Boosting for Postpartum Depression Prediction

Sriraam Natarajan[1], Annu Prabhakar[1,2], Nandini Ramanan[1], Anna Baglione[1], Kay Connelly[1], Katie Siek[1]
[1]School of Informatics and Computing, Indiana University- Bloomington
[2]School of Information Technology, University of Cincinnati
{natarasr, asprabha, nramanan, abaglion, connelly, ksiek}@indiana.edu

***Abstract*—Pregnancy and childbirth are important transitional life events for women. Like many other transitional life events, the effects of pregnancy and childbirth can have significant impact on a mother's physical and mental well-being. Sometimes they can even lead to Postpartum Depression (PPD). If left untreated, PPD can be debilitating for the mother and can adversely affect her ability to take care of herself and her infant. Since PPD is not clinically diagnosable, we consider the problem of predicting PPD from survey data about demographics, depression, and pregnancy etc. We adapt the successful functional-gradient boosting algorithm that can handle class imbalance in a principled manner. Our results demonstrate that the proposed machine learning approach can outperform the baseline classifiers and, consequently, demonstrate the potential of machine learning in predicting PPD.**

I. Introduction

Pregnancy and childbirth can have a significant impact on the physical and mental well-being of new mothers. Following their baby's birth, many new mothers experience mood disorders, which include a severe condition called postpartum depression (PPD). PPD is a form of depression that typically begins in the first month after giving birth and is characterized by symptoms including sadness, fatigue, changes in eating and sleeping patterns, reduced libido, crying episodes, anxiety, and irritability [1].

Estimates of prevalence of PPD range from 13% to 19% of new mothers [2] to up to 22.9% of mothers with one child *four years after the birth* [3]. The Centers for Disease Control and Prevention (CDC) estimates that approximately 15% of all mothers experience symptoms that meet criteria for a diagnosis of PPD in the United States [4]. In a recent study of 10,000 women in the United States, 14% screened positive for PPD [5]. This results in nearly $560,000$ cases of PPD annually in the United States alone, some of which may not be diagnosed early. To put this in perspective, more women will suffer from postpartum depression and related illnesses in a year than the combined number of new cases of both leukemia and breast cancer. If left undiagnosed or untreated, PPD can cause adverse effect on parenting abilities, mother-child interactions, breastfeeding, and behavioral and cognitive health outcome for the infant. It has been suggested that 80 percent of women with PPD symptoms do not report them, and when mothers feel seriously depressed, they do not seek help [6], [7], [8], [9]. Therefore, an early diagnosis of PPD could make a big difference for new mothers, caregivers, and clinicians.

While this is an important task, clinical diagnosis of PPD is not well established. Beck and Robertson et al. identified a number of risk factors for PPD, including history of depression or anxiety during pregnancy, history of mental illness, stressful life events, and inadequate social support [10], [1]. Additionally, factors such as marital status, infant's health, and socioeconomic status play a role and contribute to higher than normal occurrence rate of PPD. For example, PPD was found among 23.4% of financially impoverished, inner-city women [11]. In another study, approximately 36% of mothers of infants in the Neonatal Intensive Care Unit experienced symptoms that meet the PPD diagnosis [12]. Since clinical diagnosis for PPD is not well established, effects of social and environmental risk factors such as health and temperament of the baby, lack of self esteem and history of depression need to be evaluated to diagnose PPD.

The Edinburgh Postnatal Depression Scale (EPDS) is widely used by researchers as well as clinicians to measure PPD [13]. The EPDS screening is focused on postpartum mood disorder which does not take into consideration anxiety, irritability, and other symptoms that have been shown to be recurrent among the women, particularly during reproductive-related periods. Demographic information as well as the social support information are also not taken into consideration. Hence, some suggest that the EPDS may not be able to detect the considerable wide range of pre- and postpartum symptoms and disorders [14], [15]. Another study finding suggests a lower positive predictive value using EPDS in a normal population than in the validation study samples [16].

Inspired by the above studies, we propose a machine learning based approach for PPD prediction and diagnosis from survey information. The survey data consists of the demographic information and the PPD risk factors mentioned above. The big question that we ask in this paper is, *Is it possible to predict PPD effectively from non-clinical observations?* We combine the effectiveness of the state of the art machine learning techniques with carefully designed survey questions to answer this question affirmatively.

Previous approaches have used social network data such as tweets and Facebook status messages and phone usages to analyze depression in general [17], [18], [19]. W.r.t PPD, there are fewer studies and most of them employ Facebook and Twitter data; see for instance the work of De Choudhury et al. [20], [21]. While the social network data provides interesting analysis, the information using other validated instruments such as the Revised Postpartum Depression Predictors Inventory (PDPI-R) [22] used in this work merit deeper analysis. They provide a realistic solution to this challenging task of early diagnosis and development of treatment plans.

We make several key contributions in this work. First, we identify the factors that are useful in diagnosis of PPD and collect survey data focusing on these risk factors. Second, we explore the use of novel machine learning techniques on learning from these data to build robust models for PPD diagnosis. Third, we extend the state-of-the-art gradient boosting methods to learn from these data and build an interpretable tree from the set of trees learned using these methods. Finally, we perform extensive quantitative and qualitative analyses on the collected data. Our results conclusively demonstrate the potential of these machine learning methods in advancing early detection of PPD.

