the indication that could interfere with DNN training. The women's postpartum achievement stage is produced by merging time-series, correlation and women's demographic data using the FC layers in SSDNN. Lastly, the FC layer uses the Softmax layer to perform the categorization. The Mean Square Error (MSE) is the loss operation stated in Eq. (8):

$$loss == \frac{1}{n}\sum_{u=1}^{N}(v-u)^2 \qquad (8)$$

Here, N is the aggregate integer of women data samples, v denotes an actual probability of women data instances and u is the predicted probability with of u^{th} instances. Moreover, OOA [15] is used to select the features and optimizes the hyperparameter of CNN, LSTM and SSDNN for efficient PPD detection. The figure 2 depicts Block structure of SSPPDnet model for PPD prediction and classification. The below Algorithm provides the complete framework of SSPPDnet.

Algorithm: Proposed SSPPDnet Model

Input: Collected Data (Mental health and behavioral changes attributes of women from hospital records and online tweets and comments)

Output: Process large samples and enhance model performance for PPD prediction

Step 1: The collected data will be pre-processed using NLP to remove the irrelevant data.

Step 2: Then, the pre-processed images will be divided into training and testing stages.

Step 3: LIWC and LSA models are used to extract behavioral, physiological and demographic features.

Step 4: Extracted features are trained using prediction model by combining the LSTM, 2DCNN and SSDNN algorithms.

Step 5: The LSTM network is used to retrieve the time-sequence features from each data diversity.

Step 6: Temporal properties are turned into a feature tensor, which is inputted into 2DCNN to learn the relationship among several variables.

Step 7: Consolidated attribute vector is formed by merging temporal features, association attributes, and women's demographic features.

Step 8: The concentrated attributes will be fed into SSDNN to predict women's PPD.

Step 9: The trained model of SSDNN is used to estimate the desired labels

Step 10: The classified label of test data is validated by verifying actual label of test data

Step 11: By using the corresponding labels, prediction and classification of PPD is achieved.

