



A machine learning model to predict the risk of perinatal depression: Psychosocial and sleep-related factors in the Life-ON study cohort

Corrado Garbazza^{a,b,c,1,*}, Francesca Mangili^{d,1}, Tatiana Adele D'Onofrio^d, Daniele Malpetti^d, Silvia Riccardi^a, Alessandro Cicolin^e, Armando D'Agostino^{f,g}, Fabio Cirignotta^h, Mauro Manconi^{a,i}

^a Sleep Medicine Unit, Neurocenter of Southern Switzerland, Lugano, Switzerland

^b Centre for Chronobiology, University of Basel, Basel, Switzerland

^c Research Cluster Molecular and Cognitive Neurosciences, University of Basel, Basel, Switzerland

^d Istituto Dalle Molle di Studi sull'Intelligenza Artificiale (IDSIA), USI/SUPSI, Lugano, Switzerland

^e Sleep Medicine Center, Department of Neuroscience, University of Turin, Turin, Italy

^f Department of Mental Health and Addiction, ASST Santi Paolo e Carlo, Milan, Italy

^g Department of Health Sciences, Università degli Studi di Milano, Milan, Italy

^h University of Bologna, Bologna, Italy

ⁱ Faculty of Biomedical Sciences, Università della Svizzera Italiana, Lugano, Switzerland

the “Life-ON” study group

ARTICLE INFO

Keywords:

Depression
Sleep
Women
Pregnancy
Postpartum
Risk factors

ABSTRACT

Perinatal depression (PND) is a common complication of pregnancy associated with serious health consequences for both mothers and their babies. Identifying risk factors for PND is key to early detect women at increased risk of developing this condition. We applied a machine learning (ML) approach to data from a multicenter cohort study on sleep and mood changes during the perinatal period (“Life-ON”) to derive models for PND risk prediction in a cross-validation setting. A wide range of sociodemographic variables, blood-based biomarkers, sleep, medical, and psychological data collected from 439 pregnant women, as well as polysomnographic parameters recorded from 353 women, were considered for model building. These covariates were correlated with the risk of future depression, as assessed by regularly administering the Edinburgh Postnatal Depression Scale across the perinatal period. The ML model indicated the mood status of pregnant women in the first trimester, previous depressive episodes and marital status, as the most important predictors of PND. Sleep quality, insomnia symptoms, age, previous miscarriages, and stressful life events also added to the model performance. Besides other predictors, sleep changes during early pregnancy should therefore be assessed to identify women at higher risk of PND and support them with appropriate therapeutic strategies.

1. Introduction

Perinatal depression (PND) refers to the occurrence of a major depressive episode during pregnancy or within 4 weeks after childbirth, although most experts agree that any depressive episode up to one year postpartum should be considered as PND (Dagher et al., 2021). Common symptoms of PND include depressed mood and energy, weepiness, reduced appetite or overeating, either excessive or disrupted sleep,

feelings of unworthiness and overworry about the well-being of the baby, and even thoughts of harming oneself or the baby (Van Niel and Payne, 2020). Due to its high prevalence rate (ca. 12 % of women affected worldwide) (Woody et al., 2017) and the detrimental impact on the health of mothers, children, and their families, PND represents one of the most serious complications of pregnancy (Dagher et al., 2021). Moreover, given the resulting socioeconomic burden, PND is considered a priority target of health prevention strategies globally (Howard and

* Corresponding author at: Centre for Chronobiology, Psychiatric Hospital of the University of Basel, Wilhelm Klein-Strasse 27, CH, 4002 Basel, Switzerland.
E-mail address: corrado.garbazza@upk.ch (C. Garbazza).

¹ These authors contributed equally.

Khaliq, 2020). However, it has been estimated that approximately 50 % of women suffering from PND still remain undetected (Yawn et al., 2012), calling for greater efforts in early diagnosis and treatment. Finally, the management of PND is complicated for two main reasons: 1) the lack of clearly defined predictive factors that allow early identification of women at risk; 2) the limitations related to the poor acceptance of the use of drugs during pregnancy due to their safety profile.

Among the currently available screening instruments for PND, the Edinburgh Postnatal Depression Scale (EPDS) (Cox et al., 1996) is the most widely used tool internationally, although heterogeneous results have been reported in different countries and clinical settings (Gibson et al., 2009). Some risk factors for PND occurrence, including a previous history of psychiatric disorders, domestic violence, socioeconomic status, and obstetric factors, have been identified in perinatal women, but their predictive power seems to be limited and difficult to implement in a clinical context (Cellini et al., 2022).

In recent years, machine learning (ML) techniques have been increasingly applied to analyze large datasets, such as those based on digital information gathered during the perinatal period. By overcoming the limitations of standard statistical approaches (Dwyer et al., 2018), ML methods provide a more practical and accurate stratification of at-risk patients (Rajkomar et al., 2018) and are therefore acquiring greater relevance towards the development of predictive medicine and personalized psychiatry (Shatte et al., 2019).

To date, only a few studies, recently reviewed by Cellini et al. (Cellini et al., 2022), applied ML techniques to predict postpartum depression, with promising albeit limited findings.

Here we present a first ML model to predict the risk of PND occurring at any time from the second trimester of pregnancy up to six months after delivery, by using data collected from a large population of healthy women in the first trimester. Due to the well-known relationship between sleep disorders and depression, in addition to medical, psychological and sociodemographic factors, in our analysis we consider a set of subjective and objective sleep variables, in order to investigate the possible role of sleep quality in PND prediction and prevention.

2. Methods

2.1. Data collection - The “Life-ON” project

Data for this analysis were derived from the “Life-ON” project, a multicenter cohort study on sleep and mood changes across the perinatal period conducted from 2016 to 2019. The study centers were located in three large and wealthy cities in Northern Italy (Milan, Turin and Bologna, among the cities with the highest GDP in the country) and one

in Switzerland (Lugano). Four hundred and thirty-nine women (age 33.7 ± 4.2) were recruited during the first trimester of pregnancy and regularly followed-up until one year postpartum by multidisciplinary teams of investigators, including psychiatrists, neurologists, and psychologists, all of them also having expertise in sleep medicine. Women using antidepressants at time of screening or in the 6 months prior to screening were excluded from participation. At study entry (visit 1), demographic information, as well as clinical, gynaecological, psychiatric, and sleep-related data were collected from the participants according to the “Life-ON” protocol, which has been detailed in Baiardi et al. (Baiardi et al., 2016) and illustrated in Fig. 1. A detailed list and description of the assessment tools used is included in the supplementary materials section (Table S4).

Polysomnographies (PSG) were recorded during weeks 20–25 of pregnancy in the participants’ domestic environment (level 2), using a portable device (Embletta®, NeuroLite), applied by expert sleep technicians at the participants’ home.

