



HHS Public Access

Author manuscript

J Affect Disord. Author manuscript; available in PMC 2022 January 15.

Published in final edited form as:

J Affect Disord. 2021 January 15; 279: 1–8. doi:10.1016/j.jad.2020.09.113.

Development and Validation of a Machine Learning Algorithm for Predicting the Risk of Postpartum Depression among Pregnant Women

Yiye Zhang, PhD MS^{a,*†}, Shuojia Wang, MS^{b,c,†}, Alison Hermann, MD^a, Rochelle Joly, MD^a, Jyotishman Pathak, PhD^a

^aWeill Cornell Medicine, Cornell University, New York, NY, USA

^bSchool of Public Health, Zhejiang University, HangZhou, Zhejiang, China

^cTencent Jarvis Lab, Shenzhen, China

Abstract

Objective—There is a scarcity in tools to predict postpartum depression (PPD). We propose a machine learning framework for PPD risk prediction using data extracted from electronic health records (EHRs).

Methods—Two EHR datasets containing data on 15,197 women from 2015 to 2018 at a single site, and 53,972 women from 2004 to 2017 at multiple sites were used as development and validation sets, respectively, to construct the PPD risk prediction model. The primary outcome was a diagnosis of PPD within 1 year following childbirth. A framework of data extraction, processing, and machine learning was implemented to select a minimal list of features from the EHR datasets to ensure model performance and to enable future point-of-care risk prediction.

Results—The best-performing model uses from clinical features related to mental health history, medical comorbidity, obstetric complications, medication prescription orders, and patient demographic characteristics. The model performances as measured by area under the receiver operating characteristic curve (AUC) are 0.937 (95% CI 0.912 – 0.962) and 0.886 (95% CI 0.879–0.893) in the development and validation datasets, respectively. The model performances were consistent when tested using data ending at multiple time periods during pregnancy and at childbirth.

Limitations—The prevalence of PPD in the study data represented a treatment prevalence and is likely lower than the illness prevalence.

*Corresponding author: Yiye Zhang, PhD, MS, 425 East 61st Street, New York, NY, 10065, USA. yiz2014@med.cornell.edu, Phone: (646) 962-9437.

†Both authors contributed equally to the manuscript

Publisher's Disclaimer: This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

Conflict of Interest Disclosures

Authors AH, RJ, JP, and YZ have equity ownership at Iris OB Health Inc., New York. AH serves as a consultant for Sage Therapeutics.

Conclusions—EHRs and machine learning offer the ability to identify women at risk for PPD early in their pregnancy. This may facilitate scalable and timely prevention and intervention, reducing negative outcomes and the associated burden.

Keywords

Postpartum depression; machine learning; electronic health records

1. Introduction

Postpartum depression (PPD) is a potentially life-threatening mental health condition that occurs up to one year following childbirth (Stewart and Vigod, 2016). The prevalence of PPD is estimated to affect as many as 1 in 7 mothers in the US (Hahn-Holbrook et al., 2017; Wisner et al., 2013), but underdiagnosis and lack of treatment for PPD are common, especially among women with low socioeconomic status (Biaggi et al., 2016; O'Connor et al., 2019). Long-term health effects of PPD to mothers, children, and family include increased maternal and infant mortality, increased hospitalizations, impaired mother-child bonding, and impaired long-term child development (Field, 2010; Jacques et al., 2019; Stein et al., 2014; Weobong et al., 2015). The disease mechanism of PPD is multifactorial. Clinically, a history of mental illness is the most significant risk factor (Meltzer-Brody et al., 2018; Stewart and Vigod, 2016). Social determinants of health (SDoH), including poor marital relationship, low socioeconomic status, and stressful life events are also known contributors to increased PPD risk (Biaggi et al., 2016). New research indicates that there may be additional biomarkers associated with the risk for developing PPD such as excessive proinflammatory immune system activation, possible disruptions in fatty acid metabolism, disruptions in hypothalamic-pituitary-adrenal (HPA) functioning, altered neurosteroid physiology, and genetic and epigenetic signatures (Serati et al., 2016).

The importance of PPD prevention and timely intervention cannot be overstated. The American College of Obstetricians and Gynecologists (ACOG) (Committee, 2018), the American Academy of Pediatrics (AAP) (Earls et al., 2019), the US Preventive Services Task Force (O'Connor et al., 2019), and several other organizations (Stewart and Vigod, 2016) have guidelines and recommendations for universal PPD screening as part of usual care during pregnancy and the postpartum period. Current PPD prevention strategies focus on secondary rather than primary prevention, using questionnaire-based screening instruments such as the Edinburgh Postnatal Depression Scale (EPDS) (Cox et al., 1987) and Patient Health Questionnaire-9 (PHQ-9) (Löwe et al., 2004) to detect symptoms. Primary prevention techniques intervene in an illness course prior to symptom onset while secondary prevention techniques intervene soon after the symptom onset, but prior to the full manifestation of the illness. Unfortunately, it has been demonstrated that in women known to be at high risk of PPD, delaying intervention until the onset of symptoms only mildly attenuates risk for depression, while intervening with appropriately targeted prevention before the onset of symptoms substantially mitigates depression relapse risk (Cohen et al., 2006). In addition to being “too little, too late” from a clinical perspective, these screening tools present major feasibility problems for both large and smaller health systems (Beck and Gable, 2000; Gjerdingen and Yawn, 2007). In order to come into compliance with current

screening recommendations, obstetric practices often require substantial change, including not just changes to clinical workflows, but also staffing changes, new electronic health records (EHR) workflow builds, collaboration with referral networks, and investment in staff and provider training. Even then, further challenges persist such as mental health-related stigma, limitations in provider time, attention, and expertise, and scarcity in specialized mental health treatment resources.

We argue that taking a primary prevention approach has the promise of reducing the investment and resources required to address PPD while at the same time reducing the incidence of PPD rates. In this work, to identify signals that may suggest elevated future risk of PPD, we propose a primary prevention approach that is data-driven, leveraging machine learning applications to EHR data (Loudon et al., 2016; Wang et al., 2019). EHR data can be collected and analyzed routinely on a large scale using machine learning, as demonstrated by successful data-driven clinical decision support (Shortliffe and Sepúlveda, 2018) applications that assist with decision making across clinical conditions (Goldstein et al., 2017; Liang et al., 2019; Rajkomar et al., 2018; Tomašev et al., 2019).

We developed an end-to-end framework (Fig. 1) to extract features from EHR data for processing, including demographics, clinical diagnoses, medication prescriptions, laboratory results, and unstructured clinical notes. These data are sent to an optimization process to select important features and incorporated in multiple machine learning algorithms including regularized logistic regression, random forest, decision tree, extreme gradient boosting (XGboost), and multilayer perceptron (MLP) (Bishop, 2006) to predict the risk of PPD. The framework was implemented and evaluated using data available at different time intervals during pregnancy (12-week, 18-week, 24-week, and 30-week) during pregnancy and after childbirth.

We aim to demonstrate that the data-driven primary intervention approach provides an opportunity for individualized therapeutic interventions such as changing screening timelines, engaging with appropriate preventive strategies, or tailoring clinician PPD counseling time according to a patient need. To the best of our knowledge, this study is among the first in developing an EHR-based machine learning framework for identifying women at risk for PPD (Jiménez-Serrano et al., 2015; Tortajada et al., 2009; Wang et al., 2019; Zhang et al., 2020).