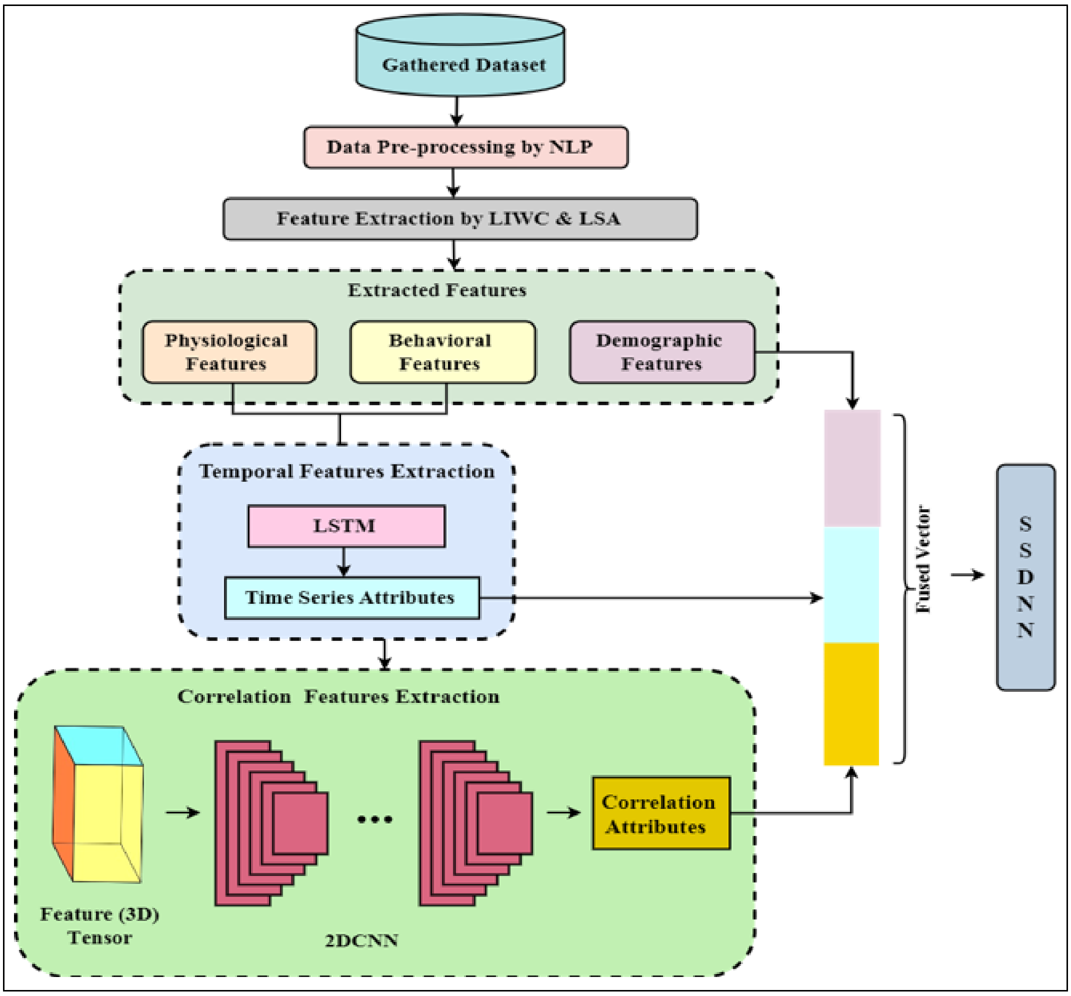

Figure 2: Block Structure of proposed SSPPDnet Model

4. Result and Discussion

4.1 Dataset Description

Social Media Post of Postpartum Depression Dataset (Dataset 1): This dataset [27] consists of social media text from Twitter and Facebook. Using an API scraper (apifi.com), 1589 text data records from Twitter and 1500 from Instagram were collected. Psychological factors are then extracted using word2vec from these social media texts. By using this details, three types of attributes like social media behavioural factors, demographic and psychological factors are gathered. From the collected dataset, 3089 instances from the dataset, with 1853 used for training and 1236 for testing, for a total of 60% and 40%, respectively.

Self-constructed Dataset (Dataset 2): The self-constructed dataset is designed to explore PPD in women during their postnatal phase. The database is built using standardized psychological tests, such as EPDS and PHQ-9 and socio-demographic questions. Women in the postnatal period who were enrolled by obstetrics and gynecology departments of the government and private hospitals in Coimbatore, Tamil Nadu, India who responded to the questions provided were identified as 15,000 women. Women with follow-ups on child delivery and post-natal visits to these departments were invited to participate. The written consent was received by either the hospital staff or researchers based on the institutional ethics approval procedure. This dataset consists of 25 variables that represent emotional state, stress, mood, sleep, appetite, concentration, family and social support, and demographical variables such as age, education, occupation and socioeconomic background which give an extensive picture of clinical and social variables concerning PPD. Furthermore, official APIs of Twitter and Instagram were used to gather 15,000 social media posts. The predictions of sentiment scores of each individual post were made using Word2Vec, and were harmonized with the psychometric data. The social media data set consists of behavioral and emotional measures such as sadness, anxiety, or hopelessness, user specific measures, such as frequency of posting and language usage, online activity trends and event-related references to family stressors or childcare. The combination of these scores on the sentiment and psychometric assessment gives a deeper picture of maternal mental health. Data were gathered according to the specifications of the department and de-anonymized, ensuring the privacy. Reliable data: To prevent redundant records with arbitrary data, records with missing or inconsistent data were deleted and statistical screening conducted to identify possible biases such as demographic or sampling bias. Relevance and authenticity of social media posts were also checked. Such precautions will make sure that the dataset is collected ethically, reliable and capable of training to predict PPD. From the collected dataset, 15,000 instances from the dataset, with 9000 used for training and 6000 for testing, for a total of 60% and 40%, respectively.