Restless Legs Syndrome (RLS) evaluation involved a clinical assessment based on the 5 diagnostic criteria established by the International Restless Legs Syndrome Study Group, followed (in case of a positive diagnosis) by administration of the International Restless Legs Study Group Severity Rating Scale (IRLS). Blood tests consisted of repeated measurements of prolactin, progesterone, ferritin, estrogen, iron throughout the study period, as indicated in Fig. 1.

Due to a relevant number of dropouts and missing data encountered during the last part of the 18-month follow-up, for our analysis we considered only the first nine study visits, corresponding to an observation period ranging from the first gestational trimester up to sixth months after delivery.

For this time frame, >1000 variables were included in the “Life-ON” database, a subset of which were selected for analysis based on the following criteria:

- All variables that were not considered relevant to the outcome by expert judgment were removed. This set includes variables such as each single item of the questionnaires, which are summarized in validated and clinically used scores (e.g., HDRS, EPDS, IRLS, etc.) or metadata collected for administrative purposes (e.g., dates, measurement units).
- Among the remaining variables, those with more than 9% of missing values were removed.
- For pairs of strongly correlated variables (Pearson correlation coefficient above 0.85) only one, selected by an expert, was retained in the dataset.

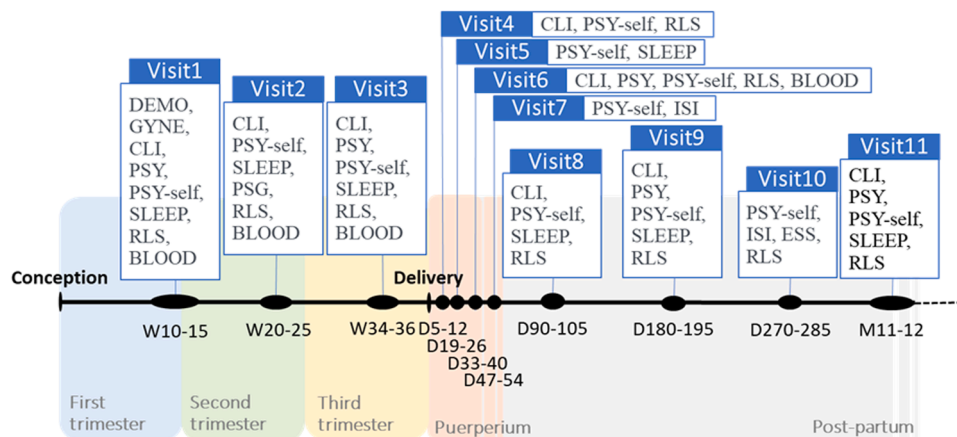


Fig. 1. Timeline of the Life-ON visit schedule. Variables are grouped as follows. DEMO: demographic assessment, GYNE: gynaecological evaluation, CLI: clinical evaluation, PSY: structured or semi-structured psychiatric assessment (MINI, HDRS, MADRS), PSY-self: self-administered psychiatric assessment (EPDS, VAS), SLEEP: sleep questionnaires (PSQI, ISI, ESS), PSG: polysomnography, RLS: restless legs syndrome evaluation, BLOOD: blood tests. W=week, D=day, M=month.

2.2. Data analysis and machine learning

The primary outcome of our analysis was a binary variable, hereafter referred to as PND12, which was considered TRUE for women with an EPDS score above the cutoff of 12 at least once at any time point from visit 2 to visit 9, and FALSE otherwise. However, a whole series of 8 consecutive EPDS scores was only available for 12,76 % of the 439 women included in the study. Since missing EPDS scores were ignored by the above definition of PND12, it was implicitly assumed that they did not exceed the depression threshold, which implies some degree of underestimation of the PND prevalence and risk. Indeed, if a woman experienced PND but her EPDS score above the threshold was missing, this was classified as FALSE. To limit this drawback, 63 cases of women for whom less than 3 out of 8 EPDS scores were available were excluded from the analysis. Therefore, 376 cases entered the final analysis.

2.2.1. Preliminary analysis and preprocessing

A bivariate analysis of the association between each predictor and the outcome variable PND12 was performed based on *t*-tests for continuous variables and chi-squared tests for categorical variables. Yates' correction for continuity was used in the chi-squared tests for binary variables. P-values were identified as significant using the Benjamini-Hochberg procedure (Benjamini and Hochberg, 1995), with a false discovery rate (FDR) of 0.05. No imputation of missing values was performed in this first, explorative, step of the analysis.

Since only 211 women in the study had complete data, to avoid an excessive reduction in the dataset size for ML analysis, we opted for imputation of missing values after performing analysis of their distribution (see the supplementary material for more details). For this purpose, we used the Multivariate Imputation via Chained Equations (MICE) method, with predictive mean-matching as the imputation procedure (Buuren and Groothuis-Oudshoorn, 2011). Missing data were imputed only once, before starting the ML analysis.

2.2.2. Predictive models

We compared three machine learning algorithms, penalized logistic regression, random forests and support vector machines (SVM) (see supplementary material for details). The following analyses refer to the selected method, i.e., the SVM (Noble, 2006). The main goal of ML was to develop a model detecting women at risk of developing PND using only observations collected during their first trimester of pregnancy. We used the Radial Basis Function (RBF) kernel and optimized the regularization strength by a grid search procedure on a nested 4-fold cross validation. The classification threshold was set to yield a specificity of at least 80% on the training data, which would guarantee a reasonable number of unnecessary treatments (considering that no dangerous treatments are administered during pregnancy or lactation, a 20 % of false positives was considered acceptable). In order to simplify the model for the benefit of its interpretability, we introduced a feature selection (FS) step retaining only the 10 most important features. The importance of the predictors was assessed based on their permutation importance (Breiman, 2001) computed on a nested 4-fold cross-validation.

The performances of the model learned were estimated in 5 repetitions of a 3-fold cross-validation. The metrics used to compare algorithms are the area under the ROC curve (AUROC), the area under the precision recall curve (AUPRC), the sensitivity (proportion of positive cases correctly identified or true positive rate) and specificity (proportion of FALSE cases correctly classified or true negative rate). In all cross-validations implemented, sampling was stratified on the PND12 outcome. To interpret the final model decisions, we employed SHapley Additive ExPlanations (SHAP), a method rooted in game theory that utilizes Shapley values. We calculated the SHAP value for each variable and instance, and we ranked the variables based on the mean absolute value of the SHAP values.

Finally, we examined the possibility to achieve a more accurate PND

risk prediction across the perinatal period, by adding the information becoming available at each follow-up visit. For this purpose, six other classifiers were trained, one for each visit (*v*) from the second to the seventh. Each of these classifiers receives as input, in addition to the variables observed during the first trimester, also those collected during all subsequent visits, up to and including the visit (*v*) at which the classifier should be used to predict the risk that the perinatal woman shows an EPDS > 12 at least once in the remaining visits (i.e. from visit *v* + 1 to 9). For instance, a model is trained to be used during the second trimester of pregnancy (*v*2) to predict the risk of developing PND during the last trimester (*v*3) or later, up to six months after delivery (*v*9). Such a model will use for the prediction all the observations collected during both the first and the second trimesters (*v*1 and *v*2). The training dataset of each of these classifiers included only those women for which at least 2 EPDS scores had been recorded between visits *v* and 9 (reduced to 1 for the classifiers built from *v*5).