2. Methods

2.1 Inclusion Criteria

All pregnant women with fully completed antenatal care procedures who had live births of infants were included in the study. The exclusion criteria were (1) maternal age below 18 or above 45, or (2) lack of outpatient, inpatient or emergency room encounter information in the EHR data within 1 year following childbirth. Participants with a prior history of mental illness and participants with active mental illness were not excluded to ensure clinical applicability in real implementation (Fig. 2). The study was approved by the Institutional Review Board at Weill Cornell Medicine (IRB protocol# 1711018789). Data extraction and analysis were performed in 2019.

2.2 Study Design

2.2.1 Outcome—The outcome is defined as having a diagnosis of PPD within 1 year of childbirth. A PPD diagnosis was defined using Systematized Nomenclature of Medicine (SNOMED) codes and the use of antidepressants within 1 year following childbirth (Dietz et al., 2007; Stewart and Vigod, 2016). The specific SNOMED codes for PPD definition are listed in Appendix (Table A1). The use of antidepressants was defined by Anatomical Therapeutic Chemical (ATC) codes under N06A (Petersen et al., 2018). To ensure that antidepressants were primarily used for treatment of mental health conditions, and not for other indications such as pain, we further excluded the following medications: *Amitriptyline, Clomipramine, Duloxetine, Flupentixol, and Nortriptyline* (Schofield et al., 2016).

2.2.2 Data Sources—For algorithm development, EHR data including demographics, diagnoses, medication prescriptions, procedures, laboratory measurements, and social determinants of health (SDoH) including the built environment characteristics such as distance to public transportation and green space on eligible patients were obtained at Weill Cornell Medicine (WCM) and NewYork-Presbyterian Hospital in New York City, USA between January 2015 and June 2018. For algorithm validation, EHR data was derived from multiple health systems across New York City affiliated to the Patient-Centered Outcomes Research Institute funded New York City Clinical Data Research Network data (NYC-CDRN) between August 2004 and October 2017 (Kaushal et al., 2014). We randomly selected 80% of the data from WCM as the training set including cross-validation and model tuning, and held the remaining 20% as the test set individually. The NYC-CDRN data was used solely as a validation set.

Both datasets were represented using Observational Medical Outcomes Partnership (OMOP) Common Data Model to record patient demographics, encounter records, diagnostic codes, procedures, prescription medications, and laboratory measurements (Overhage et al., 2012). Diagnoses, laboratory measurements, and procedures are represented as SNOMED codes, Logical Observation Identifiers Names and Codes (LOINC), and Current Procedural Terminology (CPT) codes, respectively. Medications were standardized using the ATC classification system. In addition, marital status was extracted from unstructured clinical notes using regular expression-based searches, and individuals were classified as married or not married (single/divorced/widowed) at the time of childbirth. Age was calculated as the time difference between childbirth and delivery dates. Mental health history before pregnancy was defined as having at least one diagnosis including organic disorders, substance-related disorders, schizophrenic/psychotic disorders, mood disorders, anxiety disorders, personality disorders, and other psychiatric disorders (Canada, 2015). Features with frequencies below 10 were omitted from the study to remove rare events during pregnancy. Mean values were used to perform the imputation of missing numerical values. Discrete features, such as clinical diagnoses, prescribed medication, were coded as dummy features (Rodríguez et al., 2018). Numeric features were normalized in the scale of -1 to 1 . Statistical comparison across the PPD and non-PPD group was performed using Stata 14. Independent sample T-test assuming unequal variances and Chi-Square test was used for continuous and categorical variables as appropriate.

2.3 Machine learning

2.3.1 Framework—The schematic diagram of our PPD prediction framework is shown in Fig. 1 (Schematic diagram of our PPD prediction framework) that describes the various steps involved in data preprocessing and risk model development. The machine learning model training was optimized using sequential forward selection (SFS) (T. and G., 2015) - a greedy search algorithm that searches for the combination of features that returns the maximum algorithm discriminatory power (Bradley, 1997). Starting with an empty feature set, SFS iteratively examines each feature combination such that the algorithm's performance can be maximized until the stopping criteria for the search is reached (Fig. A1) (T. and G., 2015). Five machine learning algorithms were trained, including random forest, decision tree, extreme gradient boosting (XGboost), regularized logistic regression, and multilayer perceptron (MLP). These algorithms were developed by iteratively splitting the data available to detect collective patterns across features in the subset of the data that maximally discriminate outcome classes, followed by testing the performance on the held-out data. This training process allowed us to develop prediction algorithms that are generalizable to unseen data.

Algorithm parameters were determined using a grid search for each algorithm that comprehensively searched for the best hyperparameters and parameters that resulted in the highest model performance as measured by area under the receiver operating characteristic curve (AUC). The stopping criteria for SFS were defined as 1) no increase in the AUC by 0.001 after 10 consecutive iterations, or 2) the predetermined maximum number of feature set has been reached. SFS was performed separately for women with, and without, mental health history to ensure that the model can predict for both types of patients when in actual use. We combined features selected from both SFS into a single feature set such that a single algorithm can be used for patients with and without a history of mental illness. Using the combined features, each of the machine learning algorithms was trained using 5-fold cross-validation.

2.3.2 Expert adjudicated feature selection—Clinicians in our study team (AH and RJ) reviewed the selected features in the best performing algorithm to validate feature inclusion and ensure algorithm interpretability. Starting with the entire list of features selected by SFS, we iteratively eliminated features that were determined to be irrelevant, re-constructed the algorithm using the adjudicated features, and measured the algorithm performance. This iterative process was performed while keeping the minimum AUC at 0.8. Features that were changed during this process are listed in the Appendix (Table A2).

2.3.3 Evaluation—The evaluation was performed using the held-out data set at WCM and the entire dataset from NYC-CDRN using AUC, sensitivity, specificity, and the Brier score (Hanley and McNeil, 1982). AUC is an aggregate measure of the algorithm's ability to discriminate outcome classes across all possible classification thresholds. The Brier score measures the accuracy of prediction (Rufibach, 2010). As such, higher AUC and lower Brier score indicate better prediction performance. To evaluate the algorithm performance in a simulated gestational period where data are being accumulated during pregnancy, we computed evaluation metrics using data available up to 5 different periods. Starting with

each patient's first available pregnancy encounter, we created a test dataset ending at 12-week, 18-week, 24-week, and 30-week during pregnancy, and also at childbirth assuming that data at 12-week pregnancy and childbirth contain the least and the most complete information, respectively. Lastly, error analyses were conducted by manual chart review using patients' medical records for up to 2 years after childbirth for 150 false positives and negatives. Machine learning algorithms were trained and evaluated using Scikit-learn and Seaborn in Python (3.6.5).

3. Results

A total of 15,197 deliveries from January 2015 to June 2018 were included in our analysis, excluding 124 women below age 18 or above age 45 at the time of delivery, and 2,312 women without records of clinical encounters within 1 year following childbirth (Fig. 2). Study data were randomly split into training (N=12157) and testing (N=3040) using cross-validation. The validation set contained 53,972 deliveries from August 2004 to October 2017, after excluding 1,903 deliveries by women below age 18 or above age 45 and 15,141 deliveries without encounters recorded within 1 year after childbirth (Fig. 2). The prevalence of depression was 6.7% (N=1,010) and 6.5% (N=3,513) in the WCM and NYC-CDRN datasets, respectively.