4.2 Experimental Setup and Performance Metrics

The implementation of both existing and proposed model was carried out in Python 3.11 and executed on a system with an Intel® CoreTM i5-4210 CPU @ 3GHz, 4GB RAM, and a 1TB HDD running on Windows 10 64-bit with the same datasets which is illustrated in section 4.1. The parameter settings of proposed model are listed in Table 1.

- **Accuracy:** It is determined as the ratio of accurately anticipated samples to the overall number of samples.

$$Accuracy = \frac{True\ Positive\ (TP) + True\ Negative\ (TN)}{TP + TN + False\ Positive\ (FP) + False\ Negative\ (FN)} \tag{9}$$

In Eq. (9), TP refers to the scenario in which the model accurately predicted a positive PPD cases when the actual case was indeed positive. TN refers to the scenario where the model accurately predicted negative PPD cases despite the actual negative sentiment. FP denotes the case where the model inaccurately predicted a positive PPD cases sentiment when the actual cases was negative. FN indicates the scenario where the model inaccurately predicted a negative PPD case when the actual case was positive.

- **Precision:** It evaluates the ratio of accurate TP predictions (positive PPD cases) out of all TP predictions determined by the model.

$$Precision = \frac{TP}{TP + FP} \tag{10}$$

- **Recall:** It calculates the proportion of TP predictions among all TP cases in the dataset.

Table 1: Parameter Settings of the Proposed Framework

Module	Parameter	Value
ASHN [23]	Batch size	80
	Dropout rate	0.1
	Hidden size	93
	Activation function	tanh
IBi-LSTM-CNN [25]	Pooling size	2,2
	Convolution layer	2
	LSTM units	64
	Dense layer	32
Common settings (existing and Proposed model)	Word embedding size	256
	Learning rate	0.1
	Optimizer	SGD
	Epochs	100
	Dropout rate	0.7
	Weight decay	0.0006
	Batch size	96
	Momentum	0.20
	Activation function	ReLU
	Loss function	MSE

$$Recall = \frac{TP}{TP + FN} \tag{11}$$

- **F1-score:** It is the cyclical average of precision and recall.

$$F1 - score = \frac{2 \times Precision \times Recall}{Precision + Recall} \tag{12}$$

- **Area Under the Curve (AUC):** It measures the model's ability to distinguish between normal and PPD affected patterns across different threshold settings. AUC is computed using the Receiver Operating Characteristic

(ROC) curve, which plots the True Positive Rate (TPR) against the False Positive Rate (FPR)

$$TPR = \frac{TP}{TP+FN} \quad (13)$$

$$FPR = \frac{FP}{FP+TN} \quad (14)$$

AUC score is obtained by integrating the ROC curve:

$$AUC = \int_0^1 TPR(FPR)\, d(FPR) \quad (15)$$

AUC score (closer to 1) indicates better model performance showing strong discrimination between PPD and Non-PPD.

4.3 Performance Evaluation on Dataset 1

This section evaluates the performance of SSPPDnet in comparison to other existing models, namely ASHN [23], LSTM-CNN [24], IBi-LSTM-CNN [25], and FCNN-DNN [26] on Dataset 1 which is given in section 4.1. Confusion matrix for Dataset 1 (testing set) on proposed model is given in Figure 3.

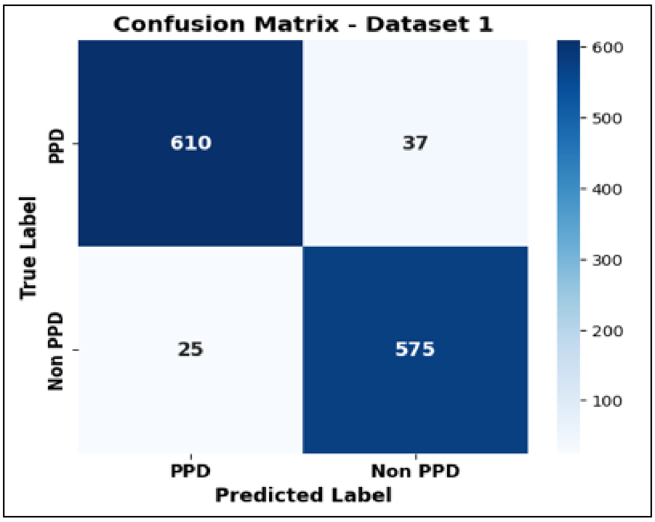

Figure 3: Confusion matrix for Dataset 1 (testing set)

