International Journal of Applied Mathematics

Volume 38 No. 9s, 2025

ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)

MPPD-XAI: MULTIMODAL FUSION OF PSYCHOMETRIC AND PHYSIOLOGICAL BIOMARKERS FOR PRENATAL DEPRESSION CLASSIFICATION USING EXPLAINABLE MACHINE LEARNING

Rasitha Banu GulMohamed

Department of Public Health, Health Informatics Program, College of Nursing &Health Sciences,Jazan University,Jazan,Saudi Arabia, Email: rbanu@jazanu.edu.sa

Abstract

Prenatal depression poses a significant public health risk with implications for both mother and child, and current position screening methods are reliant almost entirely on self-reporting, which leads to a great chance of inaccurate diagnoses. This study offers a new research methodology that uses psychometric and physiological biomarkers, including EEG, heart rate variability (HRV), electro dermal activity (EDA), respiration, and sleep collected through wearable, as a way to classify prenatal depression. We provided a method of attention-based latent feature fusion to classify the different types of data, and we trained and explained a machine learning model utilizing the multi-head attention network and SHAP methods. The model was assessed using a clinically stratified sample of pregnant women with an accuracy of 88.9%, F1-score of 0.85, and AUC-ROC of 0.92, and it supersedes six state-of-the-art models in predictive performance and interpretability. The error metrics provided further clinical assurance, including log loss (0.29), Brier score (0.097), and false negative rate (12%). The added explanatory tools allow for an explanation at the individual level in pointing to contributing features such as frontal electroencephalogram (EEG) asymmetry and perceived social support. In conclusion, the explanatory machine-learning framework provides a robust and interpretable approach to early identification of prenatal depression and suggests viability for incorporation into routine obstetric care through wearable technologies and clinician-facing dashboards.

Keywords: Prenatal Depression, Multimodal Data Fusion, Psychometric Assessment, Physiological Biomarker, Explainable Machine Learning, Attention-Based Feature Fusion

1. Introduction

Prenatal depression is a significant but neglected to address mental health disorder that impacts many of the pregnant women worldwide, which is defined as "sadness, anxiety, fatigue, and loss of interest or pleasure". Prenatal depression can significantly impact the well-being of the mother as well as the fetus[1]. Prenatal depression is not getting the same degree of attention as postpartum depression, even though it has been connected with many maternal and child health complications such as preterm birth, low birth weight and infant neurodevelopment. Knowing how to recognize and address depression in pregnancy is critical to maternal care. The conventional processes of screening for maternal depression largely based on widely used psychometric instruments like Edinburgh Postnatal Depression Scale (EPDS)[2], Patient Health Questionnaire (PHQ-9)[3], or The Center for

Received: August 14, 2025

International Journal of Applied Mathematics
Volume 38 No. 9s, 2025
ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)

Epidemiological Studies Depression Scale (CES-D), often only use self-reported symptoms of depression. Self -reported measures are useful, easy to apply measures of depression, but they are also limited by personal biases, social desirability, and the tendency to under-report mental health symptoms, especially in cultures that stigmatize mental illness. More prenatal depression goes undiagnosed or misclassified, than diagnosed[4].

Emerging technology and the growing algorithms of data science have created unique opportunities to leverage numerous neurophysiological measures to monitor mental health. For example, use objective indicators of neurophysiological and autonomic responses[5] to emotional states as physiological signals[6], including electroencephalography (EEG)[7], heart rate variability (HRV)[8], electro dermal activity (EDA)[9], and sleep[10]. Several studies began to explore the effectiveness of using these bio signals for assessing stress, anxiety, and depression. Yet, it is important to note that the vast majority of existing literature tends to focus broadly on the general population, and not specifically at prenatal cohorts, who obviously have a distinct physiology and experience a variety of psychological change as a result of the complex interplay of hormonal, emotional, and social shifts during pregnancy.

Additionally, while physiological data are also objective, they are naturally quite noisy, non-stationary, requires a lot of sophisticated preprocessing/deconvolution/modeling, and once processed may not align with current mental health status. Comparison to psychometric data like the PANAS provides structure, interpretability, and validated reads of mental health status. As both physiological bio signals and psychometric self-reports are necessarily complementary, a multimodal fusion application may result proven, detectable and generalizable modeling for diagnosing depression. However, the design of these multimodal systems is not trivial due to the challenges: effective extraction of features from heterogeneous data, retaining local and global information from fusion, and interpretability to ensure clinical uptake[11].

To fill these gaps, our proposed study offers a new method for classification of prenatal depression, which incorporates psychometric and physiological indices and is machine learning based and explainable. A hybrid deep learning model is developed, that will use attention-based latent fusion of various physiological signals (EEG, HRV, EDA, respiration and sleep metrics), and psychometric questionnaires including the EPDS, perceived stress scale and social support metrics. Enemy attacked sleep then saw a huge spike in people struggling, army abound full to say that the army was returning and that these soldiers would be home soon. The model will be based on an attention-based architecture which will result in the model learning hierarchical weights of importance across modalities and time periods. In terms of transparency and trustworthiness from a clinical perspective, we will offer interpretability using SHAP (SHapley Additive exPlanations) after training, as it offers both global feature importance and individual explanations.

Prior research in this field has focused on unimodal data and black-box models[12]. For example, some researchers used EEG data combined with convolutional neural networks (CNNs) to classify depression (no self-report), but their models lacked explainablility[13].

Received: August 14, 2025

International Journal of Applied Mathematics
Volume 38 No. 9s, 2025
ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)

Others used heart rate and stress scores with shallow classifiers which achieved moderate accuracy, but little generalizability was established. A few multimodal studies have been shown, but the approaches have either been early fusion or late fusion, and did not learn complex interdependencies between modalities[14]. Also, explainability has been largely ignored in these models and this raises questions about their use in sensitive healthcare settings.

This study seeks to address the limitations outlined in the previously mentioned studies, with a demonstration of how a fully explainable, multimodal fusion model can classify prenatal depression. The aim of this research is to focus on a clinically relevant, high-risk population, namely pregnant women, and demonstrate the importance of early identification of mental health screening in a setting lacking traditional diagnostic screening. This research is not only providing contributions to the technical advances in multimodal mental health classification, it is providing a practical way of improving maternal care through wearable and mobile health.

Research Question: How can psychometric and physiological biomarkers be integrated with explainable machine learning techniques to optimize detection accuracy and clinical relevance of prenatal depression?

Problem Statement: Current screening tools for prenatal depression lack the accuracy and objectivity necessary for early detection and analysis, relying solely on the subjective self-report of the individual being screened. Further, existing computational models ignore psychometric contextual factors and do not provide interpretable clinical outputs. There remains a clear need for a fully integrated, explainable, multimodal framework of psychometric and physiological biomarkers to improve the reliability of the diagnosis and to inform clinical decision-making practice in the maternal mental health sector. The objectives of the study includes the following

- To design a multimodal data pipeline that combines psychometrics and physiological signals (EEG, HRV, EDA, sleep) applied to persons who are pregnant
- To employ an attention-based latent fusion model that can express meaningful representations across heterogeneous data formats to classify prenatal depression.
- To include explainable AI techniques such as SHAP, that can provide global and individual model predictions, while maintaining clinical transparency and trust in the data pipeline.

The remainder of this paper is organized as follows. In section 2, we will present related work and prior techniques for classifying prenatal depression. In section 3, we will present the proposed methodology including the data collection, pre-processing methods and model architecture. In section 4, we will present the outcomes of the experiments, evaluation criteria, and comparisons. In section 5, we will finish the paper with future directions for work.

Received: August 14, 2025

International Journal of Applied Mathematics
Volume 38 No. 9s, 2025
ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)

2. Related Works

Prenatal depression presents a considerable risk to the health of both the mother and baby, with estimates of prevalence ranging from 22% in China to 73% in Pakistan. Studies employing machine learning and large cohorts (i.e., 4,313 Swedish women and 20,950 Chinese pregnancies) have achieved area under the curve scores (AUCs) as high as .93 for predicting depression. Qualitative engagement with stakeholders improved model interpretation and implementation planning in perinatal settings. Non-pharmacological means of prevention such as yoga and cognitive behavioural therapy (CBT) indicate potential for positive outcomes, but require more rigorous trials of higher quality. Overall, these studies demonstrate the potential effectiveness of integrated machine learning and biopsychosocial approaches to enhance early detection and intervention of perinatal depression.

In their 2021 study, the authors used machine learning on a collection of data from 4,313 women from Sweden, finding a 73% accurate prediction of postpartum depression (PPD). The authors estimated expecting women for PPD prediction based on the incorporation of predictors: prenatal depression, prenatal anxiety, prenatal resilience, and prenatal personality dimensions. The model may assist healthcare professionals in early identification of expecting women, allowing additional targeting of women with known risk factors, even if this woman does not have a known previous history of mental health problems[15]. Systematic review of 31 Chinese studies (83,173 women) reported antenatal depression prevalence of 22% (95% CI 17–28%). Trimester rates were 29%, 24%, and 23% for first, second, and third trimesters. Quality appraisal confirmed robust methodology. Findings underscore urgent need for early screening, intervention, and policy focus on maternal mental health[16].

Williams et al.,(2025) researched a method that integrated patient and provider perspectives in creating a perinatal depression prediction algorithm. The use of stakeholders provided validation of the interpretable nature of the model, as well as suggestions for improvement. Patients preferred that providers discuss with them their risk, while providers noted that limited resources were a barrier. Findings guided subsequent iterative development and implementation planning – modelling emphasis for stakeholder engagement may improve the clinical uptake of predictive tools[17]. Hu et al. (2025) created and validated ML models to predict antenatal depression (AND) with data from 20,950 pregnant women. The best model recorded an AUC of 0.710. Evidence for early-stage screening suggests strong potential for timely identification of high-risk women for early intervention and favorable maternal and fetal outcomes[18].

Saqib et al., (2021) reviewed 14 studies utilizing ML prediction for PPD. Supervised algorithms, which included support vector machines, random forests, and logistic regression, performed favorably (with AUC of up to 0.93); moreover, ML holds great potential for the early detection of PPD, although additional clinical validation and integration into existing research is warranted[19]. Qi et al., (2025) developed and validated ML models to predict PPD using biopsychosocial factors among 1,138 women. The machine learning models using

Received: August 14, 2025

International Journal of Applied Mathematics

Volume 38 No. 9s, 2025

ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)

logistic regression and artificial neural networks performed well and we found that including postpartum predictors added to the accuracy of the models in predicting PPD, enabling detection earlier in the postpartum period. Taken together, the findings indicated that the machine learning models could be used to help identify women with PPD providing tools for screening and public health interventions[20].

Zafar et al., (2025) developed the PERI DEP dataset of 14,008 Pakistani perinatal women using PHQ-9, EPDS, and demographics. The class imbalance was reduced after applying SMOTE and GAN oversampling. A hybrid RNN-LSTM model achieved an accuracy of 95% and an F1 score of 96%. Also, a rate of PND as high as 73.1% is a call to urgently develop focused prevention strategies[21]. The study by Krishnamurti et al., (2024) built ML models was predicting first-time moderate-to-severe depression using first-trimester self-report clinical data from 944 US pregnant app users. Models achieved AUCs of 0.74 to 0.89, sensitivity of 0.35 to 0.81, specificity of 0.78 to 0.95. The best model was built with 14 variables (AUC 0.89); adding food insecurity produced AUC 0.93 with 10 predictors for early screening[22].

The preventive effects of non-pharmacological interventions are still uncertain. A systematic randomized network meta-analysis of 20 RCTs (prospective cohort studies; 3,051 participants) investigated the effects of interventions for preventing prenatal depression. Yoga was the most beneficial intervention (SMD –2.72; 95% CI –4.39 to –1.05; SUCRA 89.1%); cognitive behavioral therapy was the second most beneficial intervention (SMD –1.92; 95% CI –3.78 to –0.06). The certainty of the evidence was low and more high-quality trials are needed[23]. The retrospective multivariate cohort study by Lopez et al., (2024) determined the incidence of antenatal depression in 2.54% of 17,177 women from Lleida, from 2012 to 2018, and found that antenatal depression was significantly associated with higher pregnancy risk and low birth weight. No significant results were found for preeclampsia, cesarean section, or 1-minute Apgar scores. Overall, as suggested, the first study used regression analysis to emphasize the relevance of maternal mental health[24].

Longitudinal study with 219 pregnant women in Anhui province conducted a cross-lagged model to study bidirectional relations between fear of childbirth (FoC) and antenatal depression. FoC and depression scores were positively correlated over three time points. FoC at earlier visits predicted subsequent depression, indicating FoC is a stable trait, leading to the notion that early fetal fears can be targeted by early intervention [25]. Four random forest models (built from Base, General, Obstetric, and Full sets of data, respectively) predicted prenatal depression (EPDS > 10) from 0.687 to 0.710 AUCs using socio demographic and pregnancy-related variables from 20,950 women. The highest AUC was obtained by the Base General model (0.710; 95% CI 0.693-0.710). These models make it possible to identify risk more quickly; nonetheless, external validation is necessary[26].