Table 1 provides the descriptive statistics of the two datasets. We found significant differences in age, the number of emergency department (ED) visits, and racial distribution between PPD and non-PPD groups in the training and validation data, respectively. The average age at the time of delivery was 33.68 (SD=4.54) in the non-PPD group and 34.56 (SD=4.39) in the PPD group of patients in the WCM dataset (p-value<0.001), and 28.87 (SD=6.20) and 30.70 (SD=6.13) in the CDRN dataset, respectively (p-value<0.001). The number of emergency room visits in the PPD group was higher than the non-PPD group in both the WCM (1.68 ± 1.55 vs. 1.32 ± 1.24 , p-value<0.001) and NYC-CDRN (6.30 ± 9.97 vs. 5.37 ± 6.87 , p-value<0.001) datasets. The training and validation datasets had different distribution of PPD across racial groups. In the WCM data, the incidence rate of PPD was the highest among White women (8.8%) and the lowest among Asian women (3.0%). In the CDRN data, the rate of PPD was the highest among White women (12.43%), Black patients had the lowest rates (4.76%).

Using SFS, 32 features were selected to be incorporated in the algorithm related to patient demographic statuses, health service utilization, mental health history, newly diagnosed mental health conditions during pregnancy, other obstetric and/or medical diagnoses during pregnancy, and vital signs. As shown in Table 2, the majority (28 out of 32) of the features included in the algorithm have statistically significant association with the outcome. Features that are indicative of past and current mental health conditions and being single mothers were associated with higher odds of a PPD diagnosis. Additionally, complications during pregnancy such as palpitations, diarrhea, vomiting, and abdominal pain also were associated with higher odds of a PPD diagnosis. Health service utilization including medication prescriptions such as Beta blocking agents, delivery by cesarean, and emergency department (ED) visits were also associated with higher odds of a PPD diagnosis. Having an

Asian race was associated with lower odds of a PPD diagnosis. Fig. A2 in the Appendix shows the Pearson correlation among the features.

Evaluation results of the algorithm performance are shown in Table 3. Logistic regression with L2 regularization was found to be the best performing algorithm using data available up to childbirth. The AUC was 0.937 (95% CI: 0.912–0.962) and 0.886 (95% CI: 0.879–0.893) in the WCM and NYC-CDRN datasets, respectively. The AUC was lower in the validation dataset potentially due to the lack of certain features such as marital status which was available only in the WCM dataset. While evaluating algorithms at different periods during the pregnancy, we observed a steady performance with respect to AUC of 0.921, 0.919, 0.922, 0.921, and 0.937 using data extracted up to 12 weeks, 18 weeks, 24 weeks, 30 weeks of gestation, and at childbirth, respectively. The steady performance may be explained by the early availability and invariability of the predictive features (see Table 2). In the NYC-CDRN dataset, we observe an increase in algorithm performance as more data accumulate over time, with an AUC of 0.810, 0.817, 0.821, 0.824, and 0.886 at 12 weeks, 18 weeks, 24 weeks, 30 weeks of gestation, and lastly at childbirth, respectively. Additionally, we report positive and negative predictive values in Table 3. While negative predictive values are close to 1 for nearly all models across time periods, we find that positive predictive values are low especially in the validation site. This could be explained by the relative low prevalence of PPD and the high frequency of the patients who were not diagnosed to have PPD (based on our criteria), but were predicted so.

False-positive and false-negative results from the algorithm were evaluated by manual chart reviews of a randomly selected 150 cases that were incorrectly classified by the logistic regression classifier. The cases had 140 and 10 false positives and false negatives, respectively. PPD diagnosis after the study period and lack of proper coding were identified as two potential reasons for the false positives and negatives. For example, the manual chart review identified that 45% of the patients incorrectly predicted to develop PPD by the prediction algorithm were in fact women who were noted to be suffering from PPD in the clinical notes. Furthermore, 34% of the PPD mentions in the notes were made one year after childbirth, beyond our study period. Thus, the incorrect predictions were due to the lack of good coding practices for PPD, a phenomenon that is frequently observed in other observational mental health studies using EHRs (Stewart et al., 2019). The availability of predictors related to mental health history also presented challenges. For example, the error analysis identified the history of anxiety and depression on 36.4% of false negative cases through manual chart review. For these patients the mental health history was not coded in the structured EHR data. Extraction of features using natural language processing techniques may facilitate higher performance by the algorithm in future studies.

4. Discussion

Results from this study suggest a promising direction to leverage routinely collected EHR data to identify pregnant women at risk for PPD. Selected EHR-driven predictors characterize women's health history, pregnancy health, demographics, and healthcare utilization. Several known PPD risk factors from the literature were represented by variables extracted in the sequential feature selection process, including history of anxiety, mood

disorder, and other mental disorders, antidepressant use, incidental mental health illnesses during pregnancy, cesarean section, and single motherhood.(Forman et al., 2000; Stewart and Vigod, 2016) Our model further identifies additional comorbid predictors, including palpitations, diarrhea, vomiting during pregnancy, hypertensive disorders and hypothyroidism. Among these comorbidities, thyroid dysfunction and hypertensive disorders have been associated with PPD onset in previous literature.(Le Donne et al., 2017; Strapasson et al., 2018) Palpitation, a common cardiac symptom, may also be a symptom of depression that was discovered by the model.(Alijaniha et al., 2016; Barsky et al., 1994) In addition, medication prescriptions of beta blocking agents and antihistamines were identified as predictors. Literature has reported the use of both beta blockers and antihistamines in association with depression although not conclusively (Yudofsky, 1992)(Gerstman et al., 1996; Ozdemir et al., 2014). Related to mode of delivery, our model selected cesarean section as a risk factor for PPD, as also studied in the previous literature.(Carter et al., 2006; Xu et al., 2017) Lastly, the number of ED visits during pregnancy and postpartum may be an indicator of a lack of proper access to primary and obstetric care.(Sheen et al., 2019)

As seen in our experiments, the risk computed by the PPD prediction algorithm updates in response to the new health information that accumulates overtime with repeated visits during pregnancy, thereby potentially allowing care providers to take timely actions according to the risk evolution. (Committee, 2018; Earls et al., 2019) With these automatically extractable features, an EHR-based prediction tool may assist with existing EHR interventions for screening to minimize variations across clinical practices in screening and information collection.(Long et al., 2019)Previous studies have reported that while the rates of screening and referrals for mental health care can be high when obstetricians recognize a risk for PPD, but they are low if symptoms are unnoticed by the care provider.(Goodman and Tyer-Viola, 2010) Our risk prediction model, by identifying women with elevated risk, may assist with tacitly raising clinician awareness of PPD and potentially increasing screening and referral rates.