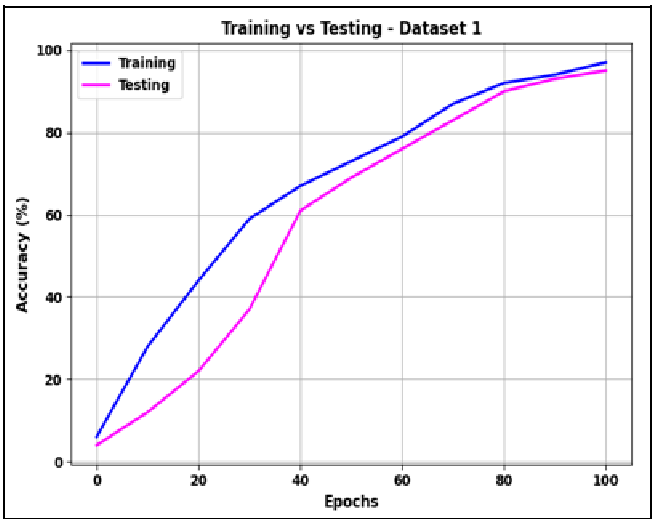

Figure 4: Training vs. Testing Accuracy on Dataset 1

Figure 4 shows the training and testing accuracy for Dataset 1. The training accuracy improves from approximately 44% to 93% between epochs 20 and 100, while the testing accuracy increases from around 22% to 95.03%, demonstrating stable learning and strong generalization capability of the proposed SSPPDnet model.

Figure 5 depicts the overall comparison of the proposed SSPPDnet model with existing models like ASHN, LSTM-CNN, IBi-LSTM-CNN, and FCNN-DNN on Dataset 1. From the above analysis, it is determined that the accuracy of SSPPDnet achieves higher performance compared to other classification models. The accuracy of SSPPDnet is 13.6%, 9.92%, 6.48%, and 3.17% higher than those existing models, respectively. Similarly, the precision of SSPPDnet is

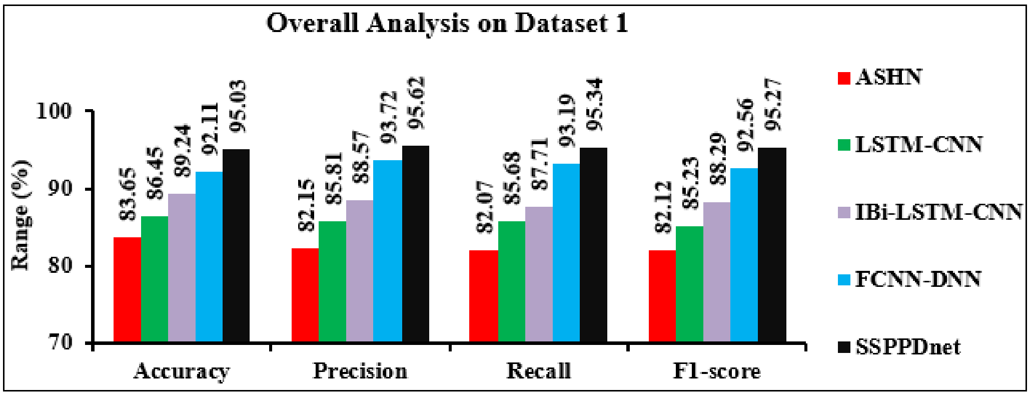

Figure 5: Overall Comparison of Proposed Model with Existing Models on Dataset 1

16.39%, 11.43%, 7.95%, and 2.02%, recall is 16.16%, 11.27%, 8.69%, and 2.3%, and F1-score is 16.01%, 11.77%, 7.9%, and 2.92% higher than those existing models, respectively.