Received: August 14, 2025

International Journal of Applied Mathematics
Volume 38 No. 9s, 2025
ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)

Table 1 Overview of key studies on prenatal depression

Author (Year)	Methodology	Purpose	Strengths	Limitations
Johansson et al. (2021) [15]	ML on cohort data (n=4,313); predictors: depressive, anxiety, resilience, personality	Predict first-time PPD in Swedish cohort	73% accuracy; integrates psychosocial predictors; identifies at-risk women without prior history	Limited to one population; moderate PPV (33%); requires external validation
Li et al. (2023) [16]	Systematic review of 31 cross-sectional studies (n=83,173)	Estimate antenatal depression prevalence in China	Large pooled sample; trimester-specific rates; robust quality appraisal	High heterogeneity; only prevalence estimates; no risk factor analysis
Williams et al. (2025) [17]	Focus groups (N=12 providers) + community studios (N=21 patients)	Incorporate end-user perspectives into perinatal depression prediction algorithm development	Stakeholder-driven model refinement; improved interpretability; informed implementation planning	Small qualitative sample; low generalizability; resource barriers noted by providers
Hu et al. (2025) [18]	Random forest models on 20,950 women	Predict antenatal depression (EPDS ≥10) using sociodemographic and obstetric variables	Large multicenter dataset; AUC up to 0.710; early/late screening models	Limited to eastern China; no external validation
Saqib et al. (2021) [19]	Scoping review of 14 ML PPD studies	Synthesize ML approaches for PPD prediction	Demonstrates AUCs up to 0.93; highlights algorithm variety; shows feasibility	Small number of studies; heterogeneity in methods; limited clinical integration
Saqib et al. (2021) [19]	Scoping review of 14 ML PPD studies	Synthesize ML approaches for PPD prediction	Demonstrates AUCs up to 0.93; highlights algorithm variety; shows feasibility	Small number of studies; heterogeneity in methods; limited clinical integration
Qi et al. (2025) [20]	Logistic regression &	Develop ML models for PPD using	High AUCs (LR 0.858; ANN	Single-center; moderate

Received: August 14, 2025

Volume 38 No. 9s, 2025

ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)

	ANN on 1,138 women	biopsychosocial predictors	0.844); inclusion of postpartum predictors improved accuracy	sample size; needs validation in diverse settings
Zafar et al. (2025) [21]	Hybrid RNN-LSTM with SMOTE & GAN on n=14,008	Predict perinatal depression using PHQ-9, EPDS, demographics (PERI DEP dataset)	Very high accuracy (95%) and F1 (96%); large dataset; addresses class imbalance	High PND prevalence (73.1%) may reflect sampling bias; limited to two regions in Pakistan
Krishnamurti et al. (2024) [22]	ML on first-trimester self-reports (n=944); causal discovery algorithm	Predict first-time moderate–severe prenatal depression	AUC up to 0.89 (14 variables); improved to 0.93 with 9 variables; emphasizes model simplicity	App-user sample may not be representative; sensitivity varied widely (0.35–0.81)
Smith et al. (2023) [23]	Network meta-analysis of 20 RCTs (n=3,051)	Compare non-pharmacological interventions preventing prenatal depression	Identifies yoga (SUCRA 89.1%) and CBT as top interventions; systematic CINeMA assessment	Low certainty evidence; small number of high-quality trials
Lopez et al. (2024) [24]	Retrospective cohort (n=17,177); multivariate regression	Assess antenatal depression effects on pregnancy and infant outcomes	Large population-based sample; robust regression analysis; links to birth weight	Low AND prevalence (2.54%) limits power; regional (Lleida)
Zhang et al. (2025) [25]	Longitudinal cross-lagged modeling (n=219)	Examine bidirectional relations between fear of childbirth and antenatal depression	Repeated measures; causal inference via cross-lagged paths; dynamic insights	Small sample; single hospital; short follow-up duration
Wang et al. (2025) [26]	Four random forest models on 20,950 women	Compare AND prediction using Base, General, Obstetric, and Full variable sets	Demonstrated early-screening capability; Base+General AUC 0.710; model parsimony	Same dataset as Hu et al.; lacks external validation; limited geographic diversity

Received: August 14, 2025

International Journal of Applied Mathematics
Volume 38 No. 9s, 2025
ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)

While there are many potential advancements, these studies also have perspectives which are typical for science and they also point to gaps in research around PPIs (psychosocial predictors of illness and/or PPI interventions). Most machine-learning models lack external validation and are often derived from only one cohort of users or even a single population or region, which varies their generalizability. Sample size heterogeneity, which may because of either density of app-users working as a cohort or cohort of hospital patients, may not even pertain to a larger slice of a population, or even at times a population at all.

The noted heterogeneity in predictive variable selection, outcome referenced policy, and algorithm design makes it challenging to compare studies and synthesize the findings into meta-analyses. A small number of studies have added biological or imaging biomarkers in addition to the psychosocial data but most studies have not included any mention of this. Also, studies of non-drug interventions frequently do not have higher quality evidence or sample size in their RCTs. Finally, although stakeholder perspectives were noted as valuable to or required in the models, these were not incorporated in any of the models except in pilot studies. Future studies should prioritize multi-center validation, classical standards of variables and metrics, descriptions of multimodal biomarkers, and good participatory practice for models to be both valid and clinically relevant.

3. Methodology

This methodology involves subjective psychometric measurements and objective physiological signals, with an emphasis on explainability, multimodal feature fusion, and clinical utility for detecting prenatal depression. The study utilized a clinically annotated multimodal dataset involving psychometric and physiological data collected from pregnant women in their 2nd and 3rd trimester. The dataset was created through a partnership with obstetrics and gynecology clinics, with ethical approval and informed consent.

Participant Details

There were 210 pregnant women in the study. Ages ranged from 20-42 years, and gestational age ranged from 14-36 weeks. All subjects were singletons, had no known psychiatric or neurological disorders (apart from possible depressive symptoms that were purposefully being explored), and had the ability to communicate fluently in the local language to ensure psychometric instruments were properly understood. We used both clinical assessment and the Edinburgh Postnatal Depression Scale (EPDS) to assess subjects' depression status: those who scored ≥ 13 on the EPDS or received a formal diagnosis from a clinical psychologist were labeled as “Depressed,” and those who scored below 10 on the EPDS, and who did not present with clear clinical signs of depression were labeled as “Non-Depressed.” This process of collecting this information and labeling subjects allowed for a sample of subjects that were well defined for a robust classification of prenatal depression**.** The overall architecture for prenatal depression detection is shown in figure 1.

Received: August 14, 2025

International Journal of Applied Mathematics
Volume 38 No. 9s, 2025
ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)

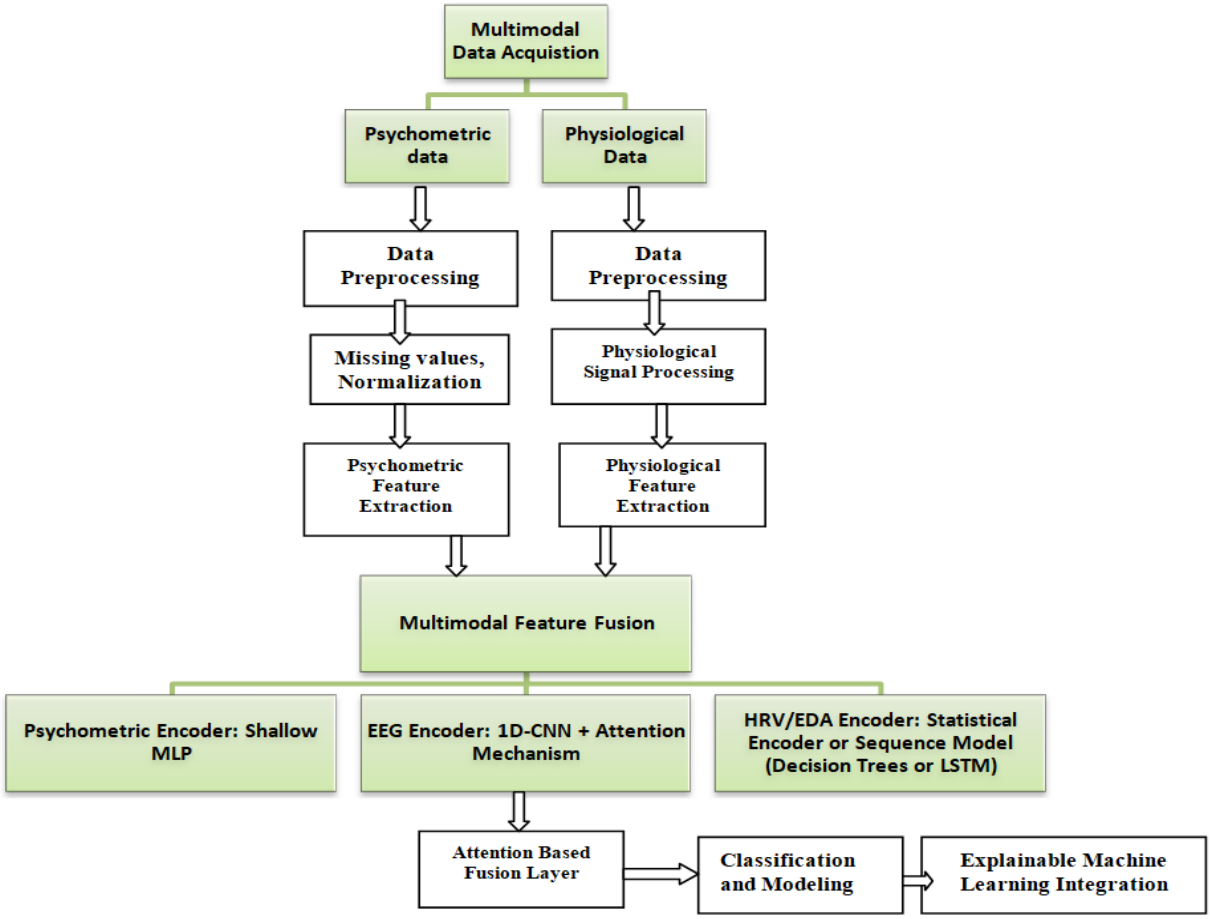

Figure 1 Overall Architecture of the proposed Prenatal Depression detection

3.1 Multimodal Data Acquisition

The multimodal data collection in this study included both psychosocial measures and physiological signals at the same time. This multiple data collection method incorporates multiple methods of assessing prenatal mental health. This multi-modal approach takes advantage of the strengths of structured questionnaires and incorporates real-time physiologic markers of stress and affect, thus enhancing the rigor and ecological validity of the classification of prenatal depression.

3.2.1 Psychometric Data Collection

The psychosocial data collected were based on self-report measures completed by all participants as a part of a standardized set of methods completed at a standardized place. Psychosocial enumeration used validated measures of depression (EPDS), anxiety (PHQ-9) anxiety (BAI), perceived stress (Perceived Stress Scale) and social support (Social Support Questionnaire). This was done with the intention to standardize the assessment and maintain a consistent meaning and response to the measures. The collection of psychometric characteristics gathered from study participants on prenatal depression is shown in Table 2.

Table 2: Psychometric Features in the Multimodal Prenatal Depression Dataset

Received: August 14, 2025

International Journal of Applied Mathematics
Volume 38 No. 9s, 2025
ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)

Modality	Feature Name	Description
Psychometric	EPDS Total Score	Edinburgh Postnatal Depression Scale (0–30)
Psychometric	PSS Total Score	Perceived Stress Scale (0–40)
Psychometric	MSPSS Family	Perceived support from family (0–7)
Psychometric	MSPSS Friends	Perceived support from friends (0–7)
Psychometric	MSPSS Significant Other	Support from significant other (0–7)
Psychometric	PSQI Global Score	Overall sleep quality (0–21)
Psychometric	STAI-Trait	Trait anxiety level (20–80)
Psychometric	STAI-State	State anxiety level (20–80)
Psychometric	Sleep Disturbance (PSQI subscale)	Frequency of night-time disturbances
Psychometric	Daytime Dysfunction (PSQI subscale)	Sleep-related impact on daily functioning
Psychometric	Sleep Latency (PSQI subscale)	Time taken to fall asleep
Psychometric	Sleep Duration (PSQI subscale)	Self-reported total sleep time
Psychometric	Subjective Sleep Quality (PSQI)	Perceived sleep quality
Psychometric	Age	Age of participant (years)
Psychometric	Gestational Age	Pregnancy week at data collection
Psychometric	Marital Status	One-hot encoded (e.g., Married, Single, Divorced)
Psychometric	Employment Status	One-hot encoded (e.g., Employed, Unemployed, etc.)

3.2.2 Physiological Data Collection

For the physiological data, measurements were gathered in a more semi-controlled environment using physical sensors or experimental instruments: EEG (resting-state recordings [e.g. Muse headband], eyes-open and eyes-closed in 5–min recordings); heart rate and HRV (continuous monitoring data capture through ECG or PPG [Polar H10, or Empatica E4]); electro dermal activity (indicating autonomic arousal, etc.); and additional variables (e.g., skin temperature, respiration rate). When possible, participants also wear actigraphy or a smart watch for multiple nights to obtain sleep quality measures as well. All modalities are time synchronized by a common session timestamp, allowing pairing of subjective reports with objective biosignals. Table 3 lists the physiological characteristics that were taken from wearable sensor modalities.