The rest of the paper is organized as follows: First, we present the details of the survey design and participant demographic information. Then, we discuss the baseline Machine Learning (ML) approaches for this task. We next outline our gradient-boosted approach. Subsequently, we present our empirical analyses of this challenging problem. Finally, we conclude the paper by discussing areas for future research.

II. Survey Design and Data Collection

We created and distributed a survey to collect demographic information as well as PPD risk factors to predict postpartum depression. This study was approved by the Institutional Review Board (IRB) at Indiana University.

A. Survey Design

The multi-part survey consisted of demographic questions, known PPD risk factors and potential symptoms of PPD. The demographic section collected information such as participant age, infant age, nationality, ethnicity, combined family income, and level of education. The Postpartum Depression Predictors Inventory (PDPI-R) [22], a validated instrument, was used to collect PPD risk factors. This section consisted of the following questions: marital status, socioeconomic status, self-esteem, prenatal depression, prenatal anxiety, whether the pregnancy was unplanned or unwanted, history of previous depression, social support, marital satisfaction, life stress, child care stress, infant temperament, and ”maternity blues.” We also used the post-delivery questions in PDPI-R, which includes infant health, feeding, and sleeping problems, whether the baby is fussy, cries a lot, or is difficult to soothe, and whether the participant has been tearful or experienced mood swings immediately (1 week) after delivery. The list of all the questions and their descriptions are presented in the Appendix.

Since we are in a supervised learning setting, we require labelled data. In order to categorize participants as either having PPD or not, we followed the procedure[1] used by the CDC [4], [23]. Accordingly, we added the following two questions in the survey. 1: ”Since your new baby was born, how often have you felt down, depressed, or hopeless?” 2: ”Since your new baby was born, how often have you had little interest or little pleasure in doing things?” (Answer options for both questions included: always, often, sometimes, rarely, never) [4]. As per CDC procedures, participants were classified as having reported PPD symptoms if they answered ’Always’ or ’Often’ to either of the above questions.

[1] https://www.cdc.gov/reproductivehealth/depression/treatments.htm

B. Survey Responses

We recruited new mothers via social media - specifically, via Facebook groups and Twitter. The inclusion criteria to participate in the survey were as follows: new mothers who were 18 years or older and had a child less than one year old.

We received 207 responses, of which 34 were omitted as their responses were incomplete. We used the rest of the responses ($N = 173$) for our analysis.

25% of the respondents (N=43) were classified as having PPD symptoms. This is higher than the prevalence rate of PPD in the United States, reported by CDC [4]. This may be a result of our respondents coming from different parts of the world, where the PPD prevalence rate may be higher. For example, the PPD rate for India is reported to be 23% [24], and 33% of our respondents were mothers of Indian nationality. The fact that PPD is under-reported and under-diagnosed may also be a contributing factor for higher than US national rate of PPD symptoms in our data. In a recent study of 48 pregnant and new mothers, 33% of the participants found to have PPD symptoms [25].

95% of the mothers who responded to our survey were married, 63% were employed and 51% were first time mothers (see Figure 1 for the distribution of the values of some features). 50% percent were White and 40% were Asian or Asian Americans. Four respondents were African American or Hispanic. Figure 2 presents this distribution. Extending the work to collect responses from races that mirror the national rates of different races remains an interesting future research direction.

III. Boosting Postpartum Depression Predictions Using Machine Learning

Given the responses to the survey questionnaire, our aim is to evaluate if these questions are sufficient for predicting the occurrence of PPD. We turn this problem into a machine learning one. More specifically, we aim to estimate $P(PPD|responses)$, i.e., the conditional distribution of PPD given the responses to the questions. Our hypothesis is that this will allow us to potentially diagnose PPD early in women and develop appropriate treatment plans.

To estimate this conditional distribution, in addition to employing standard machine learning classifiers such as Decision-trees, Naive Bayes, SVM etc., we adapt the recently successful functional-gradient boosting technique [26], [27] and its adaptation to handle class imbalance [28]. In this section, we outline these two methods in greater detail. In the experimental section, we present the list of baselines used.

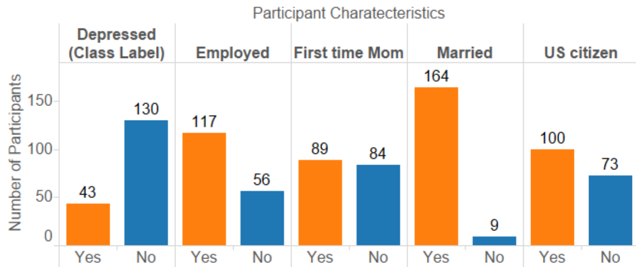

Fig. 1. Demographic Information.

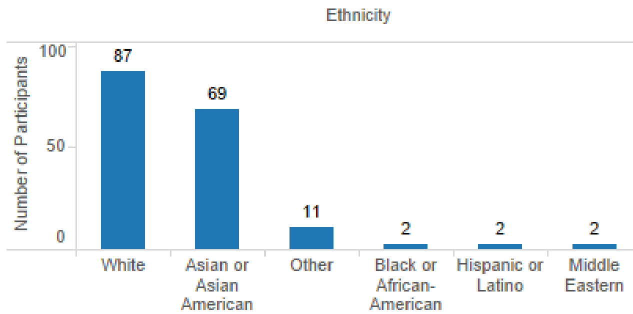

Fig. 2. Participant Ethnicity.