4.4 Performance Evaluation on Dataset 2

This section evaluates the performance of SSPPDnet in comparison to other existing models, namely ASHN [23], LSTM-CNN [24], IBi-LSTM-CNN [25], and FCNN-DNN [26] on Dataset 2 which is given in section 4.1. Confusion matrix for Dataset 2 (testing set) on proposed model is given in Figure 6.

Figure 7 shows the training and testing accuracy for Dataset 2. The training accuracy improves from approximately 45% to 99% between epochs 20 and 100, while the testing accuracy increases from around 22% to 97.63%, demonstrating stable learning and strong generalization capability of the proposed SSPPDnet model.

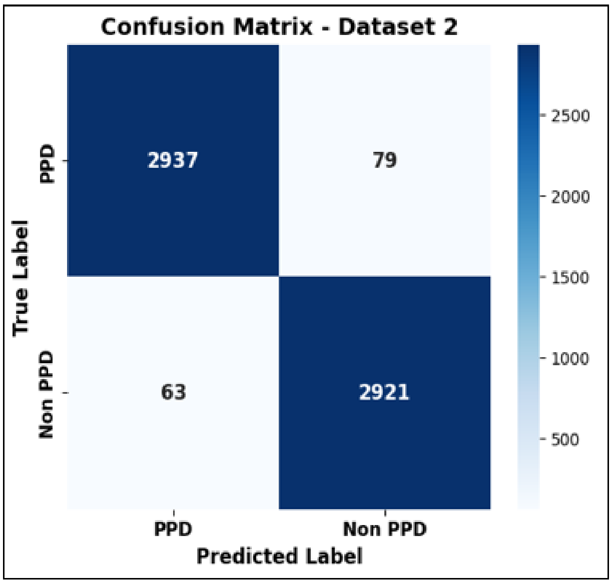

Figure 6: Confusion matrix for Dataset 2 (testing set)