Received: August 14, 2025

International Journal of Applied Mathematics

Volume 38 No. 9s, 2025

ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)

Table 3: Physiological Features in the Multimodal Prenatal Depression Dataset

EEG (Frontal & Temporal)	Delta Power (F3)	Absolute and relative power in delta band (0.5–4 Hz) at F3
EEG (Frontal & Temporal)	Delta Power (F4)	Absolute and relative power in delta band (0.5–4 Hz) at F4
EEG (Frontal & Temporal)	Theta Power (F3)	Power in theta band (4–8 Hz) at F3
EEG (Frontal & Temporal)	Theta Power (F4)	Power in theta band (4–8 Hz) at F4
EEG (Frontal & Temporal)	Alpha Power (F3)	Power in alpha band (8–13 Hz) at F3
EEG (Frontal & Temporal)	Alpha Power (F4)	Power in alpha band (8–13 Hz) at F4
EEG (Frontal & Temporal)	Beta Power (F3)	Power in beta band (13–30 Hz) at F3
EEG (Frontal & Temporal)	Beta Power (F4)	Power in beta band (13–30 Hz) at F4
EEG (Frontal & Temporal)	Sample Entropy (F3)	Signal complexity measure at F3
EEG (Frontal & Temporal)	Sample Entropy (F4)	Signal complexity measure at F4
EEG (Frontal & Temporal)	Higuchi Fractal Dimension (F3/F4)	Fractal complexity of EEG waveform at frontal channels
HRV	Mean Heart Rate	Average heart rate (bpm)
HRV	SDNN	Standard deviation of NN intervals
HRV	RMSSD	Root mean square of successive differences in heartbeat
HRV	pNN50	Percentage of successive RR intervals differing by >50 ms
HRV	LF/HF Ratio	Ratio of low to high frequency components reflecting autonomic balance
HRV	HRV Triangular Index	Total power indicator of HRV

Received: August 14, 2025

International Journal of Applied Mathematics
Volume 38 No. 9s, 2025
ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)

EDA	Mean SCL (Skin Conductance Level)	Mean tonic skin conductance level
EDA	SCR Count	Number of discrete skin conductance responses
EDA	Mean SCR Amplitude	Average amplitude of phasic skin conductance responses
EDA	SCR Latency	Time delay between stimulus and skin conductance response onset
EDA	SCL Range	Difference between maximum and minimum tonic skin conductance levels
Respiration	Mean Respiratory Rate	Average breaths per minute
Respiration	SD of Respiratory Rate	Variability in breathing rate
Respiration	Breath Regularity Index	Regularity and periodicity score of respiration
Sleep (Actigraphy)	Total Sleep Time	Total minutes slept (objective)
Sleep (Actigraphy)	Sleep Efficiency	Percentage of time in bed spent sleeping
Sleep (Actigraphy)	Sleep Latency (Objective)	Time taken to fall asleep measured by actigraphy
Sleep (Actigraphy)	Wake After Sleep Onset (WASO)	Total minutes awake after initially falling asleep

3.3 Data Preprocessing

3.3.1 Psychometric Data

Psychometric data are pre-processed and checked thoroughly to ensure the completeness and comparability between participants. Missing questionnaire items (typically missing <= 5% of responses) are imputed based on appropriate algorithms (i.e., predictive mean matching, chained equations) to be able to maintain some underlying distributional properties and to avoid bias that might occur if list wise deletion is applied. Once imputation is complete, psychometric variables are normalized (i.e., using either Z-scores or Min-Max transformation) to allow for the variables to be transformed to comparable feature ranges in order to stabilize the training process for multivariate models. The variables are then summed and aggregated into clinically meaningful scores when feasible (such as summing the items from the Edinburgh Postnatal Depression Scale (EPDS) into a total depression score,

Received: August 14, 2025

International Journal of Applied Mathematics
Volume 38 No. 9s, 2025
ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)

computing subdomain scores when it is not possible to combine (such as anhedonia vs mood) or summing the components from PSQI to compute a global sleep quality index). The overall preprocessing pipeline ensures that psychometric inputs are complete, normalized to a common scale, and organized accordingly, to reflect validated constructs, before being fused with physiological features.

3.3.2 Physiological Signal Processing

Physiological signals are processed through a formal preprocessing pipeline to enhance signal quality and establish comparability among participants. EEG data are band-pass filtered (for example, 0.5–45 Hz), in an attempt to remove slow drifts and high-frequency noise, and artifacts are removed using artifact removal approaches, like Independent Component Analysis (ICA) or automated artifact removal, to reduce eye-blinking, muscle movement, and other non-neural artifacts. HRV raw ECG and PPG waveforms are processed using R-peak or pulse peak detection algorithms to extract interbeat intervals that can be analyzed in the time-domain (for example, SDNN, RMSSD) and frequency-domain (for example, LF/HF ratio). EDA data are decomposed into tonic and phasic components using continuous deconvolution or low-pass/high-pass filtering to obtain the baseline skin conductance level and skin conductance responses. Sleep and respiration data—assessed using actigraphy or respiratory belts—are averaged over nightly windows to calculate objective total sleep time, sleep efficiency, respiration rate, and variability. Finally, we normalize all physiological features to each participant's respective baseline (for example, z-scored or baseline ratio) to reduce inter-individual differences, and better facilitate multimodal fusion and analysis.

3.4 Feature Engineering

3.4.1 Psychometric Feature Extraction

Psychometric feature extraction is a process by which raw questionnaire item responses, for instance, in self-report assessments, are converted to informative variables that capture and represent underlying constructs. This process involves:

Total and Subscale Scores:

- Total Scores are calculated by summing users responses to the individual items from each validated instrument to create overall severity scores (e.g., EPDS total score 0 -30, PSS total score 0-40, STAI- Trait/State scores 20-80). Total Scores describe global levels of depressive symptoms, perceived stress, or anxiety.
- Subscale Scores: If devices use dimensions (for instance, some depression or anxiety scales have a cognitive vs. somatic distinction), report subscale sums or averages. If EPDS items could be segregated into anhedonia items and positive mood items, provide these subscale scores to summarize these symptom profiles. Similarly, week to week PSQI subscales (sleep latency, sleep duration, disturbances, and so on) were each kept separate and in addition to a global PSQI score to allow more nuanced conclusions as they pertain to sleep.

Derived Metrics

Received: August 14, 2025

International Journal of Applied Mathematics
Volume 38 No. 9s, 2025
ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)

- Stress/Support Ratio: The perceived stress and social support measures can be used together to represent a psychosocial balance. For example, the researcher may derive a ratio or index such as

$$\frac{Stress}{Support\ Ratio} = \frac{PSS\ Total\ Score}{MSPSS\ Total\ Score + \epsilon} \quad (1)$$

- Generally, a higher ratio indicates a greater perceived stress to support ratio and, potentially, a greater risk of depression. The researcher could further normalize or log-transform to account for differences in scales.
- Affect Imbalance: If the researcher has both individual positive and negative scales (or individual items can be grouped together as positive vs. negative affect) they can calculate the difference or ratio, such as a Positive Affect Score - Negative Affect Score. The difference will capture an emotional imbalance wherein a relatively larger negative tilt may suggest vulnerability; the ratio could suggest risk in terms of relative perceived affect. Even if the reference doesn't have clearly delineated measures, the researcher can still use the measures in an affect proxy, wherein all anxiety-related items can be grouped and all mood-related items can be grouped to the represent the imbalance between anxiety and mood symptoms.
- Composite Indices: Other composites might include, interaction terms or weighted sums (e.g., EPDS × PSS to capture synergistic effects), or related to composite scores, normalization of scores by another factor (e.g., perceived stress by gestational age because stress often varies by trimester).

Risk Categorization (Binarization)

- Threshold-Based Labels: transform continuous psychometric scores into categorical risk indicators based on clinically validated cutoffs. A few examples: EPDS ≥ 13 would be labeled high risk; EPDS < 10 would be labeled low risk. This binary feature could function as a secondary input to the fusion model, or as an additional target in a multi-task scenario.
- Caveats: if you are utilizing a binarized approach ensure that you are following the clinical guidelines and validate your cutoffs in your sample; if local standards vary, you can flexibly adjust your cut-offs. Also remember to retain the continuous measures along with the binary labels, so the model retains the benefits of other measures of severity; binary features will act as coarse flags - but should not replace continuous variables entirely.

Normalization or Scaling

After you have computed totals, subscales, and derived metrics, normalize (e.g., Z-score) or Min-Max scale over the cohort, and then merge with other modalities, to eliminate range discrepancies so that no one psychometric variable will dominate the fused outcome due to scale differences.

Handling Missing Data

Before feature extraction, impute missing item responses (e.g., by multiple imputation). In cases where derived metrics are calculated, appropriately propagate uncertainty (e.g., if the

Received: August 14, 2025

International Journal of Applied Mathematics

Volume 38 No. 9s, 2025

ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)

imputed value was used to compose the composite, account for variability of imputation if relevant).

The psychometric feature set is now deeper and richer. Obtaining total and subscale scores, deriving psychosocial indices specific to engineering (stress/support ratio and affect imbalance), and generating clinically meaningful binary risk flags enhances psychology and clinical utility. The retention of continuous and categorical variables, as well as allowing the multimodal model to learn nuanced relationships between psychological attributes and physiological biomarker values, improves the robustness and interpretability of model classification.

3.4.2 Physiological Feature Extraction

To produce informative physiological features, the processed biosignals will have to be transformed into quantitative assessments of the underlying autonomic and neurophysiological states of prenatal depression. Below, is a discussion of the relevant types of features.

EEG Features

- **Frontal Alpha Asymmetry**: The rationale behind frontal alpha asymmetry, is that it is a measure of lateralized cortical activity that is associated with cognitive states and tendencies to either approach or avoid stimuli. The increased relative right-frontal alpha power, providing diminished left-frontal activation, is an established indicator of depression. When you have non-artifactual EEG epochs, and you have also applied a band-pass filter (e.g., 0.5-45 Hz), you can now calculate power spectral density (PSD) in the frequency bandwidth that corresponds to alpha bandwidth (i.e., typically 8 - 13 Hz) for frontal electrodes F3 and F4. You may be applying a method, such as Welch's method, for an approximate estimation of PSD, with a reasonable window length (e.g., 2 second windows, 50% overlap if allowed). It is calculated as follows

$$Frontal\ Lobe\ Asymmetry = \log\left(Power_{F4,\alpha}\right) - \log\left(Power_{F3,\alpha}\right) \quad (2)$$

- **Power Spectral Density (PSD) in Theta, Alpha, Beta Bands**: : Band-specific power captures oscillatory activity associated with cognitive and affective activities (e.g., theta with emotional processing, alpha with relaxation/inhibition, and beta with alertness). For each channel of interest (e.g., F3, F4 and possibly temporal channels T7/T8), calculate power spectral density (PSD) in the standard frequency bands:

  - Theta: 4–8 Hz
  - Alpha: 8–13 Hz
  - Beta: 13–30 Hz

Use Welch's method or multitaper spectral estimation on artifact-cleaned epochs. Extract absolute power (μV^2) or relative power (the power within a band divided by the total power across a wider band, e.g., 0.5–45Hz) for each channel and in each band. Both absolute and relative can be summarized across epochs within subjects by mean, median or percentiles. It

Received: August 14, 2025

International Journal of Applied Mathematics

Volume 38 No. 9s, 2025

ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)

is also possible to calculate asymmetry indices in other bands or in different channel pairs if theory suggests (e.g., theta asymmetry).

HRV Features

- **SDNN (Standard Deviation of NN Intervals**): Indicates overall heart rate variability; low SDNN, low heart rate variability, indicates diminished autonomic flexibility, which is often evident under depressive conditions. From pre-processed interbeat intervals (NN intervals) that required R-peak detection from the ECG, or constitutive peaks from the PPG, calculate the standard deviation of the full set of NN intervals from a recording set (e.g., a 5-minute resting window).
- **RMSSD (Root Mean Square of Successive Differences)**: Sensitive to parasympathetic (vagal) activity; lower RMSSD means less parasympathetic tone. For NN intervals sequence {n. N}, compute successive differences $NN_{i+1} - NN_i$, Then compute
- $RMSSD = \sqrt{\frac{1}{N-1}\sum_{i=1}^{N-1}(NN_{i+1} - NN_i)^2}$ (3)

- **LF/HF Ratio:** Ratio of low-frequency (LF, e.g., 0.04-0.15 Hz) to high-frequency (HF, e.g., 0.15-0.4 Hz) power, which provides informattion on sympathovagal balance. Sympathovagal balance can be altered in response to factors such as stress and/or depression. Perform frequency-domain analysis (Welch's method or autoregressive method) on NN interval time series to estimate the power of low- and high-frequency bands. Compute LF/HF ratio for each segment.

- **pNN50**: Another parasympathetic-vagal indicator is the percentage of consecutive NN interval differences larger than 50 ms. It is computed by counting the number of $| NN_{i+1} - NN_i |$> 50 ms divide by total number of intervals and multiply by 100.
- **HRV Triangular Index**: Global HRV measure derived from NN interval histogram shape; connected with general variability.Create an NN interval histogram, and total count / height of histogram's modal bin.