A. Gradient-Boosted PPD Prediction

Recall that the goal is to learn PPD from the survey question data. Following standard machine learning notations, us denote PPD as y and all the questions (features) as $\mathbf{x}$. The key idea in gradient-boosting is to consider a functional representation for the distribution and derive the gradients of the log-likelihood w.r.t the function. Our goal is to fit a model

$$P(y = PPD|\mathbf{x}) = \frac{e^{\psi(y=PPD,\mathbf{x})}}{1 + e^{\psi(y=PPD,\mathbf{x})}} \tag{1}$$

Standard gradient descent methods take the gradient of the log-likelihood w.r.t to the parameters of the distribution. The log-likelihood is given by $LL = \sum_i log(P(y_i|\mathbf{x_i}))$. Functional-gradient on the other hand, takes the gradient of this log-likelihood wrt ψ. Friedman(2001) suggested computing the gradient (weight) for each example separately and fit a regression function over all the weighted examples. This set of local gradients will approximate the global gradient. The functional gradient of each example $\langle \mathbf{x}_i, y_i \rangle$ w.r.t functional $(\psi(y_i = 1; \mathbf{x_i}))$ is

$$\frac{\partial \log P(y_i; \mathbf{x_i})}{\partial \psi(y_i = 1; \mathbf{x_i})} = I(y_i = 1; \mathbf{x_i}) - P(y_i = 1; \mathbf{x_i}) \tag{2}$$

, where I is the indicator function that is 1 if $y_i = 1$ and 0 otherwise. The expression is simply the adjustment required to match the predicted probability with the true label of the example. If the example is positive (PPD) and the predicted probability is less than 1, this gradient is positive indicating that the predicted probability should move towards 1. Conversely, if the example is negative (not PPD) and the predicted probability is greater than 0, the gradient is negative, driving the value the other way.

While Friedman suggested any gradient function, we employed regression trees since they are (1) easy to learn and (2) easy to interpret. Each regression tree can be viewed as defining several new feature combinations, one corresponding to each path from the root to a leaf. The resulting potential functions from all these different regression trees still have the form of a linear combination of features but the features can be quite complex. This is shown in Figure 3.

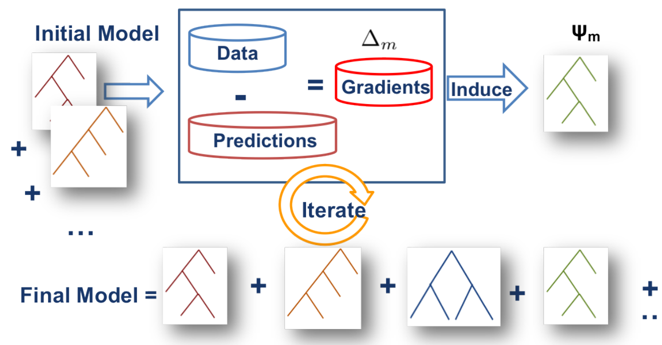

Fig. 3. Functional Gradient Boosting. This is similar to the standard gradient-boosting where trees are induced in stage-wise manner. At every iteration, the gradients are computed as the difference between observed and predicted probabilities of each example and a new regression tree is fitted to these examples.

B. Handling Imbalanced Data

While the above method is powerful, imbalanced data sets can force the algorithm to place higher weights on the majority class. While our data set is not as imbalanced as a rare disease data set [29], the ratio of PPD to non-PPD is 1:3. In order to reduce the false negative rate in FNR, we introduced a cost in the objective function. This approach has been shown to yield good results in imbalanced relational data sets [28] and in predicting rare diseases from survey data [29]. This cost explicitly weighs between predicting as false positive and as false negative, $c(y_i, y) = \mu I(y_i = 1 \land y = 0) + \nu I(y_i = 0 \land y = 1)$, where $I(y_i = 1 \land y = 0)$ is 1 for false negatives and $I(y_i = 0 \land y = 1)$ is 1 for false positives. Intuitively, $c(y_i, y) = \mu$ when a positive example (PPD example) is misclassified, while $c(y_i, y) = \nu$ when a negative example is misclassified. This is similar to the work of Macleod et al, where they considered predicting rare diseases from survey data. We adapt this algorithm in the context of learning to predict PPD from survey data.

Given this new cost function, we add this to the likelihood to obtain a modified likelihood (MLL),

$$MLL = \sum_i \log \frac{\exp\left(\psi(\mathbf{x}_i; y_i)\right)}{1 + \exp\left(\psi(\mathbf{x}_i; y') + cost(y_i, y')\right)} \tag{3}$$

Now, the gradient of this MLL w.r.t $\psi(y_i = PPD; \mathbf{x}_i)$ can be shown as:

$$\frac{\partial \log MLL}{\partial \psi(y_i = PPD; \mathbf{x}_i)} = I(y_i = PPD; \mathbf{x}_i) - \frac{P(y = PPD; \mathbf{x}_i) e^{c(y_i, y=PPD)}}{\sum_{y'_i}[P(y'_i; \mathbf{x}_i) e^{c(y_i, y'_i)}]} \quad (4)$$

The gradients of the objective function can be rewritten compactly as:

$$\Delta = I(\hat{y}_i = PPD) - \lambda P(y_i = PPD; \mathbf{x}_i) \quad (5)$$

where:

$$\lambda = \frac{e^{c(\hat{y}_i, y=PPD)}}{\sum_{y'}[P(y'; \mathbf{x}_i)\ e^{c(\hat{y}_i, y')}]}.$$