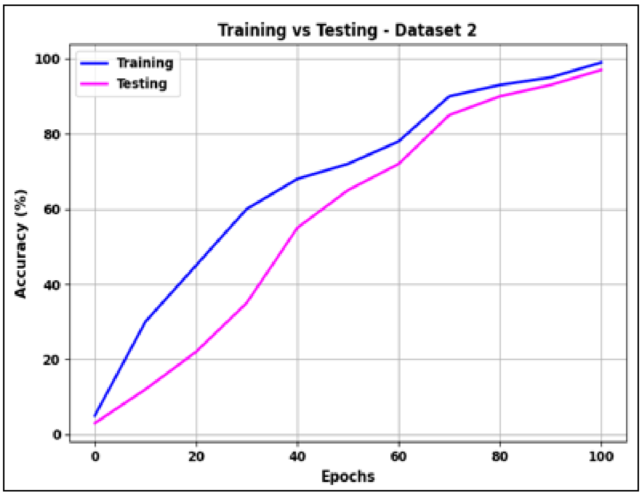

Figure 7: Training vs. Testing Accuracy on Dataset 2

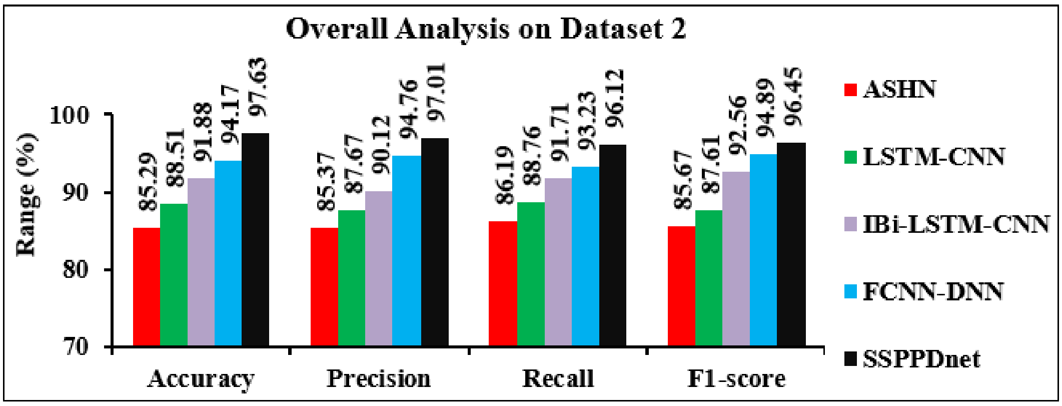

Figure 8: Overall Comparison of Proposed Model with Existing Models on Dataset 2

Figure 8 depicts the overall comparison of the proposed SSPPDnet model with existing models like ASHN, LSTM-CNN, IBi-LSTM-CNN, and FCNN-DNN on Dataset 2. From the above analysis, it is determined that the accuracy of SSPPDnet achieves higher performance compared to other classification models. The accuracy of SSPPDnet is 14.46%, 10.3%, 6.25%, and 3.67% higher than those existing models, respectively. Similarly, the precision of SSPPDnet is 13.63%, 10.65%, 7.64%, and 2.37%, recall is 11.52%, 8.29%, 4.8%, and 3.09%, and F1-score is 12.58%, 10.09%, 4.2%, and 1.64% higher than those existing models, respectively.

4.5 Ablation Study

Table 3 depicts the Ablation Study for the proposed model on Dataset 1 and Dataset 2. The ablation experiment of the proposed SSPPDnet model using dataset 1 and dataset 2. The results indicate a definite and steady growth in performance with the introduction of more modules into the model in both datasets. The performance is lowest with the baseline configuration without 2DCNN that is improved by using SSDNN (improving to around 88.12% and 87.24% with Dataset 1 and 2) and LSTM (improving to around 91.72% and 90.06% with Dataset 1 and 2) confirming the usefulness of progressive feature enhancement. Combined modules, including without 2DCNN

Table 3: Ablation Study for the proposed model

Model	Accuracy (%)	Precision (%)	Recall (%)	F1 Score (%)
Dataset 1				
Without 2DCNN	85.09	85.66	84.25	84.86
Without SSDNN	88.12	88.77	88.97	88.65
Without LSTM	91.72	91.87	90.38	90.76
Without 2DCNN + LSTM	93.41	93.55	93.86	93.68
Without SSDNN + LSTM	94.35	94.21	94.67	94.16
Proposed SSPPDnet model	95.03	95.62	95.34	95.27
Dataset 2				
Without 2DCNN	84.23	84.32	84.59	84.11
Without SSDNN	87.24	87.66	87.12	87.89
Without LSTM	90.06	90.17	90.49	90.34
Without 2DCNN + LSTM	93.55	93.77	93.24	93.16

Without SSDNN + LSTM	95.43	95.07	94.81	94.32
Proposed SSPPDnet model	97.63	97.01	96.12	96.45

+ LSTM (~93.41% and 93.55%), and without SSDNN + LSTM (~94.35% and 95.43%), also enhance the performance. The proposed SSPDnet model provides the best performance with 95.03 and 97.63 on dataset 1 and dataset 2, respectively. This shows that it has better feature extraction and generalization. In general, the proposed model achieves much higher gains in all measures compared to all the variants that have been ablated.

5. Conclusion

In this article, SSPPDnet is developed to process the large number of samples and enhance the model's performances for the efficient PPD prediction. In this method, CNN converts extracted features into feature vectors, while LSTM captures data temporal characteristics and determines time-series features for each data type. These features are fed into SSDNN which is designed for faster convergence and lower initial learning rate training. SSDNN introduces an additional scalar parameter to stabilize SGD training, which is then intercepted as a per-layer stabilizer. FC layers are used to calculate postpartum accomplishment levels, with the softmax layer used for PPD detection and classification. Finally, test results show that the SSPPDnet model obtains an accuracy of 95.03% and 97.63% on datasets respectively outperforming other models.