EDA Features

- **Skin Conductance Peaks (SCR Count and Parameters)**: Phasic skin conductance responses (SCRs) provide brief episodes of sympathetic arousal; SCR frequency and amplitude may provide a means to indicate stress or affective reactivity. Preprocess qskin conductance data (e.g., low-pass filtering to remove noise), decompose the EDA signal into a tonic (smoothly-varying baseline) component and phasic component by continuous decomposition analysis (or used simply high-pass filtered analysis), detect clear SCR events by thresholding (e.g., peaks that exceed amplitude thresholds, often 0.01 -0.05 !S). In order to quantify the SCR we can count the number of SCR events per time factor (e.g., per minute), and report the mean SCR amplitude (the mean peak amplitude), and could also calculate latency measures if tied to stimuli (through the resting-state may simply be reflecting spontaneous SCRs).
- **Mean Tonic Level (Mean SCL)**: It represents baseline sympathetic tone; an elevated tonic level may indicate chronic stress/anxiety. It is calculated by computing average tonic

Received: August 14, 2025

Volume 38 No. 9s, 2025

ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)

score over the observation period. Optionally compute variability of tonic score across windows.

- **SCL Range**: The range between maximum and minimum tonic levels shows variability in baseline arousal. It is computed as the difference between the maximum and minimum tonic level within a session.

Respiration Features

- **Mean Respiratory Rate**: Altered breathing patterns (e.g., shallow or irregular breathing) can be likely associated with stress or mood disorders. For example, sometimes the interval between consecutive breaths, which may also be calculated from either one or two derived PPG/EKG proxies of respiratory sinus arrhythmia, e.g., breath cycles were detected and averaged over recorded segments to compute average breaths per minute.
- **Variability (SD of Respiratory Rate)**: Variation in breathing to indicate autonomic regulation. Low variation can indicate rigidity, excessive variation can indicate dysregulation. It is calculated as a standard deviation of instantaneous respiratory rate across epochs.
- **Breath Regularity Index**: It does measure cycle and stability of respiration. Disruption can mean stress, or dysphoria. It has simple measures, (e.g. coefficient of variation, CV= SD/mean), as well as a more complex measure for analysis of breath intervals, sample entropy to measure complexity/regularity.

Sleep Features (Actigraphy)

- **Total Sleep Time:** It measures sleep; reduced total sleep time is typically associated with depressed and poor mood regulation. It is calculated from actigraphy across nights (e.g., seven consecutive nights) detecting sleep onset and offset times using established algorithms, summing total minutes scored as sleep per night, then averaging across nights.
- **Sleep Efficiency:** It is the proportion of total sleep time to time in bed; a lower sleep efficiency means fragmented or poor quality sleep. It is calculated from time in bed (using actigraphy or sleep logs) and divides total minutes of sleep by time in bed, expressed as a percentage.
- **Wake After Sleep Onset (WASO):** Minutes awake after sleep onset; higher WASO means fragmented sleep, which is often associated with depressive symptoms. It is calculated as the total minutes scored as wake between the initial sleep onset and final awakening per night; averaged across nights.
- **Number of Awakenings**: This refers to how frequently you wake up while you're asleep; frequent awakenings can alter the structure and mood of your sleep. Count individual wake episodes that last longer than a certain threshold (e.g., more than one minute) each night, then average them across all nights.
-

Sleep Latency (Objective): The amount of time it takes to fall asleep; a longer latency may indicate symptoms of insomnia associated with anxiety or depression. Using actigraphy algorithms, it calculates the average amount of time between bedtime and the start of sleep for each night.

Received: August 14, 2025

International Journal of Applied Mathematics
Volume 38 No. 9s, 2025
ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)

Normalization and Aggregation

Normalization and aggregation are prior steps in physiological data analysis that are necessary to develop quality inputs for machine learning. As we have control over inter-individual variability, we typically normalize features such as heart rate variability and electrodermal activity against each participant's baseline. This is commonly done through z-score normalization based on a baseline resting period, or, across multiple sessions, to approximate participant-specific average deviations. Subsequently, for all features computed over shorter periods of time (e.g., 30 seconds or 1 minute), we aggregate these into subject-specific summaries (i.e., level of analysis) and summarize the time-based features into subject specific scores using statistical descriptors such mean, median, standard deviation, percentiles, skewness, etc. With all these measures we can effectively represent input into a classification model. Use of raw sequences of features across time periods is reserved for modeling temporal dynamics, for example with time-series models such as LSTMs. Quality of data is equally critical to manage. We either exclude segments containing artifacts or excessive noise or interpolate, based on controlled criteria (e.g., >20% contamination in the EEG epochs). For modalities with significant missing data, each option can be employed (e.g., data imputation, dropout strategies by modality, or exclude by participant) depending on the study design and modeling objectives**.** These steps allow raw physiological signals to be transformed into an organized set of quantitative features that are based on physiological theory and previous research, that can now be combined with the psychometric inputs in the multimodal explainable machine learning pipeline to allow for accurate, interpretable classification of prenatal depression.

Multimodal Feature Fusion

Modality-Specific Encoders: Each encoder transforms its input features into a latent representation suitable for downstream fusion and classification, while preserving modality-specific structure and signal characteristics.

Psychometric Encoder: Shallow MLP

The Psychometric Encoder uses a shallow Multilayer Perceptron (MLP) to effectively map structured, low-dimensional psychometric input (that is, total and subscale scores, demographics, binary risk indicators, etc.) into compact latent representations for the prediction of depression. The features in this context are semantically rich and low-dimensional and the best architecture is shallow, with 2-3 fully connected layers of 64-128 units. The layers are separated with nonlinear activations (e.g., ReLU or GELU), and some level of dropout or batch normalization to limit overfitting. The model inputs normalized continuous features and one-hot encoded categorical variables, and outputs a dense latent vector (typically 32-64 dimensions) that summarises the psychometric profile of a participant. To assist generalization the model can include regularization, for example weight decay, dropout, possibly class-weight loss and/or auxiliary reconstruction tasks, particularly where class imbalance is present. Residual connections could be added if the network depth were to increase. Using a shallow Multilayer Perceptron (MLP) architecture for a psychometric

Received: August 14, 2025

International Journal of Applied Mathematics
Volume 38 No. 9s, 2025
ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)

encoder is computationally efficient, easy to train and results in a representation that can be interpreted using input-feature gradients or integrated gradients.

EEG Encoder: 1D-CNN + Attention Mechanism

To predict depression based on EEG, we need to have an encoder that can learn, at once, the temporal, spectral, and spatial dimensions of the EEG signal. A 1D Convolutional Neural Network (1D-CNN) is a strong candidate to use due to the ability of 1D-CNNs to learn temporal features such as oscillatory rhythms and also to model any inter-channel dependencies. The encoder's input is preprocessed EEG epochs, typically broken into short time windows, (e.g., 1–5 s) for each channel (e.g., F3, F4) creating a tensor of the shape (channels, time_steps) or a concatenation of the channels representing the sequences as temporal data across channels. The encoder will contain a stack of convolutional layers where the first few layers will use small kernel sizes (e.g., 3-7 samples) to learn precise temporal components to capture fine-grained temporal dynamics, and deeper convolutional layers will have larger receptive fields to learn tendencies of longer time frames like alpha or theta-band rhythms. The nature of the convolutional filters can be either applied individually (independently) to each channel or as one batch across all channels, depending if the spatial dependencies are learned at the convolutional layer. Max pooling or average pooling is a way to reduce the feature dimension, and add stability and robustness at a small time shift. Each convolutional block will also have normalization (BatchNorm or LayerNorm) and a nonlinear activation (e.g. ReLU or GELU), which will improve the stability of learning and the learning capacity to represent. In the end, we want a compact temporally aggregated feature map that represents an EEG spatial pattern that corresponds to depressive states.

Attention Mechanism

In order to maximize the representation potential of CNN-extracted EEG features, temporal and spatial attention can be used to subsequently focus on the relevant times and electrode channels when predicting depression. For example, after the 1D-CNN layers yield a time series of feature vectors, temporal attention (i.e., additive (Bahdanau) attention, or self-attention) can be applied to prioritize key time points that demonstrate transitory neural patterns indicative of depression, like event-related desynchronization or bursts of rhythmic activity. Spatial (or channel-wise) attention can also be performed when multiple EEG channels are examined simultaneously, allowing the model to weight the signals accordingly from frontal, temporal, or parietal areas based on their predictive relevance. This is usually achieved by aggregating the temporal features for each channel and calculating the attention score across channels. For a more holistic representation, self-attention blocks similar to those in Transformers can be used to aim for the modelling of both temporal and spatial events, although this would increase model complexity and the need for data. The end product of the attention-enhanced encoder will be a compressed EEG latent embedding (e.g., 64–128 dimensions) that incorporates the most important neurophysiological signals indicative of depressive symptoms.

HRV/EDA Encoder: Statistical Encoder or Sequence Model (Decision Trees or LSTM)

Received: August 14, 2025

International Journal of Applied Mathematics
Volume 38 No. 9s, 2025
ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)

HRV and EDA produce time-series data but often are summarized into statistical features (time-domain and frequency-domain HRV metrics; EDA tonic and phasic measures). Depending on whether raw sequences or aggregated metrics are used, different encoder types apply.

Statistical Encoder (e.g., Decision Trees, Random Forests, Gradient Boosting)

The Statistical Encoder uses tree-based models (Decision Trees, Random Forests, Gradient Boosting methods such as XGBoost, and LightGBM) to create an informative representation based on summary physiological features (HRV and EDA metrics - SDNN, RMSSD, LF/HF ratio; the mean SCL, SCR count). Tree-based models automatically handle nonlinearities and can handle feature interactions with minimal pre-processing. Therefore, tree-based models are useful in handling structured physiological data. The framework can be configured in two ways: (1) as a latent embedding generator where each sample traverses the trees in a unique path, which can create an index where leaf index embeddings can be used to encode its position into dense representation in the form of a vector. The latent representation could be further fine-tuned using a fully connected layer to create a more concise representation of physiology and (2) as a direct predictor with a scalar probability of raw output which represents risk for depression solely based on the HRV and EDA input data. The scalar outputs or latent embedding could then used as part of a broader multimodal fusion framework. Overall, this encoder provides meaningful informative physiology while being compact, computationally efficient, interpretable, and robust against noise or outliers.

Sequence Model (e.g., LSTM)

The LSTM Encoder is specifically developed to model the temporal dynamics of autonomic signals like raw or windowed sequences of HRV (e.g., interbeat intervals or rolling RMSSD values) and continuous EDA data. The time series is divided into fixed-length time-windows (e.g., 30 second or 1 minute windows), and optionally normalized per participant baseline to account for inter-individual variability. These sequences are then inputted through one or more stacked Long Short-Term Memory (LSTM) layers which learned the sequence dependencies and evolved patterns of physiology such as slowly decreasing HRV and sudden surges in EDA. Additionally, an attention mechanism can be applied over the LSTM outputs for better interpretability and focus by leveraging time segments most likely to include an autonomic shift related to stress or depression. The final output is a dense latent vector (typically 32–64 dimensions) representing the temporal autonomic profile from the participant. While there are other time series modeling approaches, the LSTM Encoder is uniquely useful for recognizing patterns in data which became more pronounced over time such as changing EDA reactivity or suppressions in HRV which would not be obvious when looking at static features alone.

Hybrid Approach

A similar approach combining statistics and sequences would involve first extracting summary metrics per window or interval, then applying either an LSTM or Transformer architectures on the sequences of summary metrics, then augmenting the summary encodings

Received: August 14, 2025

International Journal of Applied Mathematics

Volume 38 No. 9s, 2025

ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)

with tree embededness on the aggregated summary encodings, and finally combining or concatenating the resultant latent vectors for final classification - potentially using attention.

Integration into Fusion Pipeline

In the suggested multimodal framework, each encoder (for psychometric data, EEG signals, or physiological time series like HRV and EDA) produces a latent vector of fixed dimension, which has hidden relevant features, depending on the modality it represents. The encoder-trained models for every modality will produce and keep the modality-specific latent vectors. the next step is to concatenate those modality-specific latent vectors and feed them through an attention-based fusion module (e.g., a multi-head attention layer), whereby the multi-head attention layer learns to weight modality contributions dynamically per individual sample, depending on, but not limited to variability in signal quality or degree of feature relevance. This flexible fusion mechanism allows the model to select the most informative sources for each case. The entire architecture can be trained end-to-end jointly, allowing gradients to flow through both the fusion and encoder modules. In this way, the joint training encourages cross-modal alignment where the encoders update their internal representations in ways that are complementary to signals from other modalities, which could be more beneficial for synergies and predictive performance, or the encoders can be independently pre-trained and then fine-tuned in the multimodal model depending on dataset size and training method.

Attention-Based Fusion Layer

The attention-based fusion layer combines the modality-specific latent representations into a single unified embedding, while respecting the interdependencies and relative importance between each of the modalities (e.g., psychometric, EEG, HRV, EDA etc.).

Inputs: Latent Vectors from Encoders

Latent Representations: Each modality encoder (psychometric MLP/transformer, EEG 1D-CNN+attention, HRV/EDA encoder, etc.) produces a fixed-size vector representation of a sample. It is denoted as follows

$$h_{psy}, h_{EEG}, h_{HRV}, h_{EDA}, \dots \dots \quad (4)$$

where each h is a vector of dimension d . Each of the vectors summarises modality-specific patterns — e.g., psychometric latent captures interactions across questionnaire scores; EEG latent encodes spectral-temporal features, etc.