5. Limitations

Several limitations exist in our research. First, our study cohort as derived from the EHR in an urban academic medical center is not representative of the general US population suffering from PPD and differs from cohorts reported in previous studies with respect to PPD prevalence (Hahn-Holbrook et al., 2017). This prevalence is likely the treatment prevalence rather than the illness prevalence, as the data may not capture patients outside of the studied health system and geographical location. The prevalence may also reflect the clinician coding practices on recording a diagnosis of PPD at the study sites. Persistent stigma and social consequences of having depression coded in the EHR may prevent providers from ‘officially’ coding the diagnosis even if it is made clinically. Further, also due to stigma, patients may withhold symptom information from providers preventing accurate diagnosis. In addition to using diagnostic codes, we also defined PPD using antidepressant use while excluding those for pain indications. However, it is possible that some antidepressants were used for anxiety rather than PPD. Anxiety disorders are so frequently comorbid with depression in the peripartum period such that a diagnosis of one may even be a proxy for unidentified depression. Thus, we decided it was important not to exclude

anxiety disorder indications even at the expense of specificity, although we recognize this as a limitation of our study. Our ongoing and future work will attempt to parse these indications further by applying natural language processing to the unstructured clinical notes.

Relatedly, in this study, we did not specifically include only patients with incident depression. This decision was meant to acknowledge the powerful effect that mental health histories have on risk for developing PPD as well as to provide a clinically meaningful risk stratification for real-world obstetric providers who have large cohorts of patients with mental health histories and those who are actively seeking treatment in their practices. Due to the lack of comprehensive screening at our health systems and clinics in the study sites, we did not capture EPDS and PHQ-9 scores to define PPD. We also did not compare effectiveness of primary prevention via the prediction algorithm to current widely recommended secondary prevention efforts via EPDS or PHQ-9 screening. However, we did compare with algorithms reported in prior literature as a potential primary intervention approach, and demonstrated improved model performance. Compared to prior work by Camdeviren et al (Camdeviren et al., 2007), Tortajada et al (Tortajada et al., 2009), and Natarajan et al (S. et al., 2017), our algorithm was built by exhaustively selecting most predictive features from a larger number of candidate features from the EHR data, with an eventual goal of integrating such risk prediction models within the EHR systems and clinical workflows. Furthermore, compared to our initial pilot work (Wang et al., 2019) which did not include prior mental health diagnosis and treatment history as predictors, the prediction algorithm from this study demonstrated a significant increase in AUC, sensitivity, and specificity.

A number of future works are under preparation to address these limitations. We found White and Asian races to be predictive features in this study. However, a substantial proportion of race was unknown in both the training and validation datasets, potentially due to lack of proper documentation in the EHR (Lee et al., 2016). This is an important area for further consideration in future studies.(Sholle et al., 2019) These include a comparison of the data-driven primary intervention against usual care as a clinical trial, and additional validation work at study sites in the greater US and abroad using datasets with different PPD prevalence to evaluate the algorithm generalizability. While findings from this study present a promise for PPD risk identification using available EHR data, we realize that EHR data capture only a limited portion of patients' lives which contribute to PPD. Therefore, we will also evaluate whether the addition of patient-reported outcomes or information derived from mobile health devices, such as wearables, can contribute to higher algorithm performance. Lastly, improvement in the machine learning framework will include techniques to adjust for differing outcome distributions such that the method can be more generally applied to other populations.

6. Conclusions

In summary, this study demonstrates that a data-driven primary intervention approach using machine learning and EHR data may be leveraged to reduce the healthcare provider burden of identifying PPD risk. Methods created in this study may pave a path towards data-driven,

accurate, and scalable clinical decision support for PPD risk identification with potential benefits through early prevention, diagnosis, and intervention.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Abbreviations

PPD	postpartum depression
EHRs	electronic health records
SDoH	social determinants of health
EPDS	Edinburgh Postnatal Depression Scale
PHQ-9	Patient Health Questionnaire-9
ACOG	American College of Obstetricians and Gynecologists
AAP	American Academy of Pediatrics
XGBoost	extreme gradient boosting
MLP	multilayer perceptron
LR	logistic regression
RF	random forest
SNOMED	Systematized Nomenclature of Medicine
ATC	Anatomical Therapeutic Chemical
AUC	area under the receiver operating characteristic curve
SFS	sequential forward selection
WCM	Weill Cornell Medicine
NYC-CDRN	New York City Clinical Data Research Network

References

- Alijaniha F, Noorbala A, Afsharypuor S, Naseri M, Fallahi F, Mosaddegh M, Zadeh SF, Sadrai S, 2016 Relationship Between Palpitation and Mental Health. *Iran Red Crescent Me* 18.
- Barsky AJ, Cleary PD, Coeytaux RR, Ruskin JN, 1994 Psychiatric disorders in medical outpatients complaining of palpitations. *J Gen Intern Med* 9, 306–313. [PubMed: 8077994]
- Beck CT, Gable RK, 2000 Postpartum Depression Screening Scale: development and psychometric testing. *Nurs Res* 49, 272–282. [PubMed: 11009122]
- Biaggi A, Conroy S, Pawlby S, Pariante CM, 2016 Identifying the women at risk of antenatal anxiety and depression: A systematic review. *J Affect Disord* 191, 62–77. [PubMed: 26650969]
- Bishop CM, 2006 Pattern recognition and machine learning. Springer-Verlag New York.