References

[1] R. P. Upadhyay, R. Chowdhury, A. Salehi, K. Sarkar, S. K. Singh, B. Sinha, A. Kumar, Postpartum depression in India: a systematic review and meta-analysis, Bulletin of the World Health Organization, 95(10) (2017) 706. https://doi.org/10.2471/BLT.17.192237

[2] R. Chow, E. Huang, A. Li, S. Li, S. Y. Fu, J. S. Son, W. G. Foster, Appraisal of systematic reviews on interventions for postpartum depression: systematic review, BMC pregnancy and childbirth, 21 (2021) 1-11. https://doi.org/10.1186/s12884-020-03496-5

[3] K. Ayoub, A. Shaheen, S. Hajat, Postpartum depression in the Arab region: A systematic literature review, Clinical practice and epidemiology in mental health: CP & EMH, 16(Suppl-1), (2020) 142. https://doi.org/10.2174/1745017902016010142

[4] D. Wildali, S. Nazzal, S. Hamshari, S. Belkebir, Prevalence and risk factors of postpartum depression among women attending primary healthcare centers in northern of West Bank/Palestine: a cross-sectional study, 2022, BMC women's health, 24(1) (2024) 43. https://doi.org/10.1186/s12905-024-02887-6

[5] M. M. McKelvey, J. Espelin, Postpartum depression: Beyond the "baby blues", Nursing Made Incredibly Easy, 16(3) (2018) 28-35. https://doi.org/10.1097/01.NME.0000531872.48283.ab

[6] V. Patel, A. Rahman, K. S. Jacob, M. Hughes, Effect of maternal mental health on infant growth in low income countries: new evidence from South Asia, Bmj, 328(7443) (2004) 820-823. https://doi.org/10.1136/bmj.328.7443.820

[7] A. J. Bjertrup, M. S. Væver, K. W. Miskowiak, Prediction of postpartum depression with an online neurocognitive risk screening tool for pregnant women, European Neuropsychopharmacology, 73 (2023) 36-47. https://doi.org/10.1016/j.euroneuro.2023.04.014

[8] X. H. Zhao, Z. H. Zhang, Risk factors for postpartum depression: An evidence-based systematic review of systematic reviews and meta-analyses, Asian journal of psychiatry, 53 (2020) 102353. https://doi.org/10.1016/j.ajp.2020.102353

[9] G. Amit, I. Girshovitz, K. Marcus, Y. Zhang, J. Pathak, V. Bar, P. Akiva, Estimation of postpartum depression risk from electronic health records using machine learning, BMC Pregnancy and Childbirth, 21 (2021) 1-10. https://doi.org/10.1186/s12884-021-04087-8

[10] S. Wang, J. Pathak, Y. Zhang, Using electronic health records and machine learning to predict postpartum depression, In MEDINFO 2019: Health and Wellbeing e-Networks for All (2019) 888-892 IOS Press. https://doi.org/10.3233/SHTI190351

[11] W. Zhang, H. Liu, V. M. B. Silenzio, P. Qiu, W. Gong, Machine learning models for the prediction of postpartum depression: application and comparison based on a cohort study, JMIR medical informatics, 8(4) (2020) e15516. https://doi.org/10.2196/15516

[12] P. Cellini, A. Pigoni, G. Delvecchio, C. Moltrasio, P. Brambilla, Machine learning in the prediction of postpartum depression: A review, Journal of Affective Disorders, 309 (2022) 350-357. https://doi.org/10.1016/j.jad.2022.04.093

[13] I. Fatima, B. U. D. Abbasi, S. Khan, M. Al-Saeed, H. F. Ahmad, R. Mumtaz, Prediction of postpartum depression using machine learning techniques from social media text, Expert Systems, 36(4) (2019) e12409. https://doi.org/10.1111/exsy.12409