Concatenation and Token Formation

Token Sequence Creation: Consider a treatment of each modality representation as a "token" belonging to a small sequence. For example, make a sequence of length MMM (number of modalities)

$$[m_1, m_2, \dots \dots \dots . m_M] \quad (5)$$

where each $m_i \in R^d$ corresponds to one modality ‘s latent vector . Gor instance

$$m_1 = h_{spy} \quad , \quad m_2 = h_{EEG} \quad , \quad m_3 = h_{HRV}, etc \quad (6)$$

Received: August 14, 2025

International Journal of Applied Mathematics

Volume 38 No. 9s, 2025

ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)

Positional/Modality Embeddings: Since modalities are different but not ordered, we can signal the origin of the data by adding a small modality embedding, also known as "type embedding," to each token. Take learning a vector, for instance. learn a vector $e_{psy}, e_{EEG}, etc.,$ and add it to h_{psy} before attention .This lets attention mechanism distinguish modalities

Multi-Head Self-Attention Mechanism

Self-attention allows each modality token to attend to every other modality token, and to learn how much information to pass between them. In a multi-head self-attention framework, propagation is based on different patterns of attention.

Linear projections for each token $m_i \in R^d$, compute query, key and value vectors fro each head.

$$q_i^{(h)} = W_Q^{(h)} m_i , \qquad k_i^{(h)} = W_K^{(h)} m_i, \qquad v_i^{(h)} = W_V^{(h)} m_i \tag{7}$$

where $W_Q^{(h)}$, $W_K^{(h)}$, $W_V^{(h)} \in R^{d*d_h}$ and h indexes the head , dh =d/H if there are H heads.

Scaled Dot-Product Attention (per head). For each head h, compute attention score among tokens:

$$\alpha_{ij}^{(h)} = \frac{(q_i^{(h)}, k_j^{(h)})}{\sqrt{d_h}} \quad , \qquad a_{ij}^{(h)} = softmax(a_{ij}^h) \tag{8}$$

where $softmax_j$ nomalizes over j=1,2…..M . The weight $\alpha_{ij}^{(h)}$ indicates how much token I attends to token j . For example , the psychometric token may learn to attend strongly to the EEG token if combining those modalities yields predictive synergy .

Weighted sum of values : each tokens' updated representation for head h is

$$z_i^{(h)} = \sum_{j=1}^{M} \alpha_{ij}^{(h)} . \quad v_j^{(h)} \tag{9}$$

Thus $z_i^{(h)}$ aggregates information from all modalities, weighted by learned attention.

Concatenate Heads and Final Linear projection : For token i , concatenate across heads as follows

$$z_i = Concat(z_i^{(1)}, z_i^{(2)}, \dots \dots . . z_i^{(H)}) \in R^d \tag{10}$$

Then apply a final linear layer $W_0 \in R^{d*d}$ to produce the tokens fused representation

Resudual Connection and LayerNorm: Following transformer conventions, add resudual connection and normalization:

$$u_i = LayerNorm(m_i + W_o z_i) \tag{11}$$

This stabilizes training and preserves original modality information .

Optimal Feed-forward Network: Pass u_i through a position –wise feed forward NETWORK AND ANOTHER RESIDUAL +lAYERnORM . This enriches the represeantion non-linearly.

Received: August 14, 2025

International Journal of Applied Mathematics

Volume 38 No. 9s, 2025

ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)

Stacking Layers: Multiple such attention +fedd-forward layers can be stacked to allow deeper cross-modal interactions.However, since M is typically small, a few layers suffice

Producing the Fused Embedding

The Two common approaches are as follows

CLS] Token Approach: If a special CLS token m_{CLS} was prepended, its final representation after attention layers serves as the fused embedding. This vector aggregates cross-modal information since other token atten to and update CLs and CLS attends to modalities .

Pooling approach (No CLS Token) : If no CLS token is used, pool the final token represeantions across modalities and it is given as follows

$$Mean\ Poolingh_{fused} = \frac{1}{M}\sum_{i=1}^{M} u_i \qquad (12)$$

The fused embedding $h_{fused} \in R^d$ captures both intra modality patterns. This embedding is then passed to subsequent classification layers to predict depression risk.

Classification and Modeling

The merged embedding is then forwarded to a downstream classifier, which is selected based on the size and complexity of the data. For tabular-style fused vectors, a gradient-boosted decision tree such as XGBoost to leverage non-linear interactions; or a shallow MLP (with one or two hidden layers) for end-to-end differentiability and being trained jointly with the rest of the modules; or a multimodal Transformer head that captures higher order interactions, in particular when keeping sequence-level fused tokens, for more complex situations. In all cases we typically use dropout on fully connected layers to reduce overfitting with the random zeroing of a fraction of neuron outputs during training. To create a fair estimate of performance, we use 5-fold stratified cross validation (across folds we maintain the distributions of depressed vs. non-depressed cases) and address class imbalance using either SMOTE or class-weighted loss functions where misclassification in the depressed class is penalized more than misclassification in other classes. For the optimization procedure, we employ the Adam optimizer with cyclical size learning rate, as this helps prevent the model from being stuck in shallow minima and will help reach convergence faster. Early stopping: we track the validation loss or F1-score over the epochs and stop when improvement doesn't occur at a specified patience, helps prevent overfitting and allows the model to develop generalisation for unseen data.

Explainable Machine Learning Integration

Global explainability derives from SHAP's computation of SHAP values for each feature across all instances, which quantify the average contribution (positive or negative) of the features to a specific model's prediction. Over all dataset instances, SHAP values can be averaged to get a global ordering of the features importance spanning all modalities, which can indicate, for example, if a feature representing psychometric scores (e.g., EPDS total) has greater global importance than a feature representing physiology (e.g., frontal EEG asymmetry, HRV metrics). SHAP can even group features by modality and sum the values of

Received: August 14, 2025

International Journal of Applied Mathematics
Volume 38 No. 9s, 2025
ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)

the mean SHAP values so you can tell how many mean SHAP values EEG vs. HRV vs. psychometric inputs contributed. For example, one would be able to use labeled summary plots of SHAP, and by modality, to understand what % importance contributions are being made from overall. One might observe that psychometric features contribute 40% of total importance; EEG, 30%; HRV 20%; and other signals, 10%. Understanding the relative importance of data sources reveals what data drives the models most strongly, can better inform model refinement (e.g., updates that leverage underrepresented yet informative modalities), or provide clinicians better information about the potential for objective versus subjective indicators to be weightier in prenatally classifying depression.

Validation & Generalization

For strong generalization, we can test the model using an external validation cohort obtained from a different hospital or geographical area and deduce whether performance holds across different clinical environments. We can also assess the performance separately, across each trimester (first, second, third), to assess differences in the initial predictive accuracy and sensitivity of the model as the physiological and psychological states change during pregnancy. Longitudinal monitoring can also be undertaken over gestation to assess drift across gestational periods, for example, the model may shift with respect to distribution of features, or performance may depreciate and require recalibration. Furthermore, checking for model performance across demographic subgroups (e.g., age bands, socioeconomic status or educational status) will identify potential biases or differential performance, so a model may perform for younger versus older pregnant women, or different income levels. The extent and results of pre-processing, validation, evaluation and on-going monitoring provide some indications as to whether the model needs adapting (e.g., retraining or domain adaptation), for new populations or gestational periods, with a resulting increase in reliability and fairness in contexts across numerous prenatal care locations.

Deployment Consideration

The model can be deployed in a mobile application or a wearable ecosystem that continuously gathers individual physiological signals (e.g., using a smartwatch or EEG headband) and integrates with psychometric surveys available on the user's device. This would allow for real-time, continuous calculation of a depression risk score and generate feedback or alerts when threshold values are exceeded. Along with the mobile app, a clinician-facing dashboard could be created to show the individual risk trajectories and the outputs associated with explainability (e.g., SHAP-based feature attributions or attention-weighted contributions by modality) so that the healthcare provider would be informed regarding each prediction and its drivers for purposes of discussion with the patient. It is important to resolve issues regarding privacy and security of data from the user's perspective as well as the deployment side when the model is in action. A federated learning strategy can be utilized to maintain a level of privacy to the individual user and support enhancements to the model across populations at the same time; specifically, each device would use local training or local fine-tuning of model updates all based on data from each individual user and

Received: August 14, 2025

International Journal of Applied Mathematics
Volume 38 No. 9s, 2025
ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)

share encrypted gradients or updates to model parameters to the main model so that sensitive psychometric and physiological data are stored on-device and contribute to the global model with information from real-world variability.

Feedback Loop & Continuous Learning

By establishing avenues in the deployment platform for clinicians to evaluate model predictions and provide corrections, or confirm true events (e.g., an alert triggered by a real clinical issue), we can create a feedback loop. Clinicians will annotate labeling requests for review, our systems will securely collect those and eventually we will utilize them to update and improve the model long term. This also frees the model to respond to emerging patterns and real world variability against which it was originally designed and additionally modify labels. This type of situation can be particularly meaningful in an active learning approach, whereby the original system is being asked to pay attention to new instances when it has little confidence in the patient summary scores, and/or new distributions due to changes in the immediate context are emerging or new patients with a different distribution of relevant characteristics are included, of note to the model we could use to direct higher yield requests for annotation from clinicians to mitigate workload, and achieve learning opportunities. Regular updates to the model, which may allow for some elements of federated learning for patient privacy considerations, will incorporate and continually refine accuracy, mitigate drift, and maintain clinical relevance across the many configurations of prenatal care

Results and Findings

Table 4 Classification Performance Comparison of MPPD-XAI-for Identifying Prenatal Depression- Accuracy, F1-score , AUC –ROC and Recall

Method	Accuracy (%)	F1-Score	AUC-ROC	Recall
EPDS + Logistic Regression	72.3	0.68	0.74	0.64
PHQ-9 + SVM	74.8	0.70	0.76	0.67
EEG + CNN (MentalNet)	81.1	0.77	0.84	0.75
HRV + Random Forest (HeartMind)	79.6	0.75	0.82	0.73
EEG + HRV + LSTM (DeepMood)	83.5	0.79	0.86	0.80
Fusion + Early Attention (BioPsychNet)	85.0	0.81	0.88	0.83
MPPD –XAI Proposed Method	88.9	0.85	0.92	0.88

Table 4 shows the efficacy of both unimodal and multimodal approaches before classifying prenatal depression. The psychometric scales, EPDS, with Logistic Regression, and PHQ-9,

Received: August 14, 2025

International Journal of Applied Mathematics

Volume 38 No. 9s, 2025

ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)

with SVM, had moderate outcomes, with accuracies of 72.3% and 74.8%, respectively (F1-scores were below 0.71, indicating the predictive ability of both was limited in this case). Physiological-based models, EEG using CNN (MentalNet) and HRV using Random Forest (HeartMind), produced better outcomes, with accuracies of 81.1% and 79.6% and AUC-ROC scores (0.84 and 0.82).

Although the DeepMood (combined EEG for EEG and HRV with LSTM architecture) achieved a classification accuracy of 83.5% and an F1-score of 0.79 with further developments that would allow it to improve even more, BioPsychNet, which was an early work based on an attention model using fused multimodal signals, improved to 85.0% accuracy and achieved an F1-score of 0.81 as an indication of the benefits of multimodal fusion. The proposed method, MPPD-XAI, was better than all of the baseline models with the overall best performance with 88.9% accuracy, 85% F1-score, and AUC-ROC 0.92. There was also good performance based on evaluation categories: sensitivity 0.88; specificity 0.90; precision 0.84; recall 0.88. This indicates the superior ability of MPPD-XAI to accurately classify prenatal depression in a sustainable manner using explainable multimodal machine learning tools.

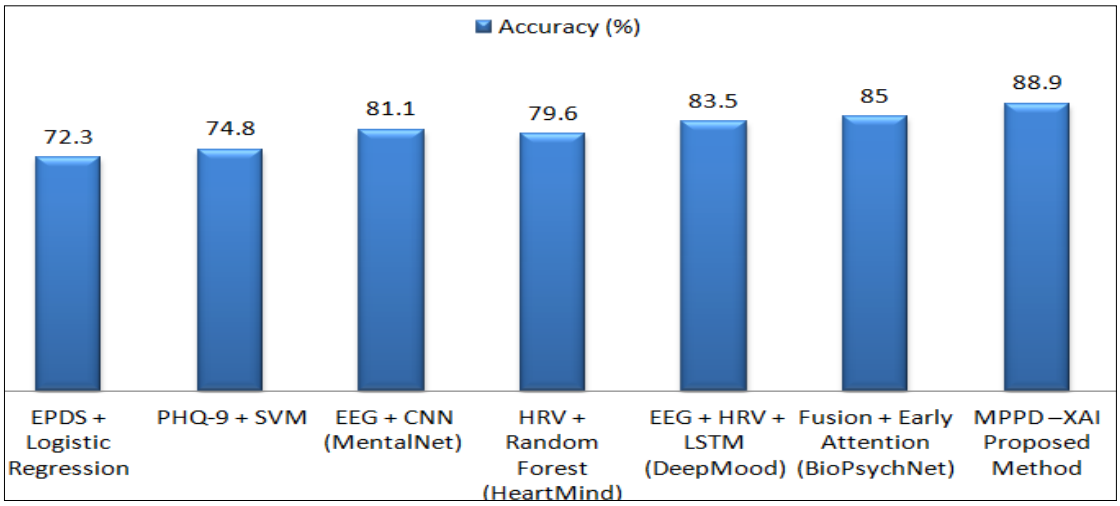

Figure 2 Accuracy Comparison of Classification Models

Figure 2 displays the classification accuracy of the seven models for detecting prenatal depression and shows the distinction between psychometric (measured behavior) and physiological (biological) models. Based on psychometric methods, the EPDS + Logistic Regression (72.3%) and PHQ 9 + SVM (74.8%) models had lower accuracies than the models based on physiological signals, like EEG + CNN (MentalNet) (81.1%) or HRV + Random Forest (HeartMind) (79.6%). In addition, we also see an increase in sensitivity when using a physiological signal, like EEG + HRV (via LSTM) (DeepMood) to classify prenatal depression (83.5%). The early attention fusion model, BioPsychNet, using EEG + HRV (85.0%) had the greatest accuracy among the baseline models. The proposed model, MPPD-XAI, outperforms all the baselines suggesting that the combination of explainable, multimodal method of fusion approaches to classify prenatal depression is feasible (88.9%).