2.2.3. Software

All analyses have been carried out using Python (version 3.8.12) and R through the Python package rpy2 (version 3.7.13).

3. Results

Longitudinal data from 439 participants of the "Life-ON" study were analyzed. Regarding the number of pregnancies, 189 of 426 women who provided this information were primiparous (43 %). For 63 of the 439 participants (14.35 %) less than 3 EPDS scores were available between visits 1 and 9. Among the 376 remaining ones, 56 (14.89 %) had an EPDS > 12 between visits 2 and 9. Fig. 2 shows the distribution of the EPDS scores obtained across the study (panel A), their mean value with 95 % confidence interval (panel B) and the fraction of women with EPDS > 12 (panel C) for each visit. Delivery occurred for all women between visits 3 and 4 (vertical dotted line in Fig. 2), after which a decrease in mean EPDS scores was generally observed.

Based on expert opinion and the statistical properties detailed in Section 2.1, 124 predictors were pre-selected. Of these, 48 were obtained during the first study visit (visit 1) corresponding to the first trimester of pregnancy (10–15th gestational week) and 33 were derived from the polysomnography (PSG) recorded during the second trimester (20–25th gestational week). The complete list of variables considered, including some descriptive measures of their distribution and the p-values of the hypothesis tests assessing their association with the PND12 outcome, is provided in Tables 1 and 2. The bivariate analysis highlighted significant associations between the PND12 variable and the first trimester scores of the questionnaires EPDS, MADRS, VAS, ISI, and PSQI ($p < 0.001$), as well as the information derived from the clinical interview history of depression (self-reported), history of diagnosed major depressive disorder (MDD, assessed by MINI) ($p < 0.001$), and history of treatment for depression (self-reported) ($p = 0.003$). Significant associations were also found for the PSG variable AI index ($p = 0.001$), as well as for neck circumference ($p = 0.002$) and marital status ($p = 0.006$). Without adjustment for multiple tests, a significant correlation ($p < 0.05$) was also found for the predictors working condition, total number of miscarriages, ESS, RLS severity, number of central hypopneas and the percentage of N2 sleep stage over the total sleep time.

A visual inspection of missing data spotted no association between the values of a predictor and the missingness of another, but showed that some variables were often missing together (see supplementary material). Furthermore, the inclusion of the PSG variables among the predictors did not result in any improvement in the performance of the SVM algorithm (see Table S2, supplementary material) and the PSG variables were therefore not included among the predictors in the ML model.

The SVM algorithm implemented after the 10-features selection (10-FS) on the MICE imputed dataset was tested in cross validation, resulting in an AUROC of 0.774 (SD 0.053), an AUPRC 0.388 (SD 0.084), to be compared with the baseline prevalence of 0.149), and sensitivity and

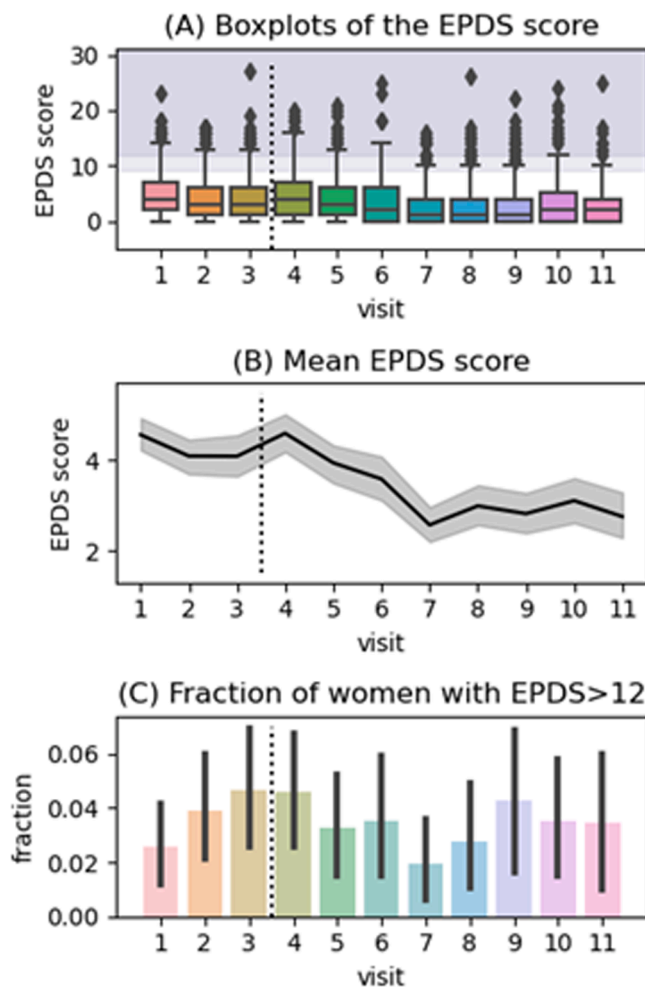


Fig. 2. Panel A): box plots of the EPDS values observed at each visit. The bottom, central and upper lines of the boxes in the plots represent the first quartile, the median and the third quartile, respectively; the whiskers extend to the maximum and minimum values, except for points that are determined to be “outliers” using a method that is a function of the inter-quartile range; diamonds represent the outliers. Panel B): mean values of the EPDS score with bootstrapped 95 % confidence intervals. Panel C): proportion of women with EPDS>12 visit by visit. Vertical dotted line indicates the time of delivery.

specificity equal, respectively, to 0.528 (SD 0.121) and 0.826 (SD 0.023).

The ten predictors selected, based on their permutation importance, by most of the 15 models learned in the cross-validation were: *history of MDD (assessed by MINI)*, *EPDS and VAS scores*, *history of alcohol abuse in the family*, *total number of miscarriages*, *marital status*, *ISI and ESS scores*, *relocation*, *age*, and *neck circumference*. Fig. 3 shows their impact on output of the final model (trained on the entire dataset), as measured by the SHAP values (Lundberg and Lee, 2017). Results showed a positive association with almost all variables except for age. Concerning marital status, where 0 stands for married and 1 for cohabitation, a higher risk was assigned to married women. Finally, the effect of a relocation on PND risk, according to this model, may be positive or negative depending on the value of the other variables, but it must be considered that its weight is rather small on average.

As described in Section 2.2, new classifiers were trained using SVM, one for each visit (from 1 to 7), of the follow-up period, using all the information collected for each patient up to that visit. The performance of the seven resulting classifiers is represented in Fig. 4. Figure S2 of the supplementary materials shows the predictors selected by at least one of the CV models and the number of models by which they were selected.

Details about the datasets used for training each classifier are provided in Table S3 (supplementary material).