For the subjects (examples) with PPD, we obtain:

$$\lambda = \frac{1}{P(y' = PPD; \mathbf{X}_i) + P(y' = NotPPD; \mathbf{X}_i) \cdot e^{\alpha}}.$$

As $\mu \to \infty$, which amounts to putting a large positive cost on the false negatives, $\lambda \to 0$ and the gradients ignore the predicted probability as the gradient is pushed closer to 1 ($\Delta \to 1$), indicating a harsher penalty on misclassified positive examples. On the other hand, when $\nu \to -\infty$, the gradients are pushed closer to 0 ($\Delta \to 0$), indicating more tolerance on misclassified negative. By setting the parameters $\mu > 0$ and $\nu < 0$, the different costs of false positive and false negative examples can be incorporated into the learning process, hence the trade-off between precision and recall can be controlled.

C. Interpretation

While the gradient-boosting method has been demonstrated to achieve superior empirical results across a variety of tasks, a key issue remains with this approach - *interpretation of the trees*. In other words, the quantitative performance of these trees do not particularly aid in the qualitative analysis of the results. This is due to the fact that each subsequent tree is learned as a result of the previous trees i.e., like boosting they “fix” the errors due to the previous trees. So the resulting trees cannot be interpreted without one another. More importantly, since they are all weak learners, considering only a few of them will not be sufficient to explain the full model.

Consequently, we take two approaches to solve this problem - exact and approximate. In the exact method, inspired by the research inside the Arithmetic circuits community [30], we define addition operators on the trees. The intuition is that we add the second tree to every leaf of the first tree (and sum their regression numbers at the leaves) and then post-prune them to reduce the redundant parts of the tree. This is shown in Figure 4. The other alternative is what we call the Craven approach [31] that was presented for making neural networks interpretable. The key idea is to relabel the training data based on the boosted model that we have learned and then train an overfitted tree to this labeled data. The intuition is that this new large tree will represent the decisions made by the original set of trees due to its performance on the training data. Recall that our original training data consists of Boolean labels (PPD vs not PPD). But the relabeled data consists of regression values that are being learned in the new tree. Hence, the resulting tree is closer to the original learned model as we show in our experiments in the next section.

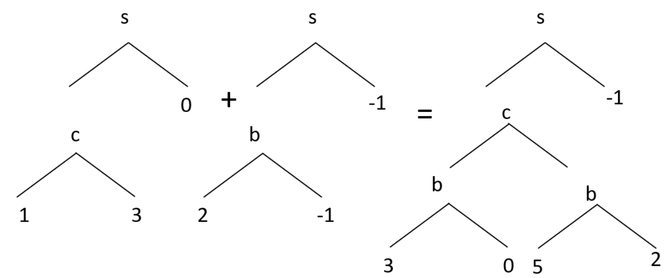

Fig. 4. Addition of two trees. As can be seen, the first two trees are adde analytically (exact). The idea is to place the second tree at each leaf of the first tree, add the corresponding regression values and prune the branches that are not reachable (due to conflicts and redundancies).

IV. Empirical Evaluation

In our experiments, we aim to ask the following questions:

Q1: Can PPD be predicted from non-clinical data?
Q2: Is machine learning a viable tool for PPD prediction?
Q3: Are there large number of weak indicators of PPD rather than a small set of strong indicators?

To answer these questions, we performed 5-fold train/test on the survey data that we collected.

a) Methods Considered: In addition to the gradient boosted tree (that we denote as *FGB*) and its cost-sensitive extension (that we denote as *Soft Margin*), we consider several baselines to rigorously analyze the data. Specifically, we consider the standard Naive Bayes, decision-trees (J48), Logistic Regression and the ensemble methods of Ada Boost and Bagging. Because the data is imbalanced, we consider undersampling of non PPD cases and oversampling of the PPD cases as well along with the state of the art minority over sampling technique called SMOTE. SMOTE [32] generates synthetic examples along the line segments joining the minority examples with their 5 nearest neighbors.

b) Metrics:: We have considered different standard metrics to quantitatively analyze the results. In addition to the standard area under the ROC curve metric, we also consider precision and recall as our metrics. Finally, as with many standard class imbalance problems, it does not suffice to simply report the standard measurements. Hence we also include F_3 and F_5 scores, where it is defined as:

$$F_\beta = (1 + \beta^2) \frac{Precision \cdot Recall}{(\beta^2 \cdot Precision) + Recall}$$

where β is the importance given to recall over precision (i.e. a higher β indicates more emphasis on recall and a smaller β indicates more emphasis on precision). We use F_3 and F_5 to increase the importance of recall over precision. We also report the confusion matrix for each experiment.

	Naive Bayes	J48	SVM	Ada Boost	Bagging	Logistic Regression	Under Sampling	Over Sampling	SMOTE	Soft Margin	FGB
ROC	0.684	0.902	0.709	0.784	0.565	0.756	0.789	0.759	0.789	0.889	0.952
Precision	0.827	0.828	0.794	0.735	0.638	0.766	0.313	0.538	0.667	0.367	0.920
Recall	0.787	0.820	0.780	0.790	0.740	0.810	0.714	0.636	0.545	1.000	0.840
F3	0.790	0.821	0.781	0.784	0.728	0.805	0.633	0.625	0.556	0.853	0.847
F5	0.788	0.820	0.781	0.788	0.735	0.808	0.680	0.632	0.549	0.938	0.843

TABLE I
RESULTS OF RUNNING THE DIFFERENT CLASSIFIER ALGORITHMS. IT CAN BE OBSERVED THAT FGB AND ITS SOFT MARGIN ADAPTATION EXHIBIT THE BEST PERFORMANCE ACROSS ALL THE MEASURES.