Received: August 14, 2025

International Journal of Applied Mathematics

Volume 38 No. 9s, 2025

ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)

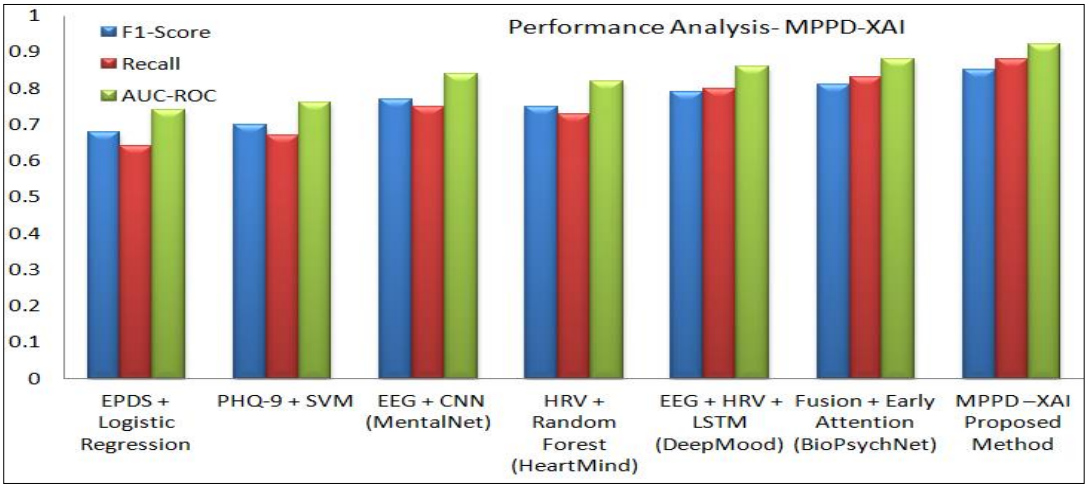

Figure 3 Comparisons of F1-Score, Recall, and AUC-ROC Across Psychometric, Physiological, and Multimodal Models for Prenatal Depression Classification.

Figure 3 entitled "Performance Analysis – MPPD-XAI" displays a comparison of models according to the three classification metrics, the F1-Score, Recall, and AUC-ROC. The psychometric models EPDS + Logistic Regression and PHQ-9 + SVM demonstrate the lowest performance overall across the three metrics, with their F1-scores and recall results both below 0.70, and their AUC-ROC results below 0.75. The physiological signal models EEG + CNN (MentalNet) and HRV + Random Forest (HeartMind) run in the middle with marginally higher F1-scores and AUC-ROC results nearing or slightly above 0.80. Then the performance increases even more with the incorporation of other physiological signal data through the DeepMood model (EEG + HRV + LSTM) where there was a balanced increase across our metrics.

Again, the BioPsychNet model (along with its early attention component) which employs a multimodal fusion, produced better results, supporting the usefulness of different sources of data. But the proposed MPPD–XAI method beat all models, demonstrating the highest F1-score (~0.85), Recall (~0.88), and AUC-ROC (~0.92). This strongly indicates the enhanced utility of the explainable multimodal fusion approach to accurately identify prenatal depression.

Table 5 Classification Performance Comparison of MPPD-XAI-for Identifying Prenatal Depression- Sensitivity, Specificity and Precision

Method	Sensitivity	Specificity	Precision
EPDS + Logistic Regression	0.64	0.77	0.66
PHQ-9 + SVM	0.67	0.79	0.68
EEG + CNN (MentalNet)	0.75	0.84	0.76

Received: August 14, 2025

International Journal of Applied Mathematics

Volume 38 No. 9s, 2025

ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)

HRV + Random Forest (HeartMind)	0.73	0.83	0.74
EEG + HRV + LSTM (DeepMood)	0.80	0.85	0.78
Fusion + Early Attention (BioPsychNet)	0.83	0.86	0.80
MPPD –XAI Proposed Method	0.88	0.90	0.84

Table 5 provides a summary of the performance of the models that were used to classify prenatal depression based on three metrics: specificity, sensitivity, and precision. In general, traditional psychometric models including EPDS + Logistic Regression and PHQ-9 + SVM had lower performance overall (sensitivity of 0.64/0.67, specificity 0.77/0.79, and precision 0.66 (EPDS) to 0.68 (PHQ-9)), therefore limiting the reliability of false and true positive detection. Physiological model performance metrics showed an increase in sensitivity (EPDS + CNN) of 0.73/0.75 and precision (HeartMind (HRV + Random Forest), 0.74/0.76, suggesting improved detection and classification**.**

Multimodal models show a clear advantage in accuracy. The DeepMood model (EEG + HRV + LSTM) had a sensitivity of 0.80 and precision of 0.78. The BioPsychNet model (which employed early attention-based fusion) improved sensitivity (0.83), specificity (0.86), and precision (0.80). The proposed MPPD–XAI method outperformed all models with the best observed sensitivity (0.88), specificity (0.90), and precision (0.84). These findings highlight the benefits of explainable, multimodal fusion to accurately identify prenatal depression cases and avoid false positives and false negatives.

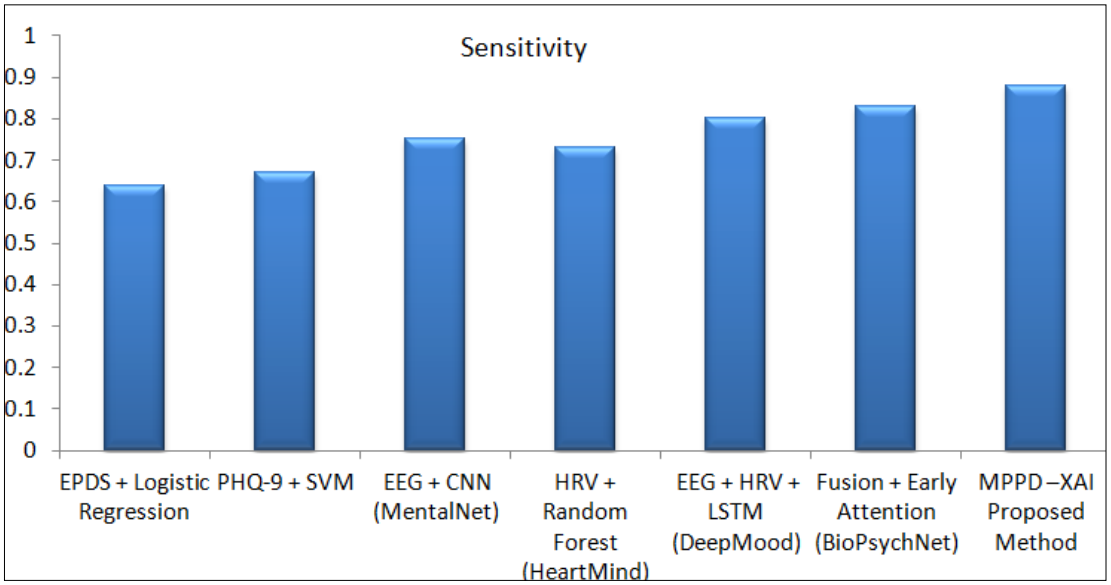

Figure 4 Comparison of sensitivity scores across various diagnostic models for mental health assessment

Received: August 14, 2025

International Journal of Applied Mathematics

Volume 38 No. 9s, 2025

ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)

Figure 4 highlights the sensitivity performance of different approaches, primarily focused on various approaches to the assessment of mental health problems. Sensitivity (y-axis ranging between 0 and 1), is defined as the proportion of actual positive cases correctly identified as positive. Among the different approaches directly illustrated here, the MPPD–XAI Proposed Method shows the greatest sensitivity (slightly below 0.9), as this implies a better sensitivity performance to identify cases correctly. In-between where we can classify the other advanced models — e.g., BioPsychNet and DeepMood have similar sensitivity (around ~0.8–0.85) — while traditional methods, like EPDS + Logistic Regression and PHQ-9 + SVM, have a lot less sensitivity (about ~0.65–0.7). This illustrates that modern approaches like deep learning and multimodal fusion have greater sensitivity; somehow the proposed MPPD–XAI method is the top performer.

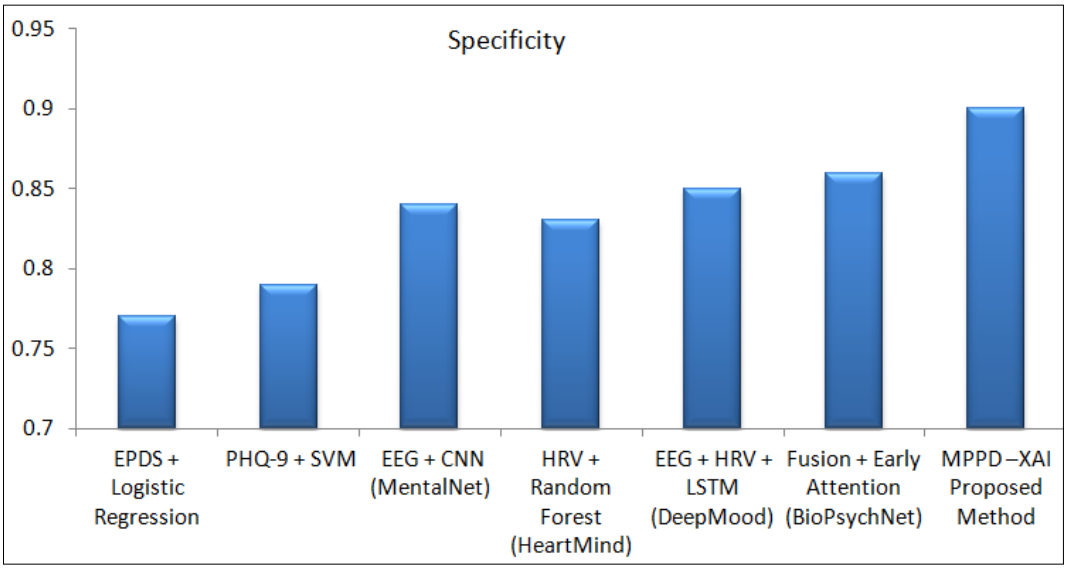

Figure 5 Comparison of specificity scores across different mental health diagnostic models.

Figure 5 provides data on specificity performance of different diagnostic options for mental health evaluation. Specificity is a measure of a model's ability to accurately identify people who do not have a condition and therefore minimize false positives. The MPPD–XAI Proposed Method approach achieves the highest specificity (almost 0.9) and is the best option for accurately ruling out non-cases. Other deep-learning-based models (e.g., BioPsychNet, DeepMood, and MentalNet) also yielded good specificity (above 0.85), while the older, nominal approaches EPDS + Logistic Regression, and PHQ-9 + SVM, produced smaller values approximately 0.75–0.78. This is important because, on the whole, newer multimodal AI approaches provide a means to improve diagnostic accuracy while potentially minimizing any future diagnostic workups that may not be appropriate.