4. Discussion

We analyzed the association between several variables collected from a large cohort of women at early gestation and PND, as defined by an EPDS score above 12, and developed a data-driven ML model to predict the risk of developing PND during pregnancy and up to 6 months postpartum.

Among all variables considered, three depression rating scales (EPDS, MADRS, and VAS) showed a strong correlation with PND (significant based on $FDR < 0.05$) and one (HDRS) a weaker one (significant based on $p < 0.05$). Current literature suggests that different clinician-rated and self-rating scales can be useful tools in identifying major depressive episodes during the perinatal period, but the predictive value of these widely used instruments had never been tested with ML techniques before.

We also found that three features of the participants’ psychiatric history were significantly associated with a higher PND risk, i.e., a previous major depressive episode (based on the MINI clinical interview), a self-reported history of depression and antidepressant treatment. MINI-assessed previous MDD, EPDS and VAS scores, as well as self-reported family history of alcohol abuse, were also found to be among the most relevant factors in predicting PND at all visits according to ML-based selection. Moreover, the importance of a previous major depressive episode (based on MINI interview) in predicting PND was confirmed by its inclusion in the SVM model obtained at visit 1 after 10-FS.

Overall, these results highlight the importance of carefully assessing the presence of a prior history of depression, as well as a current depressive episode in women during the first trimester of pregnancy, as key predictors of a later occurring PND. They are in line with current knowledge that a personal history of MDD is one of the strongest risk factors for a new onset of PND (Van Niel and Payne, 2020), and confirm the outcomes of other studies that have used ML techniques to predict postpartum depression, but not PND (Cellini et al., 2022).

Sleep in the Life-ON study was extensively investigated using both subjective (questionnaires) and objective (PSG) tools. Although a vast literature exists on sleep changes during the perinatal period and their correlation with PND (Ross et al., 2005), sleep-derived parameters had never been considered in previous, published ML models for PND prediction. The choice to perform PSG at the end of the second trimester of gestation was a thoughtful decision, motivated by the following reasons: 1) our intention was to examine sleep under the metabolic-hormonal influence of pregnancy per se, without being mainly influenced by mechanical and anatomical factors typically affecting the last trimester of pregnancy; 2) despite smaller in size, other polysomnographic data from the first and the third trimesters of pregnancy are already available in the literature, while the middle period of pregnancy is almost unexplored; 3) in late pregnancy, the sleeping position is generally less variable and rather limited by anatomical conditions; 4) since the main aim of the Life-ON study was to investigate sleep to predict PND, a PSG performed in the third trimester would have a lower predictive value than a PSG conducted in the second trimester.

Interestingly, our analysis revealed that sleep quality (PSQI), insomnia symptoms (ISI), the AI index resulting from PSG recordings and, to a lesser extent, daytime sleepiness (ESS), RLS severity (IRLS) and two other PSG measures (number of hypopneas and the proportion of the N2 sleep stage on total sleep time) were all correlated to PND. Sleep quality and daytime sleepiness also resulted to be important predictors of PND in the trained ML models, with most of them including either the ISI or the ESS score. PSQI, instead, was not selected, despite its association with the PND12 binary variable, probably due to its correlation with the ISI score (Pearson correlation coefficient=0.675, $p < 0.001$). As for the PSG-derived parameters, despite some correlations with the

Table 1

List of selected categorical predictors with p-values of the chi-squared test for the difference between their distribution in the population with PND12 = TRUE and that with PND12 = FALSE. P-values smaller than 0.05 are highlighted in bold. P-values which are significant according to the Benjamini-Hochberg procedure with FDR<0.05 (Benjamini and Hochberg, 1995) are indicated by the asterisks. The table also shows the values taken by the categorical predictors, with the absolute frequencies of each category both for the total sample and for the subset with PND12=FALSE.

Variable	Values	n	n PDE=False (%)	p- value	Variable	Values	n	n PDE=False (%)	p-value	
Demographic variables					Psychiatric assessment					
Marital status	Marriage	217	40	18.4 %	History of depression ⁽³⁾	No	319	38	11.9%	<0.001*
	Cohabitation	155	14	9.0 %		Yes	57	18	31.6%	
	Single or divorced	4	2	50.0 %		Major depressive episode in anamnesis ⁽⁶⁾	No	320	39	
Working condition	Unemployed	243	35	14.4 %	Yes	56	17	30.4%		
	Temporary job	64	6	9.4 %	History of antidepressant treatment ⁽³⁾	No	344	45	13.1%	0.003*
	Permanent job	48	14	29.2 %		Yes	32	11	34.4%	
Loss or change of job ⁽¹⁾	No	344	48	14.0 %	Family history of alcohol or substance abuse ⁽⁵⁾	No	316	44	13.9%	0.311
	Yes	32	8	25.0 %		Yes	60	12	20.0%	
Perception of poverty ⁽²⁾	No	368	53	14.4 %	Family history of suicide ⁽⁵⁾	No	351	54	15.4%	0.477
	Yes	8	3	37.5 %		Yes	25	2	8.0%	
Housing	Rent or other	128	22	17.2 %	History of other psychiatric disorders ⁽³⁾	No	314	45	14.3%	0.621
	Owner	248	34	13.7 %		Yes	62	11	17.7%	
Relocation ⁽¹⁾	No	336	46	13.7 %	Family history of depression ⁽⁵⁾	No	244	38	15.6%	0.725
	Yes	40	10	25.0 %		Yes	132	18	13.6%	
Educational level	Primary or lower secondary school	19	2	11.1 %	Sleep					
	Upper secondary school	106	19	17.9 %	History of RLS symptoms	No	332	46	13.9%	0.184
	University degree	249	35	14.1 %		Yes	44	10	22.7%	
Physical activity	None	236	34	14.4 %	Gynecological assessment					
	Occasional (<3 times/week)	38	4	10.5 %	Type of gestation	Planned	309	45	14.6%	0.726
	Regular (≥3 times/week)	95	17	17.9 %		Unplanned	64	11	17.2%	
Tobacco use (past) ⁽³⁾	No	247	41	16.6 %		Unintended	2	0	0%	
	Yes	129	15	11.6 %	Conception type	Natural	358	54	15.1%	0.902
Tobacco use (current)	No	355	54	15.2 %		Medically assisted	18	2	11.1%	
	Yes	16	2	12.5 %	⁽¹⁾ Previous 6 months					
Death of a first degree relative ⁽¹⁾	No	365	52	14.2 %	⁽²⁾ Subjectively assessed by answering the question: "Do you consider yourself poor?"					
	Yes	11	4	36.4 %	⁽³⁾ Before pregnancy					
Illness of a first degree relative ⁽¹⁾⁽⁴⁾	No	347	51	14.7 %	⁽⁴⁾ Requiring hospitalization					
	Yes	29	5	17.2 %	⁽⁵⁾ 1st or 2nd grade relatives					
					⁽⁶⁾ Assessed by the Mini International Neuropsychiatric Interview (MINI)					

outcome variable, they did not improve the predictive performance of the SVM model when included among the predictors.