As can be seen in Table 1, in every metric either FGB or soft margin outperforms the rest of the classifiers. Specifically in F_3 and F_5 scores that give higher importance to recall (that is recovery of PPD from data), the two gradient boosted methods exhibit statistically significant performance increases over the standard methods. This clearly shows that the two methods are capable of retrieving the occurrence of PPD from the data. Hence, **Q2** can be answered strongly affirmatively in that machine learning can be a viable tool for predicting PPD.

Also, notice that soft margin has a recall of 1 indicating that the false negative rate is 0. In nearly all the other metrics, these methods outperform even the under sampling and over sampling methods. This allows us to answer **Q3** positively. Given that these methods learn a large set of small models, these weak predictors appear to be more robust in PPD classification. This leads us to believe that a condition as complex as PPD does not have a single set of strong indicators but multiple risk factors that may interact in a complex manner.

Finally, one can now answer **Q1** strongly in that PPD can be potentially diagnosed outside the clinic by interacting with the concerned woman. This could potentially allow us to develop self-diagnosis tools for the women who could consult professional help when diagnosed by this tool.

We have presented the tree learned after combining the boosted trees from FGB in Figure 5. This is interesting in many respects. As a general rule, the left branch indicates that the test is satisfied and right branch is when it is not. Looking at the snippet of the tree, the first (highest level) tests are whether the baby is difficult to console and if the women is not depressed before pregnancy. This condition can fail when either the baby is not too difficult or if the woman was not depressed before. Then if the woman is employed, can share with her partner, with an infant who is at least 7 months old and the woman is in her early 30s, then the probability of not having a PPD is very high. This shows that when the woman has a trustworthy partner, has a healthy baby and is more mature, then she does not have depression. Contrast this with the middle branch where the woman is either not employed or cannot confide in her partner, and either does not have a bachelor's degree or has relationship problems; in this case, the probability of PPD is very high (0.99). Similarly, a family move or serious illness can potentially lead to PPD (with probability of 0.75). Additionally, in cases where the probability of having or not having PPD are comparable, the condition being tested at that node provides minimal or no decrease in uncertainty. Without discussing the rest of the tree, we can see that the tree has interesting observations that can allow for developing effective diagnosis and treatment of PPD.

It must be mentioned that parts of our tree confirm with findings of O'Hara and Swain [33]. For example, consider relationship with spouse, OHara and Swain reported: "These correlations were examined separately and results showed a weak association between the mother's relationship with her spouse and the incidence of postpartum depression." Our trees seem to capture the relationship between the mother and spouse as an important attribute. Of course, they mention that " Postpartum depression based on an interview was strongly negatively related to the woman's relationship with her spouse but depression based on self-report questionnaires was not significantly associated". This demonstrates yet another interesting difference between an interview and a survey. To summarize, our findings show the importance of "Martial Relationship Problems in PPD, an observation made earlier [33].

V. CONCLUSION

We considered the problem of predicting postpartum depression outside the clinical setting by analyzing demographic, behavioral, and socioeconomic information. Our initial work in this direction using more advanced machine learning methods demonstrated the effectiveness of this set of information in predicting PPD. Our results also highlighted the potential of machine learning in this challenging yet important task. Our exploration in this direction allows for developing self-diagnosis tools and treatment plans that could help women with PPD. We will explore the use of more data and more sophisticated algorithms to improve the results and realize the possibility of utilizing machine learning for this important task. An important limitation of our current approach is that we treat the CDC questions as ground truth and they are essentially a proxy for the true diagnosis. Modeling the accuracy of this approximation remains an interesting future challenge. We will explore the prediction of answers to these questions as our next research problem.

ACKNOWLEDGMENT

SN gratefully acknowledges National Science foundation grant no. IIS-1343940. SN and NR acknowledge the support of Indiana University's Precision Health Initiative.