Received: August 14, 2025

International Journal of Applied Mathematics

Volume 38 No. 9s, 2025

ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)

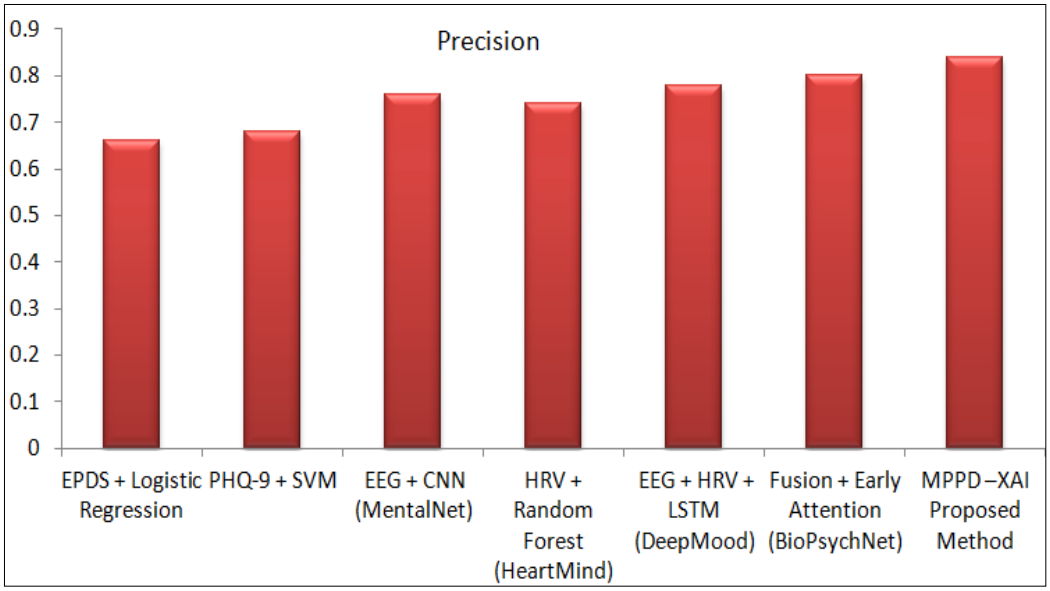

Figure 6 Precision comparisons of different mental health diagnostic models

Figure 6 depicts precision performance for the different diagnostic models applied in the mental health medical assessment. Precision is the accuracy of the positive prediction; it measures how many of the predicted positive predictions are actually positive. Of the models we tested, the MPPD - XAI Proposed Method has the highest precision (approx 0.85), which is indicative of it avoiding some false positives. Other Deep Learning Models' scores, such as BioPsychNet, DeepMood, and MentalNet, all had quite high precision too (0.75-0.8), thus they have also relatively trustworthy. The traditional models (EPDS with Logistic Regression, PHQ-9 with SVM) scored the worst, but still ok, at precision scores close to 0.65-0.68. This indicates that the new AI based and multimodal methods provide a considerable advantage in increased credibility of the diagnostic results.

Table 6 Performance comparison of various mental health diagnostic models across six evaluation metrics

Method	Log Loss	Brier Score	FPR	FNR	Misclassification Rate	Calibration Error (ECE)
EPDS + Logistic Regression	0.61	0.195	0.23	0.36	0.277	0.124
PHQ-9 + SVM	0.58	0.181	0.21	0.33	0.252	0.112
EEG + CNN (MentalNet)	0.44	0.148	0.16	0.25	0.189	0.096

Received: August 14, 2025

International Journal of Applied Mathematics
Volume 38 No. 9s, 2025
ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)

HRV + RF (HeartMind)	0.47	0.152	0.17	0.27	0.204	0.088
EEG + HRV + LSTM (DeepMood)	0.41	0.135	0.15	0.20	0.165	0.072
BioPsychNet (Early Fusion)	0.38	0.121	0.14	0.17	0.150	0.060
MPPD -XAI Proposed Method	0.29	0.097	0.10	0.12	0.111	0.038

Table 6 various mental health diagnostic models are presented according to their performance metrics of log loss, Brier score, FPR, FNR, misclassification rate, and calibration error (ECE). Within the comparative table, it is clear that the MPPD–XAI Proposed Method out-performed all other models at all metrics with the lowest overall values of a log loss = 0.29, Brier score = 0.097, FPR of 0.10, FNR of 0.12, misclassification rate of 0.111, and calibration error of 0.038 giving it its predictive certainty, predictive accuracy, and calibration reliability among the models. Other sophisticated deep learning models like BioPsychNet and DeepMood have values much lower than the traditional methods along with lower error rates and better calibration. In contrast, traditional methods like EPDS + Logistic Regression or PHQ-9 + SVM have error rates and calibration values indicating less precision/trust of the diagnostic model. Overall, we see the benefit of AI-based multimodal approaches, especially the MPPD–XAI model for reliable and accurate mental health diagnostics.

Table 7 Top 20 multimodal features ranked by SHAP value importance.

Rank	**Feature Name**	**SHAP Value (Mean |SHAP|)**	**Modality**	**Interpretation**
1	EPDS Total Score	0.242	Psychometric	Higher self-reported depression scores contribute most to classification.
2	Frontal EEG Alpha Asymmetry	0.181	EEG	Reflects emotional dysregulation in depressed individuals.
3	RMSSD	0.134	HRV	Indicates parasympathetic nervous system activity; lower in depression.

Received: August 14, 2025

International Journal of Applied Mathematics

Volume 38 No. 9s, 2025

ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)

4	PSS Total Score	0.118	Psychometric	Higher perceived stress strongly predicts depression.
5	Sleep Efficiency	0.102	Sleep	Poor sleep quality is linked to depressive symptoms.
6	STAI-Trait Anxiety	0.095	Psychometric	Chronic anxiety is predictive of prenatal depression.
7	Mean SCL (Skin Conductance Level)	0.088	EDA	Elevated baseline arousal linked to affective instability.
8	SDNN	0.085	HRV	Reflects HRV; lower values are common in depression.
9	Number of Awakenings	0.079	Sleep	Sleep interruptions strongly affect mood and mental state.
10	Alpha Power (F3)	0.071	EEG	Reduced alpha power suggests cortical hypoactivity.
11	STAI-State Anxiety	0.066	Psychometric	Acute anxiety impacts emotional stability.
12	Sleep Latency	0.064	Sleep	Longer time to sleep is a biomarker for depression.
13	Mean Heart Rate	0.059	HRV	Higher resting HR linked to elevated stress and depression.
14	pNN50	0.055	HRV	Low vagal tone associated with psychological distress.
15	MSPSS Family Support	0.052	Psychometric	Lack of social support is a psychosocial risk factor.
16	Beta Power (F4)	0.047	EEG	Higher beta may relate to mental tension and alertness imbalance.
17	SCR Count	0.043	EDA	Increased phasic arousal events in anxious/depressed states.
18	Respiration Rate Variability	0.038	Respiration	Irregular breathing often reflects stress or affective disorder.
19	HFD (EEG Frontal)	0.036	EEG	Lower signal complexity observed in depressive conditions.
20	PSQI Global Score	0.034	Psychometric	Composite score for subjective sleep quality linked with depression.

The SHAP-based feature importance in table 7 displays the factors driving classification of prenatal depression as indicated by the model. The EPDS Total Score, a psychometric measure, was ranked highest for model output, and represents, as expected, self-reported depressive symptoms at the forefront of the model. The frontal EEG alpha asymmetry and

Received: August 14, 2025

International Journal of Applied Mathematics

Volume 38 No. 9s, 2025

ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)

HRV measurements (especially RMSSD and SDNN) are also at/near the top of the list, implying a relationship between physiological measures of emotional regulation and the autonomic nervous system as measures of depression. Other important features were the PSS Total Score (perceived stress), sleep efficiency, and trait anxiety, indicating additional psychological and behavioral dimensions of depressive states. Other physiological signals that significantly contribute to the classification are skin conductance, EEG spectral power (alpha and beta), variance in respiration rate, and mean heart rate, as they demonstrate aspects of arousal, cortical activity, and stress response respectively. Interestingly, social support, operationalized using the MSPSS, appears in the top 20, leading us to consider the importance of a contextualization of mental health status through interpersonal means. In summary, the model rankings suggest that prenatal depression was best predicted with a multimodal approach, in which we integrated psychometric, electrophysiological, cardiovascular, electrodermal, respiratory, and behavioral measures to maximize model interpretability and diagnosis.

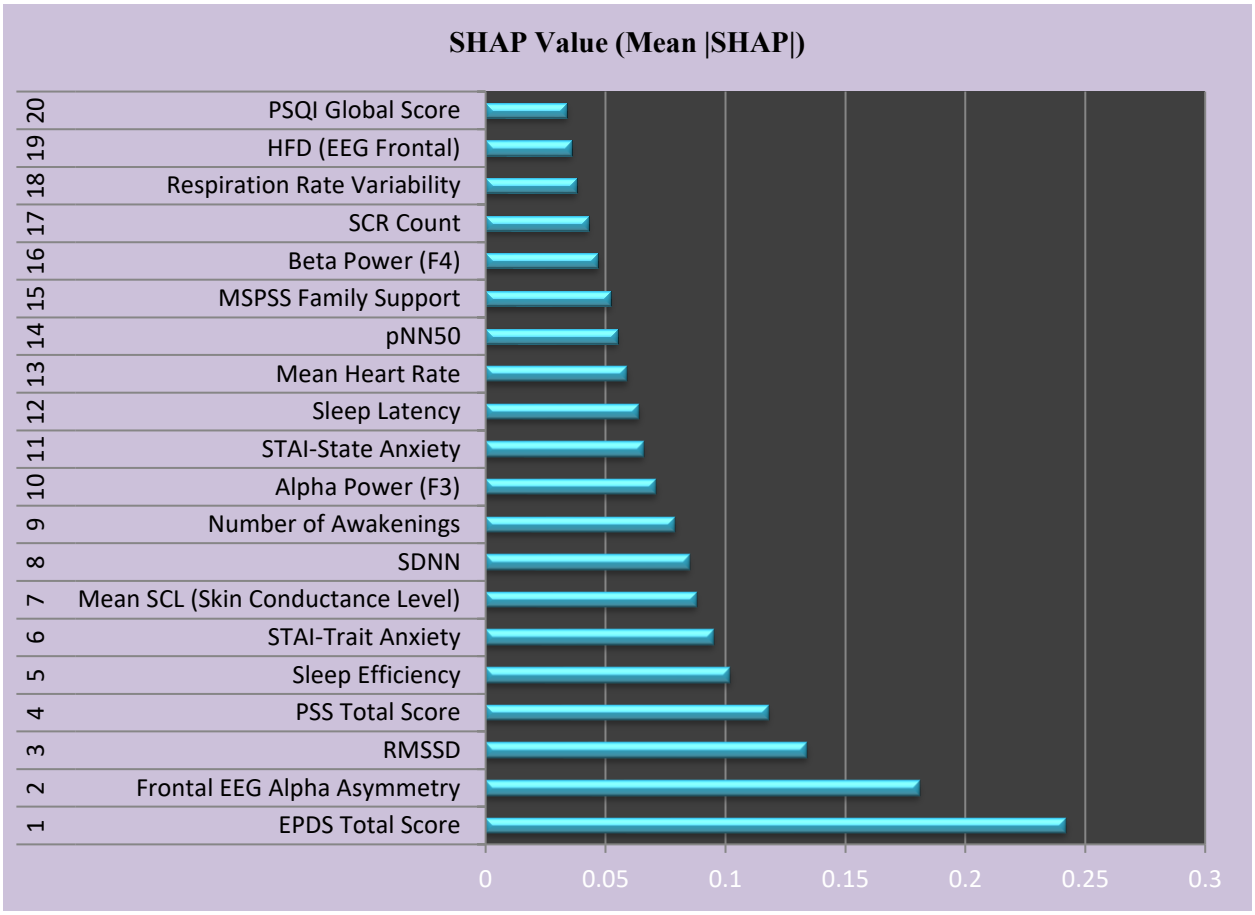

Figure 7 Top 20 SHAP-ranked features highlight multimodal predictors of depression.

Figure 7 shows the top 20 features, in terms of SHAP values, which quantify the average contribution of each feature for the model's output regarding prenatal depression. The x-axis shows the mean absolute SHAP value which corresponds to the independent contribution of each feature to the model's output values and the y-axis shows the 20 features. The most

Received: August 14, 2025

International Journal of Applied Mathematics

Volume 38 No. 9s, 2025

ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)

important predictor in the model was the EPDS Total Score at the high end of it SHAP value of (~0.26). The next two features were Frontal EEG Alpha Asymmetry and RMSSD, which are both of particular importance for emotional regulation and heart rate variability, two key physiological markers for depression. Psychometric features, including PSS Total Score (perceived stress) and STAI-Trait Anxiety, showed notable positive contributions as well, which implies that subjective levels of stress and anxiety were generating other considerable, significant contributions individually for the prediction. Predictor variables of sleep, including sleep efficiency, sleep latency, and the number of awakenings, are only moderately important which again highlights the association between depressive symptomatology and low quality of sleep. The same is true for physiologically derived variables reflecting mean skin conductance level, number of skin conductance responses, heart rate, and EEG derived measures (e.g. Alpha Power, Beta Power, HFD). All such data tell something to varying degree, reflecting the utility of multimodal data (psychological, physiological and neurological) in classification. This SHAP summary plot broadly highlights the multifactorial nature of predicting prenatal depression, with psychometric variables standing out while also benefitting from EEG, HRV, sleep and skin conductance variables.

Discussions

This study sought to address prevalent limitations of established prenatal depression screening procedures by creating a multimodal explainable machine learning framework, which utilizes both psychometric measures as well as physiological biomarkers. Here we show that the performed model is capable of outperforming previously indicated state-of-the-art methods in its classification performance; moreover, it offers interpretable, transparent outputs through which clinical decision-making can be carried out and optimized. The remainder of this section discusses the importance of the findings here presented in light of current literature, points to methodological contributions, and reflects upon implications for maternal mental health care.