- Bradley AP, 1997 The use of the area under the ROC curve in the evaluation of machine learning algorithms. *Pattern Recognition* 30, 1145–1159.
- Camdeviren HA, Yazici AC, Akkus Z, Bugdayci R, Sungur MA, 2007 Comparison of logistic regression model and classification tree: An application to postpartum depression data. *Expert Systems with Applications* 32, 987–994.
- Canada S, 2015 Mental disorders, conditions, and ICD-9 and ICD-10-CA codes. Statistics Canada.
- Carter FA, Frampton CMA, Mulder RT, 2006 Cesarean section and postpartum depression: A review of the evidence examining the link. *Psychosom Med* 68, 321–330. [PubMed: 16554400]
- Cohen LS, Altshuler LL, Harlow BL, Nonacs R, Newport DJ, Viguera AC, Suri R, Burt VK, Hendrick V, Reminick AM, Loughhead A, Vitonis AF, Stowe ZN, 2006 Relapse of major depression during pregnancy in women who maintain or discontinue antidepressant treatment. *Jama* 295, 499–507. [PubMed: 16449615]
- Committee A, 2018 ACOG Committee Opinion No. 757: Screening for Perinatal Depression. *Obstet Gynecol* 132, e208–e212. [PubMed: 30629567]
- Cox JL, Holden JM, Sagovsky R, 1987 Detection of postnatal depression: Development of the 10-item Edinburgh Postnatal Depression Scale. *Br J Psychiatry* 150, 782–786. [PubMed: 3651732]
- Dietz PM, Williams SB, Callaghan WM, Bachman DJ, Whitlock EP, Hornbrook MC, 2007 Clinically identified maternal depression before, during, and after pregnancies ending in live births. *Am J Psychiatry* 164, 1515–1520. [PubMed: 17898342]
- Earls MF, Yogman MW, Mattson G, Rafferty J, 2019 Incorporating Recognition and Management of Perinatal Depression Into Pediatric Practice. *Pediatrics* 143.
- Field T, 2010 Postpartum depression effects on early interactions, parenting, and safety practices: a review. *Infant Behav Dev* 33, 1–6. [PubMed: 19962196]
- Forman DN, Videbech P, Hedegaard M, Salvig JD, Secher NJ, 2000 Postpartum depression: identification of women at risk. *Brit J Obstet Gynaec* 107, 1210–1217.
- Gerstman BB, Jolson HM, Bauer M, Cho P, Livingston JM, Platt R, 1996 The incidence of depression in new users of beta-blockers and selected antihypertensives. *J Clin Epidemiol* 49, 809–815. [PubMed: 8691233]
- Gjerdingen DK, Yawn BP, 2007 Postpartum depression screening: importance, methods, barriers, and recommendations for practice. *J Am Board Fam Med* 20, 280–288. [PubMed: 17478661]
- Goldstein BA, Navar AM, Pencina MJ, Ioannidis JP, 2017 Opportunities and challenges in developing risk prediction models with electronic health records data: a systematic review. *J Am Med Inform Assoc* 24, 198–208. [PubMed: 27189013]
- Goodman JH, Tyer-Viola L, 2010 Detection, Treatment, and Referral of Perinatal Depression and Anxiety by Obstetrical Providers. *J Womens Health* 19, 477–490.
- Hahn-Holbrook J, Cornwell-Hinrichs T, Anaya I, 2017 Economic and Health Predictors of National Postpartum Depression Prevalence: A Systematic Review, Meta-analysis, and Meta-Regression of 291 Studies from 56 Countries. *Front Psychiatry* 8, 248. [PubMed: 29449816]
- Hanley JA, McNeil BJ, 1982 The meaning and use of the area under a receiver operating characteristic (ROC) curve. *Radiology* 143, 29–36. [PubMed: 7063747]
- Jacques N, de Mola CL, Joseph G, Mesenburg MA, da Silveira MF, 2019 Prenatal and postnatal maternal depression and infant hospitalization and mortality in the first year of life: A systematic review and meta-analysis. *J Affect Disord* 243, 201–208. [PubMed: 30245252]
- Jiménez-Serrano S, Tortajada S, García-Gómez JM, 2015 A Mobile Health Application to Predict Postpartum Depression Based on Machine Learning. *Telemed J E Health* 21, 567–574. [PubMed: 25734829]
- Kaushal R, Hripcsak G, Ascheim DD, Bloom T, Campion TR Jr., Caplan AL, Currie BP, Check T, Deland EL, Gourevitch MN, Hart R, Horowitz CR, Kastenbaum I, Levin AA, Low AF, Meissner P, Mirhaji P, Pincus HA, Scaglione C, Shelley D, Tobin JN, 2014 Changing the research landscape: the New York City Clinical Data Research Network. *J Am Med Inform Assoc* 21, 587–590. [PubMed: 24821739]
- Le Donne M, Mento C, Settineri S, Antonelli A, Benvenga S, 2017 Postpartum Mood Disorders and Thyroid Autoimmunity. *Front Endocrinol (Lausanne)* 8, 91. [PubMed: 28522989]

- Lee WC, Veeranki SP, Serag H, Eschbach K, Smith KD, 2016 Improving the Collection of Race, Ethnicity, and Language Data to Reduce Healthcare Disparities: A Case Study from an Academic Medical Center. *Perspect Health Inf Manag* 13, 1g.
- Liang H, Tsui BY, Ni H, Valentim CCS, Baxter SL, Liu G, Cai W, Kermany DS, Sun X, Chen J, He L, Zhu J, Tian P, Shao H, Zheng L, Hou R, Hewett S, Li G, Liang P, Zang X, Zhang Z, Pan L, Cai H, Ling R, Li S, Cui Y, Tang S, Ye H, Huang X, He W, Liang W, Zhang Q, Jiang J, Yu W, Gao J, Ou W, Deng Y, Hou Q, Wang B, Yao C, Liang Y, Zhang S, Duan Y, Zhang R, Gibson S, Zhang CL, Li O, Zhang ED, Karin G, Nguyen N, Wu X, Wen C, Xu J, Xu W, Wang B, Wang W, Li J, Pizzato B, Bao C, Xiang D, He W, He S, Zhou Y, Haw W, Goldbaum M, Tremoulet A, Hsu CN, Carter H, Zhu L, Zhang K, Xia H, 2019 Evaluation and accurate diagnoses of pediatric diseases using artificial intelligence. *Nat Med* 25, 433–438. [PubMed: 30742121]
- Long MM, Cramer RJ, Jenkins J, Bennington L, Paulson JF, 2019 A systematic review of interventions for healthcare professionals to improve screening and referral for perinatal mood and anxiety disorders. *Arch Women Ment Hlth* 22, 25–36.
- Loudon H, Nentin F, Silverman ME, 2016 Using clinical decision support as a means of implementing a universal postpartum depression screening program. *Arch Womens Ment Health* 19, 501–505. [PubMed: 26669601]
- Löwe B, Unützer J, Callahan CM, Perkins AJ, Kroenke K, 2004 Monitoring depression treatment outcomes with the patient health questionnaire-9. *Med Care* 42, 1194–1201. [PubMed: 15550799]
- Meltzer-Brody S, Howard LM, Bergink V, Vigod S, Jones I, Munk-Olsen T, Honikman S, Milgrom J, 2018 Postpartum psychiatric disorders. *Nat Rev Dis Primers* 4, 18022. [PubMed: 29695824]
- O'Connor E, Senger CA, Henninger ML, Coppola E, Gaynes BN, 2019 Interventions to Prevent Perinatal Depression: Evidence Report and Systematic Review for the US Preventive Services Task Force. *Jama* 321, 588–601. [PubMed: 30747970]
- Overhage JM, Ryan PB, Reich CG, Hartzema AG, Stang PE, 2012 Validation of a common data model for active safety surveillance research. *J Am Med Inform Assoc* 19, 54–60. [PubMed: 22037893]
- Ozdemir PG, Karadag AS, Selvi Y, Boysan M, Bilgili SG, Aydin A, Onder S, 2014 Assessment of the effects of antihistamine drugs on mood, sleep quality, sleepiness, and dream anxiety. *Int J Psychiat Clin* 18, 161–168.
- Petersen I, Peltola T, Kaski S, Walters KR, Hardoon S, 2018 Depression, depressive symptoms and treatments in women who have recently given birth: UK cohort study. *BMJ Open* 8, e022152.
- Rajkomar A, Oren E, Chen K, Dai AM, Hajaj N, Hardt M, Liu PJ, Liu X, Marcus J, Sun M, Sundberg P, Yee H, Zhang K, Zhang Y, Flores G, Duggan GE, Irvine J, Le Q, Litsch K, Mossin A, Tansuwan J, Wang D, Wexler J, Wilson J, Ludwig D, Volchenboum SL, Chou K, Pearson M, Madabushi S, Shah NH, Butte AJ, Howell MD, Cui C, Corrado GS, Dean J, 2018 Scalable and accurate deep learning with electronic health records. *NPJ Digit Med* 1, 18. [PubMed: 31304302]
- Rodríguez P, Bautista MA, González J, Escalera S, 2018 Beyond one-hot encoding: Lower dimensional target embedding. *Image and Vision Computing* 75, 21–31.
- Rufibach K, 2010 Use of Brier score to assess binary predictions. *J Clin Epidemiol* 63, 938–939; author reply 939. [PubMed: 20189763]
- S., N., A., P., N., R., A., B., K., S., K., C., 2017 Boosting for Postpartum Depression Prediction, 2017 IEEE/ACM International Conference on Connected Health: Applications, Systems and Engineering Technologies (CHASE), pp. 232–240.
- Schofield P, Das-Munshi J, Mathur R, Congdon P, Hull S, 2016 Does depression diagnosis and antidepressant prescribing vary by location? Analysis of ethnic density associations using a large primary-care dataset. *Psychol Med* 46, 1321–1329. [PubMed: 26879871]
- Serati M, Redaelli M, Buoli M, Altamura AC, 2016 Perinatal Major Depression Biomarkers: A systematic review. *J Affect Disord* 193, 391–404. [PubMed: 26802316]
- Sheen JJ, Smith HA, Tu B, Liu Y, Sutton D, Bernstein PS, 2019 Risk Factors for Postpartum Emergency Department Visits in an Urban Population. *Matern Child Hlth J* 23, 557–566.
- Sholle ET, Pinheiro LC, Adekanattu P, Davila MA, Johnson SB, Pathak J, Sinha S, Li C, Lubansky SA, Safford MM, Champion TR, 2019 Underserved populations with missing race ethnicity data differ significantly from those with structured race/ethnicity documentation. *Journal of the American Medical Informatics Association : JAMIA* 26, 722–729. [PubMed: 31329882]