Pre-pregnancy RLS, especially in its severe form, has been previously associated with a higher risk of PND (Wesström et al., 2014). In our study, prior RLS symptoms and RLS severity emerged as useful predictors of PND only at visit 3, which also represents the time point where RLS incidence is the highest.

In summary, the ML-guided analysis of sleep variables collected during pregnancy indicates that poor sleep quality, insomnia and daytime sleepiness symptoms in early pregnancy are strong predictors of PND during the peripartum (Figure S2). This suggests that an early

assessment of sleep changes during early gestation using three easy-to-administer questionnaires, such as the PSQI, the ISI and the ESS, should be considered as important as a psychiatric evaluation to identify those who are at higher risk of developing PND.

Regarding the clinical, gynecological and demographic variables considered in the analysis, a positive correlation was found between neck circumference, marital status, working condition, total number of miscarriages and PND. While the association between neck circumference and PND risk is difficult to interpret, especially in a population of women mostly without clinically significant OSA, the other variables clearly indicate the well-known role of psychosocial factors in causing

Table 2

List of selected numerical predictors with p-values of the t-test for the difference between their mean in the population with PND12 = TRUE and that with PND12 = FALSE. P-values smaller than 0.05 are highlighted in bold. P-values which result significant according to the Benjamini-Hochberg procedure with FDR < 0.05 (Benjamini and Hochberg, 1995) are indicated by the asterisks. The table also shows the mean and standard deviation of the predictors in the two groups. All acronyms are defined in Table S4 of the supplementary materials. Psychiatric assessment: variables reporting historical notions ("history of") are based on the patient's subjective recall; variables reporting anamnestic notions ("major depressive episode in anamnesis") are based on structured clinical interviews (MINI).

Variable	n (n _T)	[min-max]	mean ± st.dev.		p-value	Variable	n (n _T)	[min-max]	mean ± st.dev.		p-value
			PND12=FALSE	PND12=TRUE					PND12=FALSE	PND12=TRUE	
Demographic variables											
Age	376 (56)	[19–49]	33.65 ± 4.04	33.55 ± 5.24	0.871	Sleep efficiency (SE)	332 (48)	[0.422–0.971]	0.831 ± 0.102	0.847 ± 0.080	0.287
Psychiatric assessment						Mean ECG frequency	332 (48)	[51.7–180.9]	71.48 ± 12.65	68.73 ± 16.19	0.183
EPDS score	374 (55)	[0–23]	3.84 ± 3.22	7.40 ± 4.84	<0.001*	Polysomnographies - Apneas					
ASSESSMENT MADRS	372 (56)	[0–29]	3.47 ± 3.28	5.68 ± 4.67	<0.001*	Apnea index (AI)	332 (48)	[0–0.9]	0.039 ± 0.102	0.098 ± 0.185	0.001*
ASSESSMENT VAS	372 (56)	[0–8]	0.621 ± 1.059	1.693 ± 2.082	<0.001*	Central hypopneas	332 (48)	[0–0.6]	0.016 ± 0.059	0.042 ± 0.096	0.013
HDRS	370 (56)	[0–14]	1.94 ± 2.36	2.82 ± 3.2	0.015	Mixed hypopneas	332 (48)	[0–0.6]	0.006 ± 0.038	0.0146 ± 0.0875	0.298
Sleep						Snoring time (%)	332 (48)	[0–74.2]	1.64 ± 6.49	3.30 ± 13.25	0.174
ISI	371 (55)	[0–21]	4.87 ± 4.42	8.36 ± 5.42	<0.001*	SpO2 (mean)	332 (48)	[84.8–98.1]	94.74 ± 10.06	92.61 ± 19.75	0.254
PSQI	368 (53)	[0–18]	4.42 ± 2.77	6.32 ± 3.28	<0.001*	SpO2 (min)	332 (48)	[84.8–97]	90.54 ± 10.55	88.58 ± 19.07	0.303
ESS	371 (55)	[0–23]	8.52 ± 3.89	9.80 ± 4.53	0.028	RERA index	332 (48)	[0–1.8]	0.022 ± 0.165	0.025 ± 0.108	0.898
IRLS	366 (55)	[0–23]	1.52 ± 4.45	2.98 ± 6.14	0.036	AHI	332 (48)	[0–15.6]	1.33 ± 1.84	1.49 ± 1.65	0.595
MEQ	365 (51)	[27–80]	55.9 ± 9.03	55.0 ± 9.38	0.513	AHI supine	332 (48)	[0–35.2]	2.11 ± 3.7	2.33 ± 2.96	0.696
Gynecological anamnesis						AHI (REM)	332 (48)	[0–54]	4.74 ± 5.95	6.25 ± 7.40	0.119
Number of previous pregnancies	374 (56)	[0–7]	0.840 ± 1.015	1.125 ± 1.192	0.060	AHI (NREM)	332 (48)	[0–11.9]	0.463 ± 1.088	0.419 ± 0.733	0.785
Number of miscarriages	372 (55)	[0–5]	0.435 ± 0.733	0.691 ± 1.033	0.026	RDI	332 (48)	[0–15.6]	1.356 ± 1.846	1.515 ± 1.633	0.576
Clinical variables						ODI≥3 (%)	332 (48)	[0–18.2]	0.792 ± 1.815	0.808 ± 1.247	0.954
BMI kg/m2	356 (50)	[16.16–35.16]	22.78 ± 3.36	23.63 ± 3.82	0.105	ODI≥4 (%)	332 (48)	[0–7.5]	0.200 ± 0.696	0.313 ± 0.580	0.289
Abdominal circumference (cm)	366 (53)	[60–116]	86.69 ± 10.55	86.18 ± 10.74	0.746	Polysomnographies - Stages					
Neck circumference (cm)	366 (54)	[28–39]	32.36 ± 1.96	33.28 ± 2.37	0.002*	Percent of N1	332 (48)	[0.023–0.263]	0.111 ± 0.040	0.110 ± 0.041	0.903
Systolic blood pressure (BP) at rest (mmHg)	363 (53)	[80–167]	103.29 ± 10.88	105.7 ± 14.91	0.161	Percent of N2	332 (48)	[0.230–0.651]	0.460 ± 0.065	0.485 ± 0.071	0.017
Diastolic BP at rest (mmHg)	363 (53)	[50–90]	65.5 ± 7.91	66.42 ± 8.57	0.443	Percent of REM	332 (48)	[0.082–0.317]	0.201 ± 0.039	0.202 ± 0.044	0.839
Heart rate (bpm)	366 (53)	[52–112]	74.63 ± 9.06	74.3 ± 7.79	0.802	SL-N1 (min)	332 (48)	[0.4–133]	16.2 ± 18.1	18.1 ± 20.6	0.508
Blood serum						SL-N2 (min)	332 (48)	[0–136.5]	18.82 ± 18.94	21.18 ± 21.06	0.433
Prolactin (ng/ml)	340 (50)	[5.452–239.6]	55.23 ± 35.57	61.02 ± 41.58	0.301	SL-REM (min)	332 (48)	[4.5–265]	84.49 ± 40.00	81.23 ± 40.70	0.603
Progesterone (nmol/l)	346 (51)	[40.9–448.1]	111.02 ± 40.73	113.88 ± 45.75	0.65	Arousal index (ARI)	332 (48)	[0–29.6]	12.28 ± 4.23	11.03 ± 4.35	0.06
Estrogen (pmol/l)	331 (47)	[20–41.309]	1.23e4 ± 0.65e4	1.26e4 ± 0.77e4	0.783	Autonomic ARI (AAI)	332 (48)	[0–92]	22.61 ± 21.37	20.8 ± 25.17	0.597
Ferritin (nmol/l)	346 (51)	[0.00–0.584]	0.091 ± 0.080	0.087 ± 0.080	0.711	WASO	332 (48)	[6.6–285]	66.04 ± 46.32	57.83 ± 36.74	0.244
Iron (µmol/l)	344 (52)	[3.58–42.63]	18.27 ± 6.59	18.32 ± 6.28	0.961	Awakenings (n)	332 (48)	[7–58]	23.94 ± 8.67	23.88 ± 9.25	0.964
Polysomnographies - general variables						Polysomnographies - Movements					
TST (min)	332 (48)	[158.5–534.5]	397.9 ± 64.3	401.3 ± 60.6	0.735	Movement index (MI)	332 (48)	[0.4–151]	21.81 ± 18.95	23.19 ± 18.88	0.640
Time supine (min)	332 (48)	[0–396.5]	151.1 ± 91.4	142.2 ± 84.6	0.532	Periodic limb MI (PMI)	332 (48)	[0–141.6]	10.14 ± 17.45	11.46 ± 17.77	0.630
Time supine (%)	332 (48)	[0–100]	0.376 ± 0.220	0.355 ± 0.213	0.541	Respiratory-related MI	332 (48)	[0–2.3]	0.109 ± 0.267	0.158 ± 0.292	0.239
						Arousal-related PMI	332 (48)	[0–4.5]	0.471 ± 0.752	0.531 ± 0.944	0.621