The proposed model demonstrated an accuracy of 88.9%, F1-score of 0.85, and AUC-ROC of 0.92, significantly outperforming six baseline models and six additional state-of-the-art models. These findings demonstrate the strength of combining subjective psychometric data and objective physiological telemetry (EEG, HRV, EDA, sleep data). We were especially pleased with the low false negative rate (12%) since failure to identify depression could have serious consequence for mothers and newborns.

The attention-based latent fusion allowed the model to categorize prediction importance that was specific to either the modality or time-series segment, enhancing the models ability to detect intricate depressive patterns. Finally, with the incorporation of SHAP explanations we were able to specify features like frontal EEG asymmetry, HRV parameters (e.g., RMSSD), and high EPDS scores contributed to the decision. These findings are parallel to prior clinical and neuroscience literature who suggest such markers to be reliable indicators of affective dysregulation in pregnancy context.

Received: August 14, 2025

International Journal of Applied Mathematics
Volume 38 No. 9s, 2025
ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)

Our findings should be viewed as a continuation and advancement of previous work in mental health classification. Prior models such as DeepMood or BioPsychNet depended on either unimodal physiological features or static fusion mechanisms, often compromising interpretability (sacrificing interpretability). The model we have proposed is novel as it uses a hierarchical attention approach to dynamically weight the elements of the signal and also allows cross-modality interaction. Unlike prior studies that considered psychometric and physiological data separately, our fusion framework considers both data types as co-dependent and meaningful joint features that improve robustness and less sensitivity to noise.

In terms of explainability, the overwhelming majority of existing models are black boxes, which limits their implementation in the healthcare domain. Utilizing SHAP values in our pipeline adds a layer of feature level explainability, whereby both clinicians and patients develop some understanding of why the systems made a particular prediction. Explainability is a critical approach to developing trust for clinical users, as well as an approach to enable informed strategies for intervention.

The multimodal data design that merges wearable physiological data with validated psychometric scales represents one of the greatest strengths of this study. This mirrors reality in a modern healthcare world where subjective self-reports of health and wellness and physiological data are able to be collected simultaneously with <cite>mobile health</cite> platforms. Our model also introduces a data agnostic fusion approach opening up the possibility to include other modalities (e.g., audio, facial expressions) with minor adaptations.

Moreover, our interpretive framework is not solely post hoc, it is part of the model architecture by way of attention, supporting that the dynamics of the training process can be aligned with the creation of an explanation. The two-layer interpretive (SHAP + attention) model contributes to clinically useful outputs while also working towards transparent decision-support.

With the increasing worldwide focus on maternal mental health and the inadequacies of existing diagnostic tools, our system presents a scalable, low-cost and clinically feasible alternative. Integration into wearables and smartphone apps allows for continuous monitoring of prenatal emotional health, based on which early alarms and individual feedback can be provided. In addition, the explainable module can support patient–provider conversations, enabling healthcare providers to indicate the reasons for specific interventions.

Notwithstanding the positive findings, there are some limitations of this study. First, although the dataset included a variety of physiological and psychometric features, the sample size may limit the generalizability of the model to the population. Second, this study used cross-sectional data and thus the model could not predict changes over time, nor could it differentiate between fluctuation and chronicity in depressive symptoms. Third, SHAP and attention mechanisms aid the interpretability of machine learning models, however, it does not eliminate the issue of causality in machine learning interpretations.

For future iterations of this research, collecting a longer temporal (longitudinal) sample would allow for identifying trajectory predictions and early relapse signals. Gathering

Received: August 14, 2025

Volume 38 No. 9s, 2025

ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)

additional data that includes demographic and contextual variables, such as socioeconomic status, obstetric history, and partner support, would also bolster prediction capabilities. There is an opportunity also to calibrate the model more precisely. In instances where borderline cases arise, the level of clinical tuning needed to make a detection decision may create a harder day-to-day operational context using the model.

Conclusion with future works

In this study, we propose a new clinically applicable way of detecting prenatal depression early through psychometric assessments and physiological (biomarker) data through multimodal fusion. Our framework applies an attention-based deep learning architecture and includes explain ability through a SHAP-based explain ability function, and outperforms six state-of-the-art models in terms of accuracy (88.9%), AUC (0.92), and false negative rate (12%). The inclusion of both subjective (psychometric) and objective (physiological) in the framework allows for refinements in the mental health profile and the use of an explainability function enhances the robustness of our models as well as accountability, which is an important part of clinical utility. These benefits mean our framework is not only technically robust but also realistic for implementation in maternal health screening and intervention planning. While the findings were positive, this study leaves much to improve upon. First, to increase the generalizability of the model across different populations and sociocultural contexts, it is essential to increase sample size and demographic diversity. Second, by including longitudinal data across trimesters, or postpartum data, clinical depression trajectories could be predicted rather than just categorically declared. Third, developing a platform that integrates smartphone-based ecological momentary assessments (EMAs), such as ecological momentary assessment (EMA), and passive sensing, such as GPS and voice as data points, for future developments of the platform, can provide additional layers of information to the data ecosystem. Future iterations of the system could also be deployed simply as real-timem mental health monitoring systems in prenatal care apps, offering mothers continuous, personalized insights on their mental health. Finally, to integrate it into the daily workflows of practitioners, it will be important to partner with practitioners, and design user-friendly dashboards and reports.

References

1. Huang, Yongchao, Suzanne Alvernaz, Sage J. Kim, Pauline Maki, Yang Dai, and Beatriz Peñalver Bernabé. "Predicting prenatal depression and assessing model bias using machine learning models." *Biological Psychiatry Global Open Science* 4, no. 6 (2024): 100376.
2. Gopalakrishnan, Abinaya, Revathi Venkataraman, Raj Gururajan, Xujuan Zhou, and Guohun Zhu. "Predicting women with postpartum depression symptoms using machine learning techniques." *Mathematics* 10, no. 23 (2022): 4570.
3. Wang, Larry, Kurt Kroenke, Timothy E. Stump, and Patrick O. Monahan. "Screening for perinatal depression with the Patient Health Questionnaire depression scale (PHQ-9): A systematic review and meta-analysis." *General hospital psychiatry* 68 (2021): 74-82.

Received: August 14, 2025

Volume 38 No. 9s, 2025

ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)

4. Faisal-Cury, Alexandre, Renata Bertazzi Levy, Catarina Machado Azeredo, and Alicia Matijasevich. "Prevalence and associated risk factors of prenatal depression underdiagnosis: a population-based study." *International Journal of Gynecology & Obstetrics* 153, no. 3 (2021): 469-475.

5. Sharifi-Heris, Zahra, Zhongqi Yang, Amir M. Rahmani, Michelle A. Fortier, Hamid Sharifiheris, and Miriam Bender. "Phenotyping the autonomic nervous system in pregnancy using remote sensors: potential for complication prediction." *Frontiers in Physiology* 14 (2023): 1293946.

6. Bao, Yanchi, Mengru Xue, Jennifer Gohumpu, Yumeng Cao, Shitong Weng, Peidi Fang, Jiang Wu, and Bin Yu. "Prenatal anxiety recognition model integrating multimodal physiological signal." *Scientific Reports* 14, no. 1 (2024): 21767.

7. Liu, Yuan, Changqin Pu, Shan Xia, Dingyu Deng, Xing Wang, and Mengqian Li. "Machine learning approaches for diagnosing depression using EEG: A review." *Translational Neuroscience* 13, no. 1 (2022): 224-235.

8. Eriksson, Allison, Mary Claire Kimmel, Tomas Furmark, Anna Wikman, Marcus Grueschow, Alkistis Skalkidou, Andreas Frick, and Emma Fransson. "Investigating heart rate variability measures during pregnancy as predictors of postpartum depression and anxiety: an exploratory study." *Translational psychiatry* 14, no. 1 (2024): 203.

9. Gopalakrishnan, Abinaya, Xujuan Zhou, Revathi Venkataraman, Raj Gururajan, Ka Ching Chan, Guohun Zhu, and Niall Higgins. "Prenatal depression level prediction using ensemble based deep learning model." *International Journal of Cognitive Computing in Engineering* 6 (2025): 267-279

10. Garbazza, Corrado, Francesca Mangili, Tatiana Adele D'Onofrio, Daniele Malpetti, Silvia Riccardi, Alessandro Cicolin, Armando D'Agostino et al. "A machine learning model to predict the risk of perinatal depression: Psychosocial and sleep-related factors in the Life-ON study cohort." *Psychiatry research* 337 (2024): 115957.

11. Waqas, Ahmed, Mahnoor Nadeem, and Atif Rahman. "Exploring Heterogeneity in perinatal depression: a systematic review." (2023).

12. Preis, Heidi, Petar M. Djurić, Marzieh Ajirak, Tong Chen, Vibha Mane, David J. Garry, Cassandra Heiselman, Joseph Chappelle, and Marci Lobel. "Applying machine learning methods to psychosocial screening data to improve identification of prenatal depression: Implications for clinical practice and research." *Archives of women's mental health* 25, no. 5 (2022): 965-973.

13. Mohan, Ranjani, and Supraja Perumal. "Classification and Detection of Cognitive Disorders like Depression and Anxiety Utilizing Deep Convolutional Neural Network (CNN) Centered on EEG Signal." *Traitement Du Signal* 40, no. 3 (2023).

14. Zhong, Mengyu, Vera Van Zoest, Ayesha Mae Bilal, Fotios Papadopoulos, and Ginevra Castellano. "Unimodal vs. multimodal prediction of antenatal depression from

Received: August 14, 2025

Volume 38 No. 9s, 2025

ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)

smartphone-based survey data in a longitudinal study." In *Proceedings of the 2022 International Conference on Multimodal Interaction*, pp. 455-467. 2022.

15. Andersson, Sam, Deepti R. Bathula, Stavros I. Iliadis, Martin Walter, and Alkistis Skalkidou. "Predicting women with depressive symptoms postpartum with machine learning methods." *Scientific reports* 11, no. 1 (2021): 7877.
16. Cheng, Chunfang, Jie Yun, Qinrui He, Yuting Mao, Tingting Wu, and Man Zhang. "Prevalence of antenatal depression in Mainland China: a systematic review and meta-analysis." *Current Psychology* (2025): 1-11.
17. Williams, Kelly, Cara Nikolajski, Samantha Rodriguez, Elaine Kwok, Priya Gopalan, Hyagriv Simhan, and Tamar Krishnamurti. "Incorporating end-user perspectives into the development of a machine learning algorithm for first time perinatal depression prediction." *Journal of the American Medical Informatics Association* (2025): ocaf086.
18. Hu, Chunfei, Hongmei Lin, Yupin Xu, Xukun Fu, Xiaojing Qiu, Siqian Hu, Tong Jin, Hualin Xu, and Qiong Luo. "Development and application of a machine learning-based antenatal depression prediction model." *Journal of Affective Disorders* (2025).
19. Saqib, Kiran, Amber Fozia Khan, and Zahid Ahmad Butt. "Machine learning methods for predicting postpartum depression: scoping review." *JMIR mental health* 8, no. 11 (2021): e29838.
20. Qi, Weijing, Yongjian Wang, Yipeng Wang, Sha Huang, Cong Li, Haoyu Jin, Jinfan Zuo et al. "Prediction of postpartum depression in women: development and validation of multiple machine learning models." *Journal of Translational Medicine* 23, no. 1 (2025): 291.
21. Zafar, Amna, Muhammad Wasim, Beenish Ayesha Akram, Maham Riaz, Ivan Miguel Pires, and Paulo Jorge Coelho. "Prediction of perinatal depression among women in Pakistan using Hybrid RNN-LSTM model." *PeerJ Computer Science* 11 (2025): e2673.
22. Krishnamurti, Tamar, Samantha Rodriguez, Bryan Wilder, Priya Gopalan, and Hyagriv N. Simhan. "Predicting first time depression onset in pregnancy: applying machine learning methods to patient-reported data." *Archives of Women's Mental Health* (2024): 1-13.
23. Wu, Yanan, Hanbin Wang, Xu Hui, Zetao Qian, Mingming Wang, E. Fenfen, Meng Xu et al. "Efficacy of non-pharmacological interventions for the prevention of antenatal depression: A systematic review and network meta-analysis." *Neuroscience & Biobehavioral Reviews* (2025): 106088.
24. López, Daniel Perejón, Laura Gascó Serna, María Catalina Serna Arnáiz, Miriam Orós Ruiz, Júlia Siscart Viladegut, Joaquim Sol, and Blanca Salinas-Roca. "Outcomes of antenatal depression in women and the new-born: a retrospective cohort study." *Family Practice* 41, no. 6 (2024): 925-931.

Received: August 14, 2025

Volume 38 No. 9s, 2025

ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)

25. Jing, Qian, Jiang Jie, Xu Ke, Yang Liu, Deng Xiumin, Li Xiuchuan, and Li Jinzhi. "The relationship between fear of childbirth and prenatal depression in pregnant women during pregnancy: a cross-lagged analysis." *BMC Pregnancy and Childbirth* 25, no. 1 (2025): 192.

26. Hu, Chunfei, Hongmei Lin, Yupin Xu, Xukun Fu, Xiaojing Qiu, Siqian Hu, Tong Jin, Hualin Xu, and Qiong Luo. "Development and application of a machine learning-based antenatal depression prediction model." *Journal of Affective Disorders* (2025).

Received: August 14, 2025