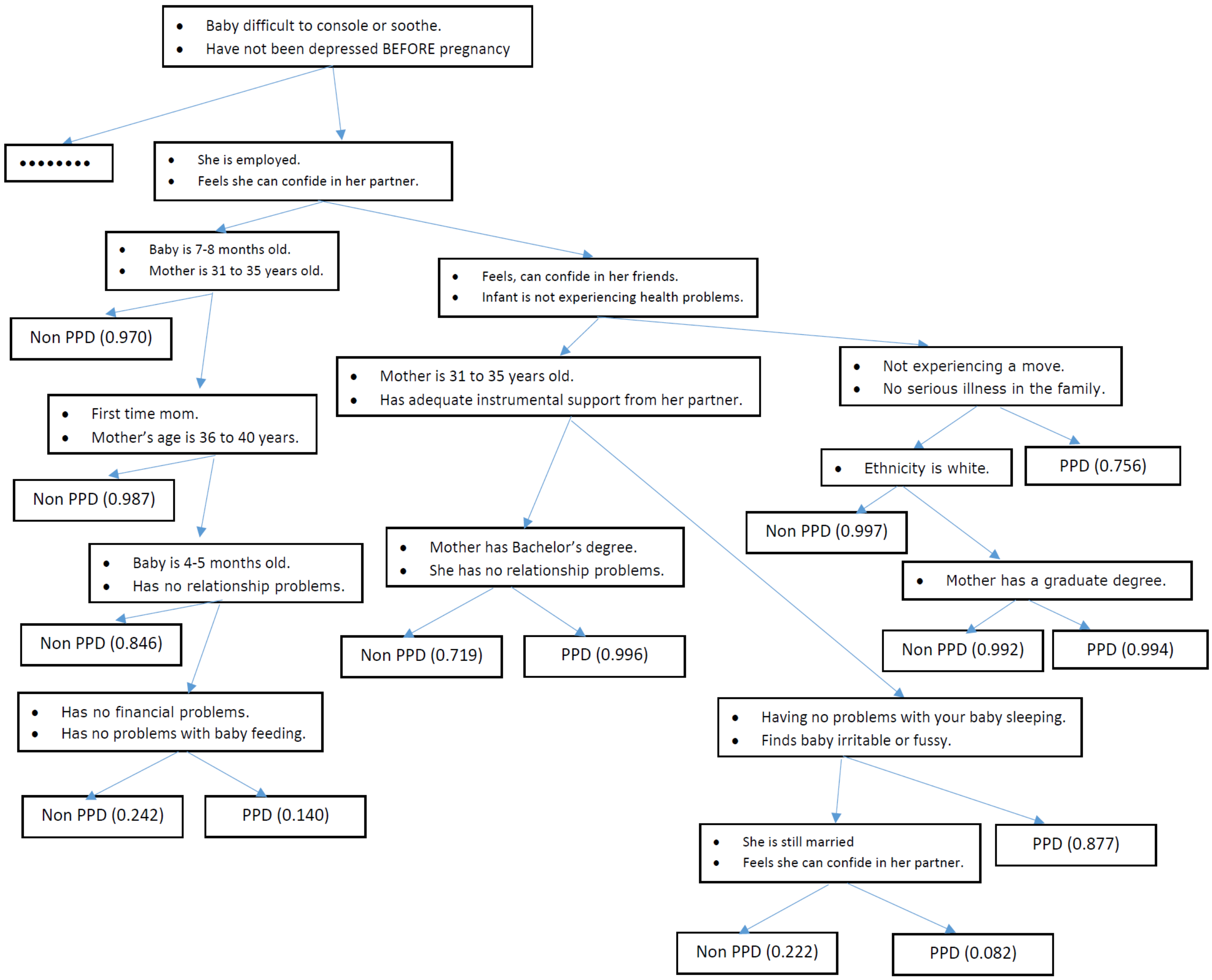

Fig. 5. Final Tree learned from the model. The left branch corresponds to the condition (test) being true and the right branch corresponds to it being false.

References

[1] C. T. Beck, "Predictors of postpartum depression: an update," *Nursing research*, vol. 50, no. 5, pp. 275–285, 2001.

[2] M. W. O'hara and J. E. McCabe, "Postpartum depression: current status and future directions," *Annual review of clinical psychology*, vol. 9, pp. 379–407, 2013.

[3] H. Woolhouse, D. Gartland, F. Mensah, and S. J. Brown, "Maternal depression from early pregnancy to 4 years postpartum in a prospective pregnancy cohort study: implications for primary health care," *BJOG: Int J Obstet Gy*, vol. 122, no. 3, pp. 312–321, Feb. 2015. [Online]. Available: http://dx.doi.org/10.1111/1471-0528.12837

[4] C. for Disease Control, P. (CDC *et al.*, "Prevalence of self-reported postpartum depressive symptoms–17 states, 2004-2005." *MMWR. Morbidity and mortality weekly report*, vol. 57, no. 14, p. 361, 2008.

[5] W. KL, S. DY, M. MC, and et al, "Onset timing, thoughts of self-harm, and diagnoses in postpartum women with screen-positive depression findings," *JAMA Psychiatry*, vol. 70, no. 5, pp. 490–498, 2013. [Online]. Available: + http://dx.doi.org/10.1001/jamapsychiatry.2013.87

[6] R. H. Kelly, D. F. Zatzick, and T. F. Anders, "The detection and treatment of psychiatric disorders and substance use among pregnant women cared for in obstetrics," *American journal of psychiatry*, vol. 158, no. 2, pp. 213–219, 2001.

[7] A. Whitton, R. Warner, and L. Appleby, "The pathway to care in postnatal depression: women's attitudes to post-natal depression and its treatment." *Br J Gen Pract*, vol. 46, no. 408, pp. 427–428, 1996.

[8] K. A. Yonkers, S. M. Ramin, A. J. Rush, C. A. Navarrete, T. Carmody, D. March, S. F. Heartwell, and K. J. Leveno, "Onset and persistence of postpartum depression in an inner-city maternal health clinic system," *American Journal of Psychiatry*, vol. 158, no. 11, pp. 1856–1863, 2001.

[9] A. MacLennan, D. Wilson, and A. Taylor, "The self-reported prevalence of postnatal depression," *Australian and New Zealand journal of obstetrics and gynaecology*, vol. 36, no. 3), 1996.