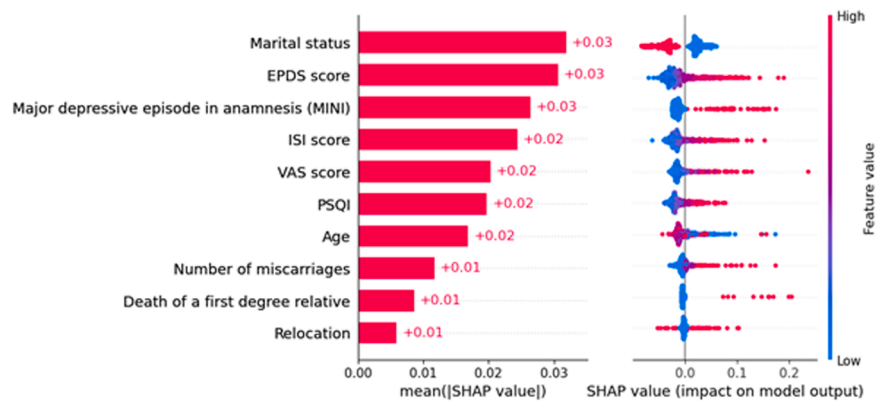


Fig. 3. (Left) Importance of the 10 selected predictors illustrated by their mean absolute SHAP value. (Right) Direction of the relationship between each predictor and the PND outcome described through the local SHAP values.

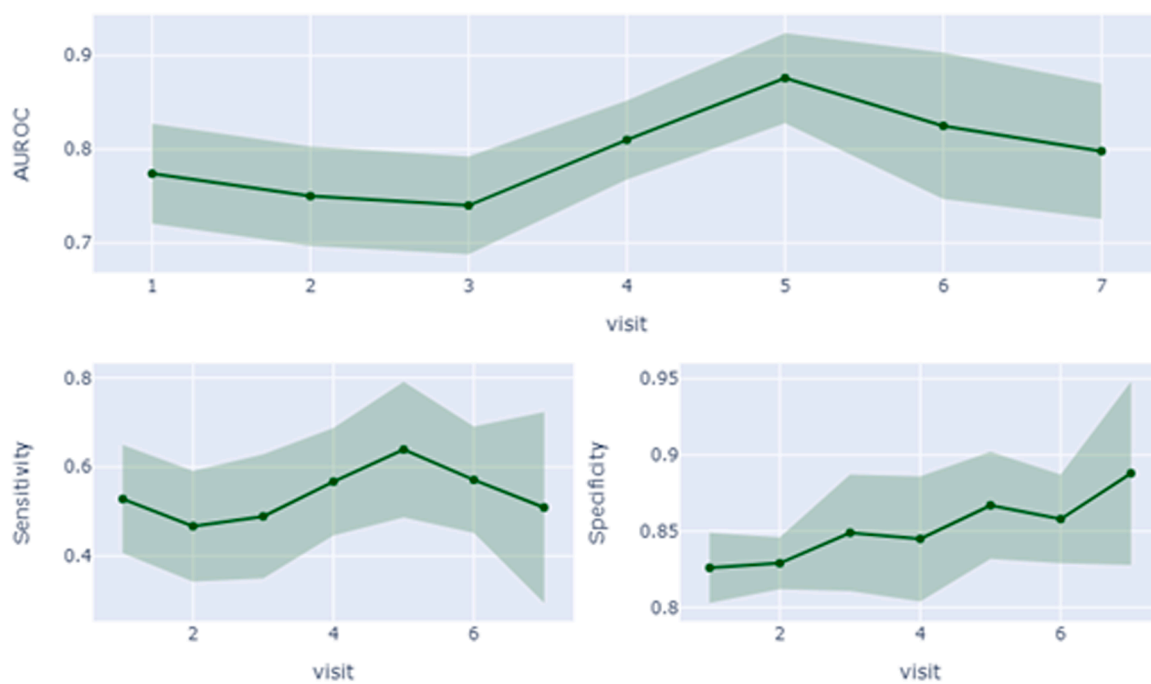


Fig. 4. AUROC (top) and specificity and sensitivity (bottom) of the SVM classifier retrained at each visit with all the available information. The shaded area represents the standard deviation of the performance metric over the 15 cross-validation models.

PND. Interestingly, number of miscarriages and marital status were found to be relevant predictors of PND only at the beginning of the pregnancy, whereas loss or change of job showed a relevant impact also during the puerperium. By looking at the variables selected by ML models, the importance of age as a predictor remains uncertain, as it was used by almost all models only at visits 1 and 3. The same applies to other clinical predictors, which were rarely selected by the models at any study visits, the most used ones being neck circumference (visits 1 and 2) and serum iron (visit 3).