- Shortliffe EH, Sepúlveda MJ, 2018 Clinical Decision Support in the Era of Artificial Intelligence. *Jama* 320, 2199–2200. [PubMed: 30398550]
- Stein A, Pearson RM, Goodman SH, Rapa E, Rahman A, McCallum M, Howard LM, Pariante CM, 2014 Effects of perinatal mental disorders on the fetus and child. *Lancet* 384, 1800–1819. [PubMed: 25455250]
- Stewart CC, Lu CY, Yoon TK, Coleman KJ, Crawford PM, Lakoma MD, Simon GE, 2019 Impact of ICD-10-CM Transition on Mental Health Diagnoses Recording. *EGEMS (Wash DC)* 7, 14. [PubMed: 31065557]
- Stewart DE, Vigod S, 2016 Postpartum Depression. *N Engl J Med* 375, 2177–2186. [PubMed: 27959754]
- Strapasson MR, Ferreira CF, Ramos JGL, 2018 Associations between postpartum depression and hypertensive disorders of pregnancy. *Int J Gynecol Obstet* 143, 367–373.
- T. L, G. T, 2015 SFS feature selection technique for multistage emotion recognition, 2015 IEEE 3rd Workshop on Advances in Information, Electronic and Electrical Engineering (AIEEE), pp. 1–4.
- Tomašev N, Glorot X, Rae JW, Zielinski M, Askham H, Saraiva A, Mottram A, Meyer C, Ravuri S, Protsyuk I, Connell A, Hughes CO, Karthikesalingam A, Cornebise J, Montgomery H, Rees G, Laing C, Baker CR, Peterson K, Reeves R, Hassabis D, King D, Suleyman M, Back T, Nielson C, Ledsam JR, Mohamed S, 2019 A clinically applicable approach to continuous prediction of future acute kidney injury. *Nature* 572, 116–119. [PubMed: 31367026]
- Tortajada S, García-Gomez JM, Vicente J, Sanjuán J, de Frutos R, Martín-Santos R, García-Esteve L, Gornemann I, Gutiérrez-Zotes A, Canellas F, Carracedo A, Gratacos M, Guillamat R, Baca-Garcia E, Robles M, 2009 Prediction of postpartum depression using multilayer perceptrons and pruning. *Methods Inf Med* 48, 291–298. [PubMed: 19387507]
- Wang S, Pathak J, Zhang Y, 2019 Using Electronic Health Records and Machine Learning to Predict Postpartum Depression. *Stud Health Technol Inform* 264, 888–892. [PubMed: 31438052]
- Weobong B, ten Asbroek AH, Soremekun S, Gram L, Amenga-Etego S, Danso S, Owusu-Agyei S, Prince M, Kirkwood BR, 2015 Association between probable postnatal depression and increased infant mortality and morbidity: findings from the DON population-based cohort study in rural Ghana. *BMJ Open* 5, e006509.
- Wisner KL, Sit DK, McShea MC, Rizzo DM, Zoretich RA, Hughes CL, Eng HF, Luther JF, Wisniewski SR, Costantino ML, Confer AL, Moses-Kolko EL, Famy CS, Hanusa BH, 2013 Onset timing, thoughts of self-harm, and diagnoses in postpartum women with screen-positive depression findings. *JAMA Psychiatry* 70, 490–498. [PubMed: 23487258]
- Xu H, Ding Y, Ma Y, Xin XL, Zhang DF, 2017 Cesarean section and risk of postpartum depression: A meta-analysis. *J Psychosom Res* 97, 118–126. [PubMed: 28606491]
- Yudofsky SC, 1992 Beta-blockers and depression. The clinician's dilemma. *JAMA : the journal of the American Medical Association* 267, 1826–1827.
- Zhang W, Liu H, Silenzio VMB, Qiu P, Gong W, 2020 Machine Learning Models for the Prediction of Postpartum Depression: Application and Comparison Based on a Cohort Study. *JMIR Med Inform* 8, e15516. [PubMed: 32352387]

Highlights

- Applying machine learning to electronic health records (EHR) data can preemptively identify women at higher risk of postpartum depression.
- Two datasets of multi-site EHR data were used as development and validation sets, respectively.
- Mental health history, number of emergency department visits, blood pressure, complications during pregnancy are among the predictors in the machine learning algorithm.
- The algorithm performances as measured by area under the receiver operating characteristic curve are 0.937 and 0.886 in the development and validation datasets, respectively.