[10] E. Robertson, S. Grace, T. Wallington, and D. E. Stewart, "Antenatal risk factors for postpartum depression: a synthesis of recent literature," *General hospital psychiatry*, vol. 26, no. 4, pp. 289–295, 2004.

[11] S. E. Hobfoll, C. Ritter, J. Lavin, M. R. Hulsizer, and R. P. Cameron, "Depression prevalence and incidence among inner-city pregnant and postpartum women." *Journal of consulting and clinical psychology*, vol. 63, no. 3, p. 445, 1995.

[12] N. N. Tahirkheli, A. P. Tackett, M. A. McCaffree, and S. R. Gillaspy, "Postpartum depression on the neonatal intensive care unit: current perspectives." *International Journal of Women's Health*, vol. 6, pp. 974–

987, 2014.
[13] J. L. Cox, J. M. Holden, and R. Sagovsky, “Detection of postnatal depression. development of the 10-item edinburgh postnatal depression scale.” *The British journal of psychiatry*, vol. 150, no. 6, pp. 782–786, 1987.
[14] U. Halbreich and S. Karkun, “Cross-cultural and social diversity of prevalence of postpartum depression and depressive symptoms,” *Journal of affective disorders*, vol. 91, no. 2, pp. 97–111, 2006.
[15] F. Abdollahi, M.-S. Lye, A. M. Zain, S. S. Ghazali, and M. Zarghami, “Postnatal depression and its associated factors in women from different cultures,” *Iranian journal of psychiatry and behavioral sciences*, vol. 5, no. 2, p. 5, 2011.
[16] M. Eberhard-Gran, A. Eskild, K. Tambs, S. Opjordsmoen, and S. Ove Samuelsen, “Review of validation studies of the edinburgh postnatal depression scale,” *Acta Psychiatrica Scandinavica*, vol. 104, no. 4, pp. 243–249, 2001. [Online]. Available: http://dx.doi.org/10.1111/j.1600-0447.2001.00187.x
[17] M. A. Moreno, L. A. Jelenchick, and K. G. Egan, “Feeling bad on facebook: depression disclosures by college students on a social networking site. depression and anxiety 28 (6): 447–455,” 2011.
[18] S. Tsugawa, Y. Mogi, Y. Kikuchi, F. Kishino, K. Fujita, Y. Itoh, and H. Ohsaki, “On estimating depressive tendencies of twitter users utilizing their tweet data,” in *2013 IEEE Virtual Reality (VR)*. IEEE, 2013, pp. 1–4.
[19] S. Saeb, M. Zhang, C. J. Karr, S. M. Schueller, M. E. Corden, K. P. Kording, and D. C. Mohr, “Mobile phone sensor correlates of depressive symptom severity in daily-life behavior: An exploratory study,” *Journal of medical Internet research*, vol. 17, no. 7, 2015.
[20] M. De Choudhury, S. Counts, E. J. Horvitz, and A. Hoff, “Characterizing and predicting postpartum depression from shared facebook data,” in *Proceedings of the 17th ACM conference on Computer supported cooperative work & social computing*. ACM, 2014, pp. 626–638.
[21] M. De Choudhury, S. Counts, and E. Horvitz, “Predicting postpartum changes in emotion and behavior via social media,” in *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. ACM, 2013, pp. 3267–3276.
[22] C. T. Beck, K. Records, and M. Rice, “Further development of the postpartum depression predictors inventory-revised,” *Journal of Obstetric, Gynecologic, & Neonatal Nursing*, vol. 35, no. 6, pp. 735–745, 2006.
[23] M. A. Whooley, A. L. Avins, J. Miranda, and W. S. Browner, “Case-finding instruments for depression. two questions are as good as many,” *Journal of general internal medicine*, vol. 12, no. 7, pp. 439–445, 1997.
[24] V. Patel, M. Rodrigues, and N. DeSouza, “Gender, poverty, and postnatal depression: a study of mothers in goa, india,” *American Journal of Psychiatry*, vol. 159, no. 1, pp. 43–47, 2002.
[25] A. Prabhakar, L. Guerra-Reyes, V. M. Kleinschmidt, B. Jelen, H. MacLeod, K. Connelly, and K. Siek, “Investigating the suitability of the asynchronous, remote, community-based method for pregnant and new mothers,” in *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, 2017.
[26] J. H. Friedman, “Greedy function approximation: A gradient boosting machine,” *Annals of Statistics*, pp. 1189–1232, 2001.
[27] S. Natarajan, T. Khot, K. Kersting, and J. Shavlik, *Boosted Statistical Relational Learners: From Benchmarks to Data-Driven Medicine*. SpringerBriefs in Computer Science, 2015.
[28] S. Yang, T. Khot, K. Kersting, G. Kunapuli, K. Hauser, and S. Natarajan, “Learning from imbalanced data in relational domains: A soft margin approach,” in *2014 IEEE International Conference on Data Mining, ICDM 2014*, 2014, pp. 1085–1090.
[29] H. MacLeod, Y. S, K. Oakes, K. Connelly, and S. Natarajan, “Identifying rare diseases from behavioural data:a machine learning approach,” in *CHASE ’16*, 2016.
[30] A. Darwiche, “A differential approach to inference in bayesian networks,” *J. ACM*, vol. 50, no. 3, pp. 280–305, May 2003.
[31] M. Craven and J. Shavlik, “Extracting tree-structured representations of trained networks,” in *NIPS*, 1996, pp. 24–30.
[32] N. V. Chawla, K. W. Bowyer, L. O. Hall, and W. P. Kegelmeyer, “SMOTE: synthetic minority over-sampling technique,” *Journal of Artificial Intelligence Research*, vol. 16, pp. 321–357, 2002.
[33] M. W. O’hara and A. M. Swain, “Rates and risk of postpartum depressiona meta-analysis,” *International Review of Psychiatry*, vol. 8, no. 1, pp. 37–54, 1996.