A recent review of studies using ML-models to predict postpartum depression (PPD) (Cellini et al., 2022) found the following parameters to be the strongest predictors of PPD: *Age, Education, Marital status, Income, Ethnicity, Depression lifetime, Depression during pregnancy, Anxiety, Smoking, Mode of delivery, Gestational age at delivery, APGAR score, BMI, Antidepressant use.*

Interestingly, besides the presence of depression pre- or during pregnancy, *marital status* and *age* were also included among the predictors in our model for visit 1. In the “Life-ON” study, *marital status* was assessed, but only distinguished between married and cohabiting

couples, due to the limited number of singles, and a higher risk of PND was observed for married women. By contrast, *education, smoking, and BMI* were not selected as relevant predictors in our models, probably due to the peculiar characteristics of our study population, which was recruited in 3 wealthy cities of northern Italy and one in Switzerland, and was generally characterized by healthy lifestyle habits, low BMI, and middle-high socioeconomic status. In fact, although the demographic variables *working condition* and *perception of poverty* were also considered in our analysis, they did not emerge as relevant predictors.

The final SVM classifier utilizes the values of the 10 main risk factors listed in Fig. 3 observed for a woman during her first trimester of pregnancy to compute an index associated with her probability of experiencing symptoms of PND. The algorithm used to train it achieved, in cross-validation, a mean AUROC of 0.774 and an AUPRC of 0.388. We estimated that, in clinical practice, such classifier could help identify (and refer for treatment) 52.8 % (sensitivity) of women at risk of PND from data collected during the first trimester of pregnancy, before they reach an EPDS score >12 during the peripartum, at the price of unnecessarily treating only 17.4 % of other (presumably healthy) women.

These findings are in line with those summarized in the review by Cellini et al. [8] where the AUC achieved by five studies focusing on outcomes based on the EPDS ranged between 0.75 and 0.83. Among them, the study which most closely resembles our analysis is that by Andersson et al. (Andersson et al., 2021), which defines depression as EPDS \geq 12 and includes in the list of predictors *psychiatric diagnoses, psychological factors, obstetric and sociodemographic factors*. Despite being based on a larger dataset, including 3736 healthy women and 577 with PPD symptoms, this study achieved an AUC between 0.79 and 0.81 and, for specificities between 0.81 and 0.89, obtained sensitivities between 0.59 and 0.51. By comparison, our study shows that ML models applied to the whole perinatal period, can achieve similar predictive performances as those addressing only the postnatal period and with a much smaller size of the training sample, suggesting that the Life-ON study achieved a sufficiently large sample size to capture the relevant associations between variables.

When adding to the model the information collected during the follow-up visits (Fig. 4), we observed that only specificity has a clear positive trend, while the AUROC and sensitivity reached a maximum when using data from visits 1 to 5, i.e., between the first trimester of gestation and 3 weeks after delivery. Thereafter, a decrease in performance with data from visits 6 and 7, was evident. However, this observation may follow from the increasing dropout rates during postpartum, leading to less observations being available to train and validate these later models.

The dropouts and the amount of missing data should be considered a study limitation, as they may introduce biases in the analysis. In fact, only 177 of the 439 study participants provided a complete EPDS questionnaire in all the 9 visits considered and for only 201 of the 376 women included in the ML analysis all the selected predictors were available. Moreover, from the data collected, it emerges that the majority of women who participated in the Life-ON study had good general health, with low BMI and clinically irrelevant sleep apnea, a stable working condition, and a high level of education. This likely reflects the socioeconomic status of the population of four wealthy cities in northern Italy and Switzerland, but may represent a bias when trying to generalize the results on a large scale or to countries with different socio-demographic characteristics. Similarly, given that most of the participating women were Caucasian, the study is lacking in terms of racial diversity.

A further limitation is represented by the fact that the “Life-ON” project did not evaluate some variables that have previously been found to predict PPD, such as pregnancy-related and pediatric complications, anxiety, neuroticism, and income. In particular, given the evidence that antenatal anxiety has been shown to be an important predictor of postnatal anxiety and mood disorders (Grant et al., 2008) assessing anxiety disorder, traits, or neuroticism by either clinical interview or maternal self-report (eg. using the State-Trait Anxiety Inventory) could be a future direction to explore for machine learning model studies exploring predictive factors for PND. Finally, another limitation is that all the assessment questionnaires and scales used have been validated in a mixed gender population, with the exception of the EPDS scale, which has been validated in pregnant women.

In conclusion, we found that PND symptoms throughout the perinatal period can be predicted by using ML methods. In our ML model, a wide range of data collected during early pregnancy, including several sleep variables, were used to identify women at higher risk of developing PND symptoms up to 6 months after delivery. Besides confirming the importance of psychiatric and sociodemographic variables, our analysis showed that insomnia symptoms, subjective poor sleep quality and daytime sleepiness are among the strongest predictors of PND. As these conditions can be easily assessed early or during pregnancy by administering self-report questionnaires that can be quickly interpreted even by clinicians who are not experts in sleep medicine, this may help to proactively identify women who are at increased risk of developing PND and to support them with targeted therapeutic strategies. While some

PSG variables were also associated with an increased risk of PND, they did not globally improve model performance, so their role needs further investigation.

Overall, the advantage of our ML tool is that it utilizes a variety of diagnostic assessments and patient features to generate a more accurate and personalized risk assessment compared to a single clinical assessment questionnaire. As a next step, the ML-based predictive model designed may be used by clinicians through an application that gathers the set of selected predictors from the patient and predicts her risk of PND, e.g. by computing a risk score.

Funding sources

The Italian centers participating in the “Life-ON” study were supported by a grant from the Italian Ministry of Health and the Emilia-Romagna region (PE-2011-02348727). The Sleep Medicine Unit of Lugano was supported by the Swiss National Science Foundation (grant 320030_160250). Philips Respironics donated the light boxes for light therapy and the actigraphic devices. Vifor-Pharma contributed to the costs of blood collection and analysis with an unrestricted grant.

Ethics

The “Life-ON” study has been approved by the respective ethics committee of the four participating centers in Italy and Switzerland. All participants gave a written informed consent prior to study entry.

Authors statement

This statement is to certify that all authors have seen and approved the final version of the manuscript being submitted.

CRediT authorship contribution statement

Corrado Garbazza: Writing – review & editing, Writing – original draft, Methodology, Investigation, Data curation, Conceptualization. **Francesca Mangili:** Writing – review & editing, Writing – original draft, Software, Methodology, Formal analysis, Data curation, Conceptualization. **Tatiana Adele D’Onofrio:** Writing – review & editing, Software, Methodology, Formal analysis, Data curation. **Daniele Malpetti:** Writing – review & editing, Software, Methodology, Formal analysis, Data curation. **Silvia Riccardi:** Writing – review & editing, Visualization, Supervision, Methodology, Investigation. **Alessandro Cicolin:** Writing – review & editing, Visualization, Supervision, Resources, Project administration, Funding acquisition. **Armando D’Agostino:** Writing – review & editing, Visualization, Supervision, Resources, Project administration, Investigation, Funding acquisition. **Fabio Cirignotta:** Writing – review & editing, Visualization, Supervision, Resources, Project administration, Funding acquisition. **Mauro Manconi:** Writing – review & editing, Visualization, Supervision, Resources, Project administration, Funding acquisition, Conceptualization.

Declaration of competing interest

None

Supplementary materials

Supplementary material associated with this article can be found, in the online version, at [doi:10.1016/j.psychres.2024.115957](https://doi.org/10.1016/j.psychres.2024.115957).

References

- Andersson, S., Bathula, D.R., Iliadis, S.I., Walter, M., Skalkidou, A., 2021. Predicting women with depressive symptoms postpartum with machine learning methods. *Sci. Rep.* 11, 7877. <https://doi.org/10.1038/s41598-021-86368-y>.

- Baiardi, S., Cirignotta, F., Cicolin, A., Garbaza, C., D'Agostino, A., Gambini, O., Giordano, A., Canevini, M., Zambrelli, E., Marconi, A.M., Mondini, S., Borgwardt, S., Cajochen, C., Rizzo, N., Manconi, M., 2016. Chronobiology, sleep-related risk factors and light therapy in perinatal depression: the "Life-ON" project. *BMC Psychiatr.* 16, 374. <https://doi.org/10.1186/s12888-016-1086-0>.
- Benjamini, Y., Hochberg, Y., 1995. Controlling the false discovery rate: a practical and powerful approach to multiple testing. *J. R. Stat. Soc. Ser. B Methodol.* 57, 289–300. <https://doi.org/10.1111/j.2517-6161.1995.tb02031.x>.
- Breiman, L., 2001. Random forests. *Mach. Learn.* 45, 5–32. <https://doi.org/10.1023/A:1010933404324>.
- Buuren, S.van, Groothuis-Oudshoorn, K., 2011. mice: multivariate imputation by chained equations in R. *J. Stat. Softw.* 45, 1–67. <https://doi.org/10.18637/jss.v045.i03>.
- Cellini, P., Pigoni, A., Delvecchio, G., Moltrasio, C., Brambilla, P., 2022. Machine learning in the prediction of postpartum depression: a review. *J. Affect. Disord.* 309, 350–357. <https://doi.org/10.1016/j.jad.2022.04.093>.
- Cox, J.L., Chapman, G., Murray, D., Jones, P., 1996. Validation of the edinburgh postnatal depression scale (EPDS) in non-postnatal women. *J. Affect. Disord.* 39, 185–189. [https://doi.org/10.1016/0165-0327\(96\)00008-0](https://doi.org/10.1016/0165-0327(96)00008-0).
- Dagher, R.K., Bruckheim, H.E., Colpe, L.J., Edwards, E., White, D.B., 2021. Perinatal depression: challenges and opportunities. *J. Womens Health* 30, 154–159. <https://doi.org/10.1089/jwh.2020.8862>, 2002.
- Dwyer, D.B., Falkai, P., Koutsouleris, N., 2018. Machine learning approaches for clinical psychology and psychiatry. *Annu. Rev. Clin. Psychol.* 14, 91–118. <https://doi.org/10.1146/annurev-clinpsy-032816-045037>.
- Gibson, J., McKenzie-McHarg, K., Shakespeare, J., Price, J., Gray, R., 2009. A systematic review of studies validating the edinburgh postnatal depression scale in antepartum and postpartum women. *Acta Psychiatr. Scand.* 119, 350–364. <https://doi.org/10.1111/j.1600-0447.2009.01363.x>.
- Grant, K.-A., McMahon, C., Austin, M.-P., 2008. Maternal anxiety during the transition to parenthood: a prospective study. *J. Affect. Disord.* 108, 101–111. <https://doi.org/10.1016/j.jad.2007.10.002>.
- Howard, L.M., Khalifeh, H., 2020. Perinatal mental health: a review of progress and challenges. *World Psychiatry Off. J. World Psychiatr. Assoc. WPA* 19, 313–327. <https://doi.org/10.1002/wps.20769>.
- Lundberg, S.M., Lee, S.-I., 2017. A unified approach to interpreting model predictions. In: *Proceedings of the 31st International Conference on Neural Information Processing Systems, NIPS'17*. Curran Associates Inc., Red Hook, NY, USA, pp. 4768–4777.
- Noble, W.S., 2006. What is a support vector machine? *Nat. Biotechnol.* 24, 1565–1567. <https://doi.org/10.1038/nbt1206-1565>.
- Rajkumar, A., Oren, E., Chen, K., Dai, A.M., Hajaj, N., Hardt, M., Liu, P.J., Liu, X., Marcus, J., Sun, M., Sundberg, P., Yee, H., Zhang, K., Zhang, Y., Flores, G., Duggan, G.E., Irvine, J., Le, Q., Litsch, K., Mossin, A., Tansuwan, J., Wang, D., Wexler, J., Wilson, J., Ludwig, D., Volchenboum, S.L., Chou, K., Pearson, M., Madabushi, S., Shah, N.H., Butte, A.J., Howell, M.D., Cui, C., Corrado, G.S., Dean, J., 2018. Scalable and accurate deep learning with electronic health records. *NPJ Digit. Med.* 1, 18. <https://doi.org/10.1038/s41746-018-0029-1>.
- Ross, L.E., Murray, B.J., Steiner, M., 2005. Sleep and perinatal mood disorders: a critical review. *J. Psychiatr. Neurosci. Jpn.* 30, 247–256.
- Shatte, A.B.R., Hutchinson, D.M., Teague, S.J., 2019. Machine learning in mental health: a scoping review of methods and applications. *Psychol. Med.* 49, 1426–1448. <https://doi.org/10.1017/S0033291719000151>.
- Van Niel, M.S., Payne, J.L., 2020. Perinatal depression: a review. *Cleve. Clin. J. Med.* 87, 273–277. <https://doi.org/10.3949/ccjm.87a.19054>.
- Wesström, J., Skalkidou, A., Manconi, M., Fulda, S., Sundström-Poromaa, I., 2014. Pre-pregnancy restless legs syndrome (Willis-Ekbom Disease) is associated with perinatal depression. *J. Clin. Sleep Med. JCSM Off. Publ. Am. Acad. Sleep Med.* 10, 527–533. <https://doi.org/10.5664/jcsm.3704>.
- Woody, C.A., Ferrari, A.J., Siskind, D.J., Whiteford, H.A., Harris, M.G., 2017. A systematic review and meta-regression of the prevalence and incidence of perinatal depression. *J. Affect. Disord.* 219, 86–92. <https://doi.org/10.1016/j.jad.2017.05.003>.
- Yawn, B.P., Olson, A.L., Bertram, S., Pace, W., Wollan, P., Dietrich, A.J., 2012. Postpartum depression: screening, diagnosis, and management programs 2000 through 2010. *Depress. Res. Treat.* 2012, 363964. <https://doi.org/10.1155/2012/363964>.