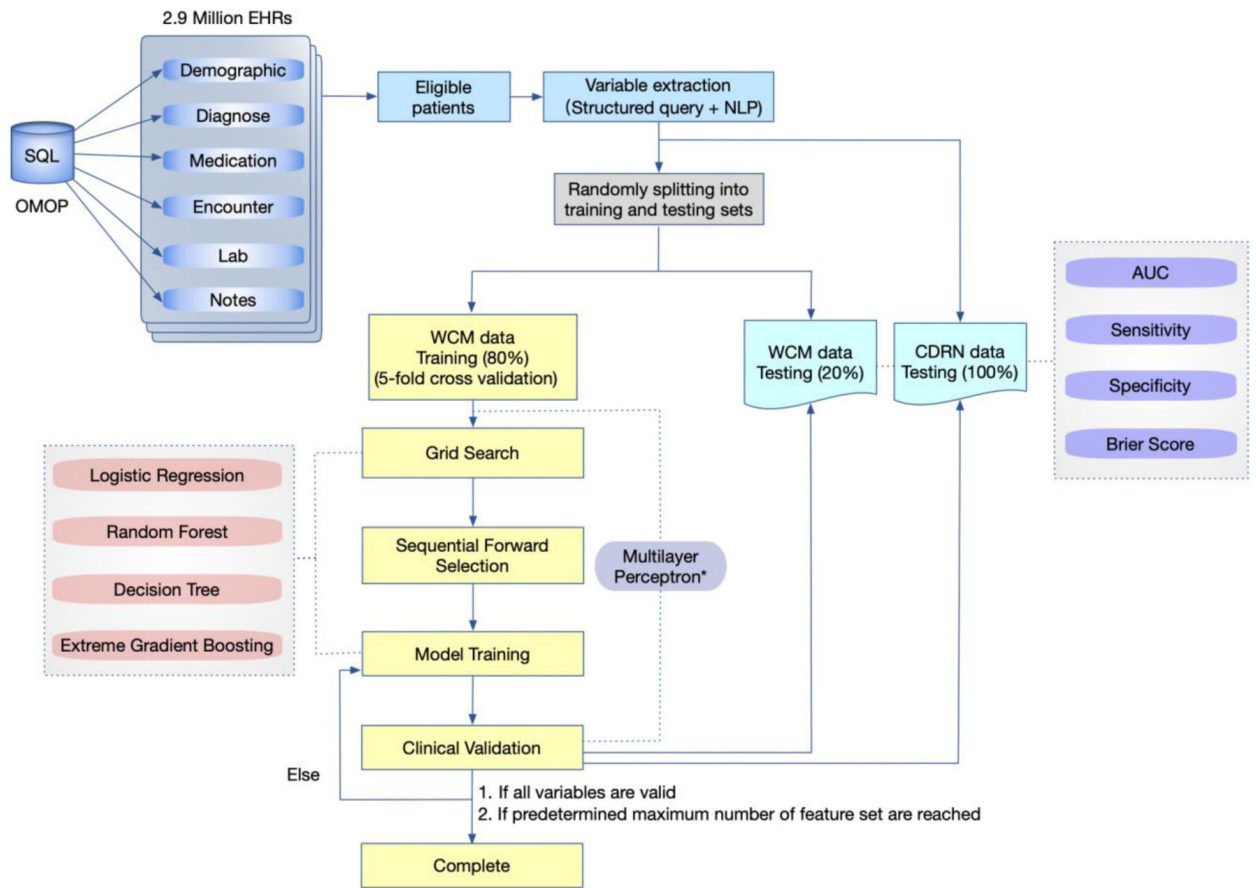


Fig. 1. Schematic diagram of our PPD prediction framework. This diagram depicts the process of the study. The modified version of multilayer perceptron (gray) is used to compare with Tortajada et al (2009).

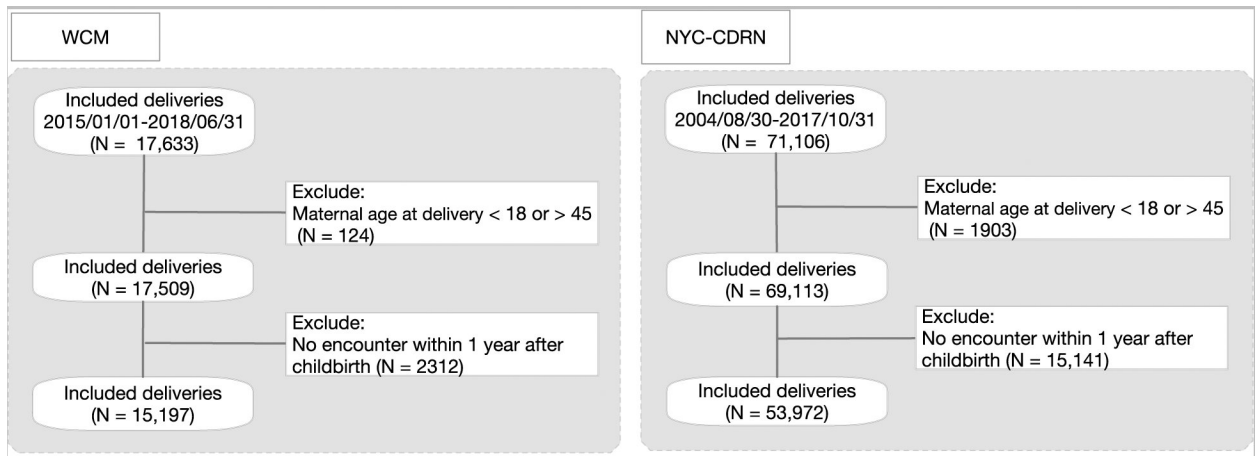


Fig. 2.
The inclusion and exclusion criteria of this study.

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

Table 1.

Demographic information of the two datasets

Variable	WCM data			CDRN data		
	non-PPD	PPD	<i>P</i>	non-PPD	PPD	<i>P</i>
N (%)	14187(93.35)	1010(6.65)		50459(93.49)	3513(6.51)	
	Mean ± SD			Mean ± SD		
Age, year	33.68 ± 4.54	34.56 ± 4.39	<0.001	28.87 ± 6.20	30.70 ± 6.13	<0.001
Pre-pregnancy BMI, kg/m ²	23.92 ± 4.32	24.07 ± 4.51	0.294	28.72 ± 6.29	28.27 ± 6.91	<0.001
# ED visits	1.32 ± 1.24	1.68 ± 1.55	<0.001	5.37 ± 6.87	6.30 ± 9.97	<0.001
	N (%)			N (%)		
Race group						
White	6794(91.23)	653(8.77)	<0.001	8275(87.57)	1175(12.43)	<0.001
Asian	2784(96.97)	87(3.03)		1535(94.58)	88(5.42)	
Black	839(93.85)	55(6.15)		13815(95.24)	690(4.76)	
Other combinations	1612(94.16)	100(5.84)		19163(94.09)	1204(5.91)	
Unknown	2158(94.94)	115(5.06)		7671(95.56)	356(4.44)	

Table 2.

The association of selected features with PPD using univariate logistic regression in WCM training data.