Table A: Survey Questions

Theme	Question	Answer
Demographic Questions	Please pick your age range	{less than 20 years, 20-24 years, 25-30 years, 31-35 years, 36-40 years, 41-45 years, more than 45 years}
	How old is your baby? (e.g., 4 weeks, 3 months etc)	Text
	Are you a US citizen?	{Yes, No}
	Please enter your nationality (e.g. India, Korea, England etc.)	Text
	Please enter the name of the country you live in	Text
	Please enter the zip code or postal code you live in	Text
	How many children do you have?	{1, 2, 3, 4 or more}
	Which of the following race/ ethnicity do you most closely Identify with?	{American Indian or Alaska Native, Asian or Asian American, Black or African-American, Native Hawaiian or Other Pacific Islander, Middle Eastern, White, Hispanic or Latino} or "Other: " and Text
	Highest level of education	{No high school diploma or GED, High School Diploma, GED, Some College, Trade or Vocational School, Associate's Degree, Bachelor's Degree, Graduate Degree, Doctoral Degree, Professional or other Terminal Degree}
	Employment status	{Employed for wages; Self-employed; Out of work for more than one year; Out of work for less than 1 year; Not employed, looking for work; Not employed, not looking for work; A homemaker; A student; Unable to work}
	Family's combined annual income (if not dollar amount, use "other" and enter the income amount in your currency - e.g 100,000 Rs)	{Under \$20,000, Between \$20,000 and \$40,000, Between \$40,000 and \$70,000, Between \$70,000 and \$100,000, Between \$100,000 and \$150,000, Greater than \$150,000} or "Other: " and Text
Marital Status	Relationship Status	{Single, Living with partner, Married, Widowed, Divorced, Separated, Never been married}
Self-Esteem	Do you feel good about yourself as a person?	{Yes, No}
	Do you feel worthwhile?	{Yes, No}
	Do you feel you have a number of good qualities as a person?	{Yes, No}
Prenatal Depression	Have you felt depressed DURING your pregnancy?	{Yes, No}
	When and how long have you felt depressed during pregnancy?	{Throughout pregnancy, In the first trimester, In the second trimester, In the third trimester}
	How mild or severe would you consider your depression was during pregnancy?	{Very mild, Mild, Severe, Very severe}
	Have you talked to your provider about depression during pregnancy?	{Yes, No}
Prenatal Anxiety	Have you felt anxious during your pregnancy?	{Yes, No}
	How long did you feel anxious?	{Throughout pregnancy, In the first trimester, In the second trimester, In the third trimester}
History of Previous Depression	Have you ever been depressed BEFORE pregnancy?	{Yes, No}
	When did you experience depression before being pregnant?	{Few months before, few years before, Throughout adulthood, In childhood, In teenage years} or "Other: " and Text

Continued on next page

Table A – Continued from previous page

Theme	Question	Answer
History of Previous Depression	Have you been under a physician's care for depression before being pregnant?	{Yes, No}
	Did the physician prescribe any medication for your depression before being pregnant?	{Yes, No}
Unwanted/Unplanned Pregnancy	Was your pregnancy planned?	{Yes, No}
	Was your pregnancy unwanted?	{Yes, No}
Social Support	Do you feel you receive adequate emotional support from your partner?	{Yes, No}
	Do you feel you receive adequate instrumental support from your partner (e.g. help with household chores or taking care of baby)?	{Yes, No}
	Do you feel you can rely on your partner when you need help?	{Yes, No}
	Do you feel you can confide in your partner?	{Yes, No}
	Do you feel you can confide in your family?	{Yes, No}
	Do you feel you can confide in your friends?	{Yes, No}
Marital Satisfaction	Are you currently experiencing any marital/ relationship problems?	{Yes, No}
	Are you satisfied with your marriage/relationship?	{Yes, No}
	Are things going well between you and your partner?	{Yes, No}
Life Stress	Are you currently experiencing any stressful events in your life (pick one or more) such as:	{Financial problems, Relationship problems, Death in the family, Serious illness in the family, Moving, Unemployment, Job Change, None}
Child Care Stress	Is your infant experiencing any health problems?	{Yes, No}
	Are you having problems with your baby feeding?	{Yes, No}
	Are you having problems with your baby sleeping?	{Yes, No}
Infant Temperament	Would you consider your baby irritable or fussy?	{Yes, No}
	Does your baby cry a lot?	{Yes, No}
	Is your baby difficult to console or soothe?	{Yes, No}
Maternity Blues	Did you experience a brief period of fearfulness and mood swings the first week after delivery?	{Yes, No}
PPD Screening Questions (CDC)	Since your new baby was born, how often have you felt down, depressed, or hopeless?	{Always, Often, Sometimes, Rarely, Never}
	Since your new baby was born, how often have you had little interest or little pleasure in doing things?	{Always, Often, Sometimes, Rarely, Never}