[14] Y. Zhang, S. Wang, A. Hermann, R. Joly, J. Pathak, Development and validation of a machine learning algorithm for predicting the risk of postpartum depression among pregnant women, Journal of affective disorders, 279 (2021) 1-8. https://doi.org/10.1016/j.jad.2020.09.113

[15] D. Suganthi, A. Geetha, Predicting Postpartum Depression with Aid of Social Media Texts Using Optimized Machine Learning Model, International Journal of Intelligent Engineering & Systems, 17(3) (2024). https://doi.org/10.22266/ijies2024.0630.33

[16] J. F. Raisa, M. S. Kaiser, M. Mahmud, A machine learning approach for early detection of postpartum depression in Bangladesh, In International Conference on Brain Informatics, (2022, July) 241-252. Cham: Springer International Publishing. https://doi.org/10.1007/978-3-031-15037-1_20

[17] T. H. K. R. Prabhashwaree, N. M. Wagarachchi, Predicting Mothers with Postpartum Depression using Machine Learning Approaches, In 2022 International Research Conference on Smart Computing and Systems Engineering (SCSE), 5 (2022, September) 28-34 IEEE. https://doi.org/10.1109/SCSE56529.2022.9905186

[18] A. Gopalakrishnan, R. Venkataraman, R. Gururajan, X. Zhou, G. Zhu, Predicting Women with Postpartum Depression Symptoms Using Machine Learning Techniques, Mathematics, 10(23) (2022) 4570. https://doi.org/10.3390/math10234570

[19] S. Rousseau, I. S. Polachek, T. I. Frenkel, A machine learning approach to identifying pregnant women's risk for persistent post-traumatic stress following childbirth, Journal of Affective Disorders, 296 (2022) 136-149. https://doi.org/10.1016/j.jad.2021.09.014
[20] Y. Huang, S. Alvernaz, S. J. Kim, P. Maki, Y. Dai, B. P. Bernabé, Predicting prenatal depression and assessing model bias using machine learning models, Biological Psychiatry Global Open Science, 4(6) (2024) 100376. https://doi.org/10.1016/j.bpsgos.2024.100376
[21] C. Wakefield, M. G. Frasch, Predicting patients requiring treatment for depression in the postpartum period using common electronic medical record data available antepartum, AJPM focus, 2(3) (2023) 100100. https://doi.org/10.1016/j.focus.2023.100100
[22] H. Liu, A. Dai, Z. Zhou, X. Xu, K. Gao, Q. Li, ... S. Wang, An optimization for postpartum depression risk assessment and preventive intervention strategy based machine learning approaches, Journal of Affective Disorders, 328 (2023) 163-174. https://doi.org/10.1016/j.jad.2023.02.028
[23] A. Gopalakrishnan, R. Gururajan, R. Venkataraman, X. Zhou, K. C. Ching, A. Saravanan, M. Sen, Attribute Selection Hybrid Network Model for risk factors analysis of postpartum depression using Social media, Brain Informatics, 10(1) (2023) 28. https://doi.org/10.1186/s40708-023-00206-7
[24] P. Srivatsav, S. Nanthini, Detecting Postpartum Depression Stages in New Mothers: A Comparative Study of Novel LSTM-CNN vs. Random Forest, SPAST Reports, 1(3) (2024). https://doi.org/10.5220/0012569700003739
[25] U. K. Lilhore, S. Dalal, N. Varshney, Y. K. Sharma, K. B. Rao, V. M. Rao, ... P. Chakrabarti, Prevalence and risk factors analysis of postpartum depression at early stage using hybrid deep learning model, Scientific Reports, 14(1) (2024) 4533. https://doi.org/10.1038/s41598-024-54927-8
[26] Y. Lin, D. Zhou, A method for predicting postpartum depression via an ensemble neural network model, Frontiers in Public Health, 13 (2025) 1571522. https://doi.org/10.3389/fpubh.2025.1571522
[27] Dataset: Social Media Post of PostPartum Depression, Kaggle (2025), available online: https://www.kaggle.com/datasets/suganthidevasagayam/social-media-post-of-postpartum-depression