Variables	non-PPD	PPD	OR (95%CI)	P
N(%)	11324(93.15)	833(6.85)		
Anxiety history	135(1.19)	473(56.78)	108.90(87.55,135.44)	<0.001
Other disorder history	156(1.38)	169(20.29)	18.22(14.46,22.96)	<0.001
Antidepressants	22(0.19)	129(15.49)	94.14(59.52,148.89)	<0.001
Mood disorder history	120(1.06)	266(31.93)	43.80(34.75,55.21)	<0.001
Depression in pregnancy	60(0.53)	108(12.97)	27.97(20.22,38.68)	<0.001
Anxiety in pregnancy	82(0.72)	141(16.93)	27.93(21.05,37.07)	<0.001
Mental disorder in pregnancy	62(0.55)	91(10.92)	22.27(15.99,31.01)	<0.001
Palpitations	147(1.30)	28(3.36)	2.65(1.75,3.99)	<0.001
Diarrhea	159(1.40)	29(3.48)	2.54(1.70,3.79)	<0.001
Vomiting in pregnancy	298(2.63)	53(6.36)	2.52(1.86,3.40)	<0.001
Hypertensive disorder	104(0.92)	17(2.04)	2.25(1.34,3.77)	0.002
Acute pharyngitis	107(0.94)	17(2.04)	2.19(1.30,3.66)	0.003
Hemorrhage in early pregnancy antepartum	92(0.81)	14(1.68)	2.09(1.18,3.68)	0.011
White	5390(47.60)	540(64.83)	2.03(1.75,2.35)	<0.001
Threatened miscarriage	478(4.22)	67(8.04)	1.99(1.52,2.59)	<0.001
Abdominal pain	699(6.17)	96(11.52)	1.98(1.58,2.48)	<0.001
Migraine	69(0.61)	10(1.20)	1.98(1.02,3.86)	0.044
Beta blocking agents	173(1.53)	24(2.88)	1.91(1.24,2.95)	0.003
Antihistamines for systemic use	661(5.84)	84(10.08)	1.81(1.43,2.30)	<0.001
Hypothyroidism	1062(9.38)	121(14.53)	1.64(1.34,2.01)	<0.001
Placental infarct	264(2.33)	31(3.72)	1.62(1.11,2.37)	0.013
Single (vs. Married)	1412(12.47)	154(18.49)	1.59(1.33,1.91)	<0.001
Deliveries by cesarean	2449(21.63)	240(28.81)	1.47(1.25,1.72)	<0.001
Direct acting antivirals	482(4.26)	50(6.00)	1.44(1.06,1.94)	0.018
Primigravida	6699(59.16)	556(66.75)	1.39(1.19,1.61)	<0.001
Pre-eclampsia	79(0.70)	8(0.96)	1.38(0.67,2.87)	0.386
Other antibacterials	489(4.32)	47(5.64)	1.33(0.97,1.80)	0.073
#ED visit	1.31 ± 1.24	1.67 ± 1.54	1.24(1.17,1.31)	<0.001
Abnormality of organs and/or soft tissues of pelvis affecting pregnancy	659(5.82)	53(6.36)	1.10(0.82,1.47)	0.52
Diastolic blood pressure in third trimester	69.30 ± 5.64	69.83 ± 5.88	1.09(1.02,1.16)	0.009
False labor at or after 37 completed weeks of gestation	270(2.38)	13(1.56)	0.65(0.37,1.14)	0.131
Asian	2227(19.67)	75(9.00)	0.40(0.32,0.52)	<0.001

Table 3.

The model performance.

Time	Classifier	AUROC	Sensitivity	Specificity	Brier Score	PPV	NPV
WCM (development site)							
12wk	Logistics regression	0.921(0.893,0.949)	0.79	0.97	0.074	0.61	0.99
	Random Forest	0.897(0.866,0.928)	0.80	0.97	0.054	0.60	0.99
	Decision Tree	0.903(0.873,0.933)	0.83	0.96	0.045	0.59	0.99
	XGboost	0.908(0.878,0.938)	0.82	0.97	0.068	0.60	0.99
	MLP	0.921(0.893,0.949)	0.63	0.98	0.028	0.71	0.98
18wk	Logistics regression	0.919(0.891,0.947)	0.79	0.97	0.074	0.61	0.99
	Random Forest	0.897(0.866,0.928)	0.80	0.97	0.056	0.60	0.99
	Decision Tree	0.890(0.858,0.922)	0.82	0.96	0.048	0.59	0.99
	XGboost	0.902(0.872,0.932)	0.82	0.97	0.097	0.60	0.99
	MLP	0.919(0.891,0.947)	0.63	0.98	0.028	0.71	0.98
24wk	Logistics regression	0.922(0.895,0.949)	0.79	0.97	0.074	0.61	0.99
	Random Forest	0.903(0.873,0.933)	0.80	0.97	0.057	0.60	0.99
	Decision Tree	0.895(0.864,0.926)	0.83	0.96	0.048	0.59	0.99
	XGboost	0.919(0.891,0.947)	0.83	0.96	0.082	0.57	0.99
	MLP	0.920(0.892,0.948)	0.63	0.98	0.028	0.72	0.98
30wk	Logistics regression	0.921(0.893,0.949)	0.79	0.97	0.074	0.61	0.99
	Random Forest	0.914(0.885,0.943)	0.83	0.97	0.056	0.65	0.99
	Decision Tree	0.887(0.855,0.919)	0.82	0.96	0.048	0.59	0.99
	XGboost	0.912(0.883,0.941)	0.82	0.96	0.085	0.57	0.99
	MLP	0.917(0.889,0.945)	0.64	0.98	0.028	0.72	0.98
Childbirth	Logistics regression	0.937(0.912,0.962)	0.83	0.96	0.082	0.59	0.99
	Random Forest	0.935(0.910,0.960)	0.84	0.96	0.067	0.57	0.99
	Decision Tree	0.911(0.882,0.940)	0.87	0.96	0.052	0.55	0.99
	XGboost	0.935(0.910,0.960)	0.87	0.94	0.101	0.46	0.99
	MLP	0.933(0.907,0.959)	0.64	0.99	0.026	0.75	0.98
CDRN (validation site)							
12wk	Logistics regression	0.810(0.801,0.819)	0.70	0.85	0.150	0.24	0.98
	Random Forest	0.788(0.779,0.797)	0.71	0.85	0.144	0.24	0.98
	Decision Tree	0.790(0.781,0.799)	0.71	0.85	0.152	0.24	0.71
	XGboost	0.789(0.780,0.798)	0.71	0.85	0.180	0.24	0.98
	MLP	0.812(0.803,0.821)	0.65	0.87	0.111	0.26	0.97
18wk	Logistics regression	0.817(0.808,0.826)	0.70	0.85	0.151	0.24	0.98
	Random Forest	0.794(0.785,0.803)	0.72	0.84	0.145	0.24	0.98
	Decision Tree	0.794(0.785,0.803)	0.72	0.84	0.152	0.24	0.98
	XGboost	0.793(0.784,0.802)	0.72	0.85	0.180	0.25	0.98
	MLP	0.817(0.808,0.826)	0.65	0.87	0.111	0.26	0.97
24wk	Logistics regression	0.821(0.812,0.830)	0.71	0.85	0.152	0.25	0.98
	Random Forest	0.800(0.791,0.809)	0.73	0.84	0.146	0.24	0.98

Time	Classifier	AUROC	Sensitivity	Specificity	Brier Score	PPV	NPV
30wk	Decision Tree	0.799(0.790,0.808)	0.73	0.84	0.152	0.24	0.98
	XGboost	0.798(0.789,0.807)	0.73	0.85	0.180	0.25	0.98
	MLP	0.824(0.815,0.833)	0.64	0.88	0.110	0.27	0.97
	Logistics regression	0.824(0.815,0.833)	0.72	0.85	0.153	0.24	0.98
	Random Forest	0.807(0.798,0.816)	0.74	0.84	0.148	0.24	0.98
	Decision Tree	0.802(0.793,0.811)	0.73	0.84	0.152	0.24	0.98
	XGboost	0.801(0.792,0.810)	0.73	0.84	0.181	0.25	0.98
Childbirth	MLP	0.827(0.818,0.836)	0.65	0.88	0.110	0.27	0.97
	Logistics regression	0.886(0.879,0.893)	0.80	0.84	0.158	0.26	0.98
	Random Forest	0.860(0.852,0.868)	0.82	0.87	0.154	0.26	0.99
	Decision Tree	0.856(0.848,0.864)	0.86	0.84	0.149	0.27	0.99
	XGboost	0.864(0.856,0.872)	0.84	0.84	0.178	0.27	0.99
	MLP	0.887(0.880,0.894)	0.66	0.88	0.105	0.28	0.97

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript