# Predicting Antenatal Depression during 3rd Trimester: A Machine Learning Approach with Feature Selection and TOPSIS Ranking

Atena Karimi[1,2], Ashkan Mozdgir[3*], Reza Tashakkori[4,2], Abolfazl Bagheri Tofighi[2,5], Shabnam Bozorgzadeh[6]

[1]Department of Industrial Engineering, Faculty of Engineering, Kharazmi University, Tehran, Iran

Atenakarimi2000@yahoo.com

[2] Barbod Artificial Intelligence Lab, Tehran, Iran

[3]Department of Industrial Engineering, Faculty of Engineering, Kharazmi University, Tehran, Iran

a.mozdgir@khu.ac.ir

[4]Department of Industrial Engineering, Faculty of Engineering, Kharazmi University, Tehran, Iran

rezatashakkory@gmail.com

[5]Department of Industrial Engineering & Management Systems, Amirkabir University of Technology (Tehran Polytechnic), Tehran, Iran

abolfazl.bagheri@aut.ac.ir

[6]Department of Obstetrics and Gynecology, Iran University of Medical Sciences, Tehran, Iran

sh.bozorgzade@gmail.com

**Abstract**

This study aimed to predict antenatal depression in erythroid women during their 3rd trimester. We investigated prediction using four feature selection methods: Extra-tree classifier, Fisher score, and PCA. We also merged the common features of the Extra-tree classifier and Fisher score and applied them to predicting antenatal depression in the 3rd trimester of pregnancy. We gathered data from 62 women and their corresponding 18 attributes and evaluated them using the Hamilton depression rating scale (HAM-D). Seven ML models were implemented to predict antenatal depression, including k-nearest neighbors, Support vector machine, random forest, decision tree, bagging classifier, multi-layer perception, and naïve Bayes. Therefore, the trained models were evaluated using various metrics, including accuracy, sensitivity, specificity, precision, F1 score, FNR, FPR, and area under the receiver operating characteristic curve. Ultimately all models were prioritized using TOPSIS with different feature selection methods, and the best model was found to be DT without implementing any feature selection. The results of this study show the most important factors in predicting depression in the 3rd trimester of pregnancy.

**Keywords:** Antenatal Depression, Pregnancy, Hamilton depression rating scale, Classification Algorithm, Decision Tree

## 1. Introduction

A woman's childbearing years are associated with the highest risk of depression in her lifetime [1]. This may be due to the fact that mental illnesses like depression and anxiety can be triggered substantial physiological changes [2], as well as worries about the well-being of the unborn child during pregnancy [3]. It is widely recognized that a significant number of pregnant women encounter emotional distress and anxiety throughout their pregnancy period [4]. It is estimated that depression is the most prevalent mental disorder during pregnancy, with prevalence rates ranging between 4% and 20% [5].

Antenatal Depression, a type of depression that impacts women during pregnancy, occurs in approximately 1 out of every 7 cases, with estimated depression rates from 7 to 37.1% [4] and increasing in frequency throughout the pregnancy [6]. Studies show that low- and lower-middle-income countries have a higher occurrence of antenatal depression [7], but in both developed and developing countries, antenatal depression, and anxiety are notable risk factors for postnatal depression [8]. Globally, expectant mothers encounter anxiety during the first, second, and third trimesters at rates of 18.2%, 19.1%, and 24.6%, respectively [9]. Another research shows that first- and third-trimester depression is more common than second-trimester depression [10]. The reason for this might be that pregnant women have to prepare themselves for the problems and challenges of their upcoming life during the first trimester, and during the third trimester, when they are about to deliver their baby, they experience a great deal of stress and anxiety for parturition safely and finally becoming a mother [11]. It was also found that antenatal depression was associated with maternal perceptions of male-child preference among Iranian women. Additionally, their husbands and families exert more concealed pressure on healthy male-child favorability, leading them to depression during pregnancy [12]. As depression is a prevalent ailment among pregnant Iranian women, it seems crucial to identify and treat depression while pregnant. [13]. The importance of antenatal depression cannot be overstated: It is one of the most common psychiatric disorders, which has a severe adverse impact on both the mother and the infant [14]. Studies show that depression can cause psychological issues due to physical and hormonal changes, as well as anxiety toward labor or fetal outcomes [15]. Both the mother and the child can be adversely affected by maternal depression, anxiety, and stress during pregnancy in the long term [16]. Also, poor offspring health outcomes are associated with antenatal depression [5], and preterm birth was more likely to occur among those with depressive symptoms [17]. It is worth highlighting that, pregnant women identified with depression face an elevated risk of low birth weight, preterm birth, and cesarean sections (CSs) [18]. Additionally, according to the findings, maternal prenatal stress negatively affects offspring's socioemotional development, whereas prenatal depression has a larger effect size than anxiety [19]. Nevertheless, it's potentially adverse effects on both the pregnant woman and her unborn child, prenatal depression are less studied and considered than postpartum depression [20]. Also, both prenatal and postnatal depression can be caused by experiencing new life responsibilities and family expectations [21], but very little research has been conducted on antenatal depression [22]. Studies say that prenatal depression is more common and wide-spreading than postnatal depression [7] and depression rates tend to be higher during pregnancy than during the first year following delivery [23]. Previous research has shown depression and anxiety following childbirth to decrease rather than increase, although postpartum depression and anxiety have historically been the focus of more research than antenatal depression [24].

Women suffering from antenatal depression are more likely to develop postpartum depression, preeclampsia, abuse of substances, hemorrhage, develop edema, premature rupture of membranes, and suffer from severe headaches [7]. A noteworthy discovery was made, indicating that depression during pregnancy was a strong indicator of experiencing depression after childbirth [4]. Furthermore, mothers who have previously undergone antenatal depression are more prone to developing postponal depression [25]. It appears that the antenatal state plays an important role in predicting postpartum depression [26], and depression during pregnancy has been identified as one of the most significant risk factors for postnatal depression in studies focusing on the perinatal period [27]. So, in order to reduce postnatal depression and anxiety, it will be most important to treat symptoms of prenatal depression and anxiety [28]. So, it is recommended that pregnant women be aware of the symptoms of antenatal depression and seek treatment if necessary. Although the cause of antenatal depression is unknown, experts believe it is due to biological, psychological, and social factors [20]. Results showed that unemployment [7], unwanted or unexpected pregnancy [1], history of depression [2], lack of social support or partner [29], marital status [20], history of abuse or violence [2], miscarriage/pregnancy loss [30], adverse experiences in life [2], and high perceptions of stress [2], being a housewife [1], the occurrence of pregnancy complications [31], satisfaction in relationship with partner [2], newborn's gender [12], poor socioeconomic status [32], and low-income were most important risk factors and features associated with antenatal depression. Moreover, one of the most crucial factors affecting and enhancing pregnant women's mental health is their empowerment in making decisions regarding financial authority, parenting, and pregnancy [33].

As antenatal and postnatal depression are highly prevalent and have severe consequences, prevention seems imperative [34]. It is possible to identify most factors contributing to antenatal depression at the beginning of pregnancy and pay more attention to prenatal health care. Also, the 'Foetal programming' theory suggests that the perinatal period is the crucial time for mental health protection efforts, and prevention models should be developed during this period [35]. So, we can modify several antenatal depression risk factors with proper interventions at the right time [20]. Unfortunately, an estimated 50% of pregnant women do not receive a diagnosis of depression or anxiety during pregnancy [36]. Besides, women who are only screened once throughout their pregnancy may also find it difficult to diagnose antenatal depression [2], but In order to prevent or treat depression during pregnancy, proper screening for depression must be conducted [20]. As a matter of fact, the majority of women who suffer from perinatal depression do not appear to be interested in receiving treatment [37]. The results of these studies suggest that it can be deduced that prioritizing prevention over treatment is of greater significance. The utilization of machine learning (ML) tools and classifiers for prediction and forecasting purposes can be extremely beneficial in in various fields of science. Ashourloo et al. predicted wheat yield based on Sentinel-2, regression, and machine learning models [38]. YAO et al. improved Support Vector Machine Regression in Multi-Step-Ahead Prediction for Tunnel Surrounding Rock Displacement [39]. In addition, several studies are using ML classifiers in healthcare problems. Sabahi et al. developed an ML model to predict hospital mortality [40]. Haghbayan et al. used machine learning algorithms to increase the efficiency of umbilical cord blood banks [41]. Tashakori et al. developed several ML models to predict hospital NICU admission [42]. According to the findings of Shin D et al. study, the random forest performed the best out of nine supervised ML classifiers, with an AUC of 88.4% in predicting postpartum depression [43]. In Javed F and al's research, MLP and SVM classifiers were used to predict antenatal depression and anxiety with an AUC of 88% and 80%, respectively [29].

We found that this common mental disorder affects many pregnant women, particularly in the later stages of their pregnancy; however, our review discovered that studies on this subject are scarce and limited. Given these findings, greater attention should be paid to pregnant women's mental health, and it is important for individuals to be aware of the symptoms and risk factors associated with antenatal depression and to seek treatment if necessary. The studies we considered also reveal that the third trimester of pregnancy is the most vital period for antenatal depression. Thus, we focused on using ML to predict antenatal depression during this period which is an essential tool for timely prevention and action. Currently, the lack of using ML tools in research on antenatal depression hinders progress in this area. Based on our review, a significant amount of research is focused on postpartum depression rather than antenatal depression. Furthermore, to prevent or effectively intervene in antenatal depression, it is crucial to understand the risk factors that may contribute to its development. Ultimately, this study seeks to bridge the identified gap by establishing a foundation for efficient prevention, treatment, and comprehensive comprehension of antenatal depression. The main contributions of this research are:

- Different ML algorithms were developed to predict antenatal depression in the 3rd trimester of pregnancy based on real data.
- For each algorithm, the most important features for predicting prenatal depression in the 3rd trimester were determined.
- The TOPSIS method was implemented as multi criteria decision making (MCDM) to rank the performance of ML algorithms based on ML evaluation metrics.

These contributions are discussed in detail in the following sections.

## 2. Experimental setup

The objective of this study is to anticipate antenatal depression during the 3rd trimester, with the aim of halting its progression. We employed a comprehensive approach utilizing seven ML algorithms, namely KNN, DT, RF, Bagging, SVM, MLP, and NB. The research is conducted through three key stages, namely pre-processing, modeling, and evaluating. The research is conducted through three key stages, namely pre-processing, modeling, and evaluating, as illustrated in Fig 1.

**#Please Insert Figure 1 around here#**

### *2.1 Dataset*

In this study, we aim to create a model for predicting antenatal depression with the Iranian Hospital, Tehran, Iran dataset. This dataset conducted a nested case-control study between 2014 and 2016 on hypothyroid pregnant women [36]. This study includes 62 women who were pregnant with either boys (49.5%) or girls (50.28%) and had normal pregnancies or prenatal depression. Among mothers with depression, 17 (27.87%) are expecting boys, and 12 (19.67%) have girls. The women who participated in the study were between 18 and 40 years old, without any autoimmune or systemic diseases field [36], and had a BMI between 20.07 and 30.44. Additionally, they had no previous history of pregnancy complications, did not use infertility medication, had a healthy socio-economic status, and did not have any history of psychotic disorders or recent experiences with traumatic events (such as the loss of a loved one, or financial difficulties) or mood disorders field [36]. The exclusion criteria included unwillingness to participate, autoimmune or thyroid disease that occurred during pregnancy, new-onset systemic disorders, medication use other than typical supplements, hyperemesis gravidarum, pregnancy-related abnormalities, and major life changes [36]. Overall, the dataset consisted of 176 attributes including answers to the Hamilton depression questionnaire and postpartum-related features that were either omitted or grouped into one feature. Table 1 depicts the details of all attributes.

**#Please Insert Table 1 around here#**

### *2.2 Data preprocessing*

The process of data preparation is a critical phase in which data is made ready for analysis through various procedures such as cleaning, transforming, and integrating. The goal of data preprocessing is to improve the quality of data for a specific data mining operation. This involves various steps such as removing outliers, handling missing values, dealing with categorical variables, standardization, and selection of relevant features to be used in building classifiers.

### *2.3 Dealing with Outliers*

Given that our dataset is relatively small, we cannot rely on outlier removal [44]. Instead, we need to replace such extreme values with alternatives. To handle outliers in the anti-TPO, we carefully examined the box plot and relied on the interquartile range to identify any anomalies. We replaced such outliers with missing data. In addition, we removed the record entirely from the delivery age when a patient indicated a miscarriage occurring at 9 weeks of gestation to maintain data consistency.

### *2.4 Dealing with missing values*

Table 2 presented the number of missing values in each feature. In this dataset, some variables had missing values that were filled using information derived from other variables of the same participant. For instance, there were no missing values in the BMI for any of the trimesters, but for some individuals, it was zero, which is not a meaningful value for BMI and thus was replaced with the Nan value. At first, we decided to fill in the missing values which were related to the weight of the patients. We found that the difference between the mother's weight before and after pregnancy in the first trimester was an average of five kilograms, which we used and fill in the missing values in these features. The same approach was applied to all weights in the different trimesters. After imputing missing values with proper values, we were able to use the patients’ height to fill in the missing values in the BMI. To determine the child's gender, we used the type of delivery, which showed that most male children were born through a procedure other than natural delivery, while most female children were born naturally. The delivery age was populated with the median of the recent miscarriage for common complaint pregnancy. Finally, we replaced missing values in the anti-TPO with its average value.

**#Please Insert Table 2 around here#**

### *2.5 Dealing with categorical columns*

In order to transform categorical features into numerical ones, we utilized the Label Encoder from the Sklearn library, given that there were a limited number of distinct values present in these features. Some of the features in our dataset were binary and contained only two values, while others were more diverse, such as the mother's education, which ranged from sub-Diploma to Ph.D. However, since these values have an inherent ranking relationship, they could be easily encoded using this method.

### *2.6 Normalization*

In order to improve the performance of our models, we utilized the Standard-Scaler function from the Sklearn library, which allowed us to scale our features to a range of 0 to 1. By doing so, we were able to ensure that all of our variables were within the same range, which is an important step in improving model performance.

### *2.7 Balance*

Based on the information presented in the pie chart in Fig 2, we can infer that the distribution of antenatal depression prevalence in the data is relatively even.

**#Please Insert Figure 2 around here#**

### *2.8 Model Development*

The dataset was divided into two distinct sets: an 80% portion designed as the training set, the remaining 20% allocated as the testing set, and a specific shuffle mode was used to maintain consistency in the results. Feature selection techniques were utilized to enhance the performance of each model employed. Additionally, the parameters of each model were adjusted to find the optimal combination that would yield the highest possible accuracy.

### *2.9 Feature Selection*

The purpose of the feature selection step is to identify a limited set of features that are relevant and useful for the analysis by removing redundant and irrelevant features [45]. In our study, we evaluated four different feature selection methods: PCA (principal component analysis), XTC (extra tree classifier), Fisher score, and a combination of XTC and Fisher score. The ranking of each feature based on each method is presented in Table 3, where a higher ranking number indicates the greater significance of the feature. Interestingly, the variable related to the father's job did not receive a ranking number in the Fisher score analysis. We selected the top 12 features based on their ranking scores for the Fisher score method and utilized the top 13 features for the Extra Tree Classifier.

**#Please Insert Table 3 around here#**

### *2.10 Feature importance*

In ML, determining feature importance as well as visualizing related results are crucial [45]. Given its straightforwardness and ability to be easily understood, feature ranking proves especially valuable in fields like biomedicine and social sciences [46]. the value of the coefficient associated with each feature is employed to ascertain the significance and ranking of each feature [47]. The feature importance or coefficient scores of these classifiers are presented in their respective sections. The ranking and importance of features are also visualized in Fig 3.

**#Please Insert Figure 3 around here#**

## 3. Model Development

The data set was partitioned into a training subset and a testing subset. Then, different classifier were developed to predict antenatal depression in the 3rd trimester of pregnancy.

## 3.1 Classification

ML classification is a key aspect of AI, allowing for precise data predictions and categorization. In this study, we utilized SVM, DT, RF, KNN, NB, MLP, and Bagging classifiers.

### *3.2 Evaluation Metrics*

Various evaluation metrics, such as accuracy (how often does a machine learning model correctly predict an outcome), recall (the portion of the positive samples is correctly labeled), specificity (percentage of true negatives correctly identified), precision (how often does a machine learning model correctly predict the possetive class), F1 score (the average of precision and recall), false possetive rate (FPR) , and false negative rate (FNR), were employed to gauge the performance of the models.

#### *3.2.1 Confusion Matrix:*

The confusion matrix is a tabular representation that assesses the performance of a classification algorithm. It provides insight into the model's performance and allows easy comparison between different classifiers. In contrast, the confusion matrix provides essential metrics such as accuracy, precision, recall, and f1-score, enabling a comprehensive evaluation of a classification model's performance [48]. Table 4 demonstrates the confusion matrix of our study. The following formulas were developed based on this study [49].

**#Please Insert Table 4 around here#**

Python 3.8.8 is used to develop and apply all supervised ML models. The entire task was executed on a Windows operating system installed on an Intel(R) Core(TM) i7-10750H CPU @ 2.60GHz, operating 2592 Mhz, with 6 Core(s) and 12 Logical Processor(s).

#### *3.2.2 Topsis*

Technique for Order Preference by Similarity to Ideal Solutions (TOPSIS) is a numerical technique utilized in the field of multi-criteria decision-making (MCDM). The mathematical model is straightforward and can be applied to a wide range of applications [50]. Furthermore, TOPSIS serves as a compensatory method that permits trade-offs among criteria, enabling a satisfactory outcome in one criterion to compensate for a subpar result in another [51]. Within this approach, the optimal solution is determined by its proximity to the positive-ideal solution and its distance from the negative-ideal solution [52]. Consequently, the alternatives are ranked based on an overall index derived from the distances between the alternatives and the ideal solution [53].

## 4. Result and Discussion

### *4.1 Result of exploratory data analysis (EDA)*

In this study, Exploratory Data Analysis (EDA) method was utilized to analyze and summarize the dataset, to gain a better understanding of its characteristics [45]. A heatmap in Fig 4 was generated to demonstrate the correlation between features and antenatal depression in 3rd trimester. The strength of the correlation is represented by the lighter color of the cell. Additionally, KDE plots were produced in Fig 5 to show the density distribution between depressed and non-depressed mothers. According to the KDE plot, mothers with a BMI between 25.8-38, age between 19-45, anti-TPO serum between 1-132, and a monthly income of 1.2-8 million toman are more likely to experience antenatal depression. Furthermore, the bar plot for discrete features is plotet and one of the bar plots in Fig 6 was created to display the probability of antenatal depression among mothers expecting male infants.

**#Please Insert Figure 4 around here#**

**#Please Insert Figure 5 around here#**

**#Please Insert Figure 6 around here#**

### *4.2 Result of machine learning analysis*

In this research, various classification algorithms were utilized including DT, bagging DT-based, RF, MLP, SVM, KNN, and NB after data preprocessing which involved detecting and replacing outliers and missing values, standardizing, and encoding. The predictor's performance was evaluated using the train-test-split technique, and assessment metrics such as accuracy, sensitivity, specificity, precision, F1 Score, AUC-ROC, FPR, and FNR were employed. The predictors were ranked using topsis for four different feature selection methods including Fisher score, XTC, PCA, and the combination of Fisher score and XTC. The most promising result was achieved using DT without implementing any feature selection methods. In each algorithm, the best result is found based on a variety of different feature selection methods which contain DT without any feature selection with the topsis rank of 1, bagging with

Fisher score with the rank of 2, SVM with no feature selection with a rank of 10, SVM without any feature selection with a rank of 10, and RF, MLP, KNN, and NB with the combination of XTC and Fisher score with a rank of 2, 4, 11 and 18. Table 5 shows the prediction result of all seven models across the feature selection methods. Among the best-chosen models in each feature selection method, topsis was used to illustrate the differences between all evaluation metrics in Figure 7. Table 6 presented the best parameters sets for the best model of each feature selection method. The top 12 most contributing features were also identified and ranked based on their inclusion in the seven models in Table 7.

**#Please Insert Table 5 around here#**

**#Please Insert Figure 7 around here#**

**#Please Insert Table 6 around here#**

**#Please Insert Table 7 around here#**

The study findings suggest that ML models can be effectively employed to achieve accurate predictions of antenatal outcomes. The top 14 features identified by our models, which included factors such as the gender of the baby, maternal education and job, maternal BMI in the third trimester, and previous pregnancy history, can be useful indicators for predicting antenatal outcomes. We employed seven different ML models, including decision trees, bagging-decision tree-based, random forests, K nearest neighbor, multilayer perceptron, support vector machine, and naive bayes, and used four different feature selection methods to optimize our model including PCA, fisher score, extra tree classifier and the common features between the fisher score and XTC. The Features returned by fisher score are preeclampsia, anti-TPO, father's education, sex, miscarriage, income, BMI in the third trimester, age, mother's education, parity, and Recent miscarriage. The Features selected by the extra tree classifier are the father's education, BMI 3, anti-TPO, fetal distress, father's job, age, monthly income, sex, mother's education, miscarriage, gravid, and mother's job respectively. We combined XTC and fisher score's selected features to build a new feature selection method and attempt to boost the models' performance. The commonly selected features among these two feature selection methods are anti-TPO, sex, monthly income, mother's education, BMI3, age, father's education, and miscarriage. Most of the models performed better when common features among XTC and fisher score were used. To find the best feature selection for each model we used topsis to prioritize ML-models combination with feature selection. The best models are DT-o (accuracy = 0.92, AUC = 0.92 ), bagging-f (accuracy = 0.92, AUC = 0.93 ), RF-c (accuracy = 0.92 , AUC = 0.93 ), MLP-c (accuracy = 0.86 , AUC = 0.85 ), SVM-o (accuracy = 0.77, AUC = 0.77 ), KNN-c (accuracy = 0.77 , AUC = 0.79 ), and NB-c (accuracy = 0.54 , AUC = 0.51) respectively. Most of the results for predicting antenatal outcomes were AUC= 0.82 [54] and AUC= 0.84 [55], whereas we achieved a higher accuracy of 0.93. Our study also revealed interesting findings related to factors that can impact antenatal outcomes, such as pregnancy-related complaints, including back pain [56], and unemployment [7]. Our study also confirms the findings of a study of Iranian women that antenatal depression is associated with a preference for male children [12].

Several strengths were demonstrated in our study. To begin with, we consulted the results of a nested case-control study conducted by Iranian Hospital on pregnant women. As part of the feature selection process for building ML-based prediction models, several significant features were selected. Despite these strengths, there are also limitations. For instance, we used a relatively small dataset, thus future studies could benefit from larger datasets to achieve even better results. Antenatal depression was based on the Hamilton depression rating scale, therefore there may be information bias. We also observed that some well-known features, such as marital status[20], level of social support [29], and history of depression[7], were not included in our study which are important features. Additionally, we only employed traditional ML methods, and future research could explore the use of more complex predictors, such as artificial neural networks. In conclusion, our study adds to the existing body of research focused on utilizing ML for predicting antenatal outcomes, shedding light on the potential of these methods to enhance the well-being of both mothers and their unborn babies.

## 4. Conclusion

Antenatal depression is a significant mental health concern that can have profound effects on both pregnant women and their children. This study aimed to evaluate the efficacy of various machine learning algorithms in predicting antenatal depression using different input variables. Seven algorithms were tested, along with four feature selection methods, to determine the most precise prediction model. The results indicated that utilizing a combination of XTC and Fisher scores produced the highest accuracy across the majority of models. Particularly, the Decision Tree and Bagging classifier demonstrated superior accuracy compared to other models within the three feature selection methods.Key features for antenatal depression prediction in the third trimester were identified, including preeclampsia, fetal distress, miscarriage, recent miscarriage, gravid, anti-TPO, monthly income, fathers' education, mothers' occupation, BMI in the third trimester, menstrual history, and fathers' occupation. However, it is important to note that this study has limitations, such as being conducted in a single hospital in Tehran and having a limited sample size. These factors may impact the generalizability of the results to a larger population. Further research in diverse locations with larger sample sizes is necessary to validate the findings and enhance the accuracy of machine learning tools for predicting antenatal depression.

**References**


1. Kheirabadi, G. R., Maracy, M. R., "Perinatal depression in a cohort study on Iranian women", *Journal of Research in Medical Sciences: The Official Journal of Isfahan University of Medical Sciences*, 15(1), pp. 41 (2010).
2. Biaggi, S., Conroy, J., Pawlby, S., et al., "Identifying the women at risk of antenatal anxiety and depression: a systematic review", *Journal of Affective Disorders*, 191, pp. 62-77 (2016). DOI: 10.1016/j.jad.2015.11.014.
3. Priya, S., Chaturvedi, S., Bhasin, S. K., et al., "Depression, anxiety and stress among pregnant women: a community-based study", *Indian Journal of Psychiatry*, 60(1), pp. 151 (2018). DOI: 10.4103/psychiatry.indianjpsychiatry_230_17.
4. Zhang, L., Yang, X., Zhao, J., et al., "Prevalence of prenatal depression among pregnant women and the importance of resilience: a multi-site questionnaire-based survey in Mainland China", *Frontiers in Psychiatry*, 11, pp. 374 (2020). DOI: 10.3389/fpsyt.2020.00374.
5. Smith, J., Twynstra, J., Seabrook, J. A., "Antenatal depression and offspring health outcomes", *Obstetric Medicine*, 13(2), pp. 55-61 (2020). DOI: 10.1177/1753495x19843015.
6. Miller, E. S., Saade, G. R., Simhan, H. N., et al., "Trajectories of antenatal depression and adverse pregnancy outcomes", *American Journal of Obstetrics and Gynecology*, 226(1), pp. 108.e1-108.e9 (2022). DOI: 10.1016/j.ajog.2021.07.007.
7. Yin, X., Sun, N., Jiang, N., et al., "Prevalence and associated factors of antenatal depression: systematic reviews and meta-analyses", *Clinical Psychology Review*, 83, pp. 101932 (2021). DOI: 10.1016/j.cpr.2020.101932.
8. Norhayati, M., Hazlina, N. N., Asrenee, A., et al., "Magnitude and risk factors for postpartum symptoms: a literature review", *Journal of Affective Disorders*, 175, pp. 34-52 (2015). DOI: 10.1016/j.jad.2014.12.041.
9. Dennis, C.-L., Falah-Hassani, K., Shiri, R., "Prevalence of antenatal and postnatal anxiety: systematic review and meta-analysis", *The British Journal of Psychiatry*, 210(5), pp. 315-323 (2017). DOI: 10.1192/bjp.bp.116.187179.
10. Bunevicius, R., Kusminskas, L., Bunevicius, A., et al., "Psychosocial risk factors for depression during pregnancy", *Acta Obstetricia et Gynecologica Scandinavica*, 88(5), pp. 599-605 (2009). DOI: 10.1080/00016340902846049.
11. Marchesi, C., Bertoni, S., Maggini, C., "Major and minor depression in pregnancy", *Obstetrics & Gynecology*, 113(6), pp. 1292-1298 (2009). DOI: 10.1097/01.aoa.0000370533.89790.14.
12. Rouhi, M., Rouhi, N., Vizheh, M., et al., "Male child preference: is it a risk factor for antenatal depression among Iranian women?", *British Journal of Midwifery*, 25(9), pp. 572-578 (2017). DOI: 10.12968/bjom.2017.25.9.572.
13. Azami, M., Badfar, G., Shohani, M., et al., "The prevalence of depression in pregnant Iranian women: a systematic review and meta-analysis", *Iranian Journal of Psychiatry and Behavioral Sciences*, 12(3) (2018). DOI: 10.5812/ijpbs.9975.
14. Alderdice, F., McNeill, J., Lynn, F., "A systematic review of systematic reviews of interventions to improve maternal mental health and well-being", *Midwifery*, 29(4), pp. 389-399 (2013). DOI: 10.1016/j.midw.2012.05.010.
15. Lin, I.-H., Huang, C.-Y., Chou, S.-H., et al., "Efficacy of prenatal yoga in the treatment of depression and anxiety during pregnancy: a systematic review and meta-analysis", *International Journal of Environmental Research and Public Health*, 19(9), pp. 5368 (2022). DOI: 10.3390/ijerph19095368.

16. Schetter, C. D., Tanner, L., "Anxiety, depression and stress in pregnancy: implications for mothers, children, research, and practice", *Current Opinion in Psychiatry*, 25(2), pp. 141 (2012). DOI: 10.1097/yco.0b013e3283503680.
17. Straub, H., Adams, M., Kim, J. J., et al., "Antenatal depressive symptoms increase the likelihood of preterm birth", *American Journal of Obstetrics and Gynecology*, 207(4), pp. 329.e1-329.e4 (2012). DOI: 10.1016/j.ajog.2011.10.538.
18. Yedid Sion, M., Harlev, A., Weintraub, A. Y., et al., "Is antenatal depression associated with adverse obstetric and perinatal outcomes?", *The Journal of Maternal-Fetal & Neonatal Medicine*, 29(6), pp. 863-867 (2016). DOI: 10.3109/14767058.2015.1023708.
19. Madigan, S., Oatley, H., Racine, N., et al., "A meta-analysis of maternal prenatal depression and anxiety on child socioemotional development", *Journal of the American Academy of Child & Adolescent Psychiatry*, 57(9), pp. 645-657.e8 (2018). DOI: 10.1016/j.jaac.2018.06.012.
20. Míguez, M. C., Vázquez, M. B., "Risk factors for antenatal depression: a review", *World Journal of Psychiatry*, 11(7), pp. 325 (2021). DOI: 10.5498/wjp.v11.i7.325.
21. Martínez-Paredes, J. F., Jácome-Pérez, N., "Depression in pregnancy", *Revista Colombiana de Psiquiatría*, 48(1), pp. 58-65 (2019). DOI: 10.1016/j.rcpeng.2017.07.002.
22. Senturk, V., Abas, M., Berksun, O., et al., "Social support and antenatal depression in extended and nuclear family environments in Turkey: a cross-sectional survey", *BMC Psychiatry*, 11(1), pp. 1-10 (2011). DOI: 10.1186/1471-244x-11-48.
23. Underwood, L., Waldie, K., D'Souza, S., et al., "A review of longitudinal studies on antenatal and postnatal depression", *Archives of Women's Mental Health*, 19, pp. 711-720 (2016). DOI: 10.1007/s00737-016-0629-1.
24. Heron, J., O'Connor, T. G., Evans, J., et al., "The course of anxiety and depression through pregnancy and the postpartum in a community sample", *Journal of Affective Disorders*, 80(1), pp. 65-73 (2004). DOI: 10.1016/j.jad.2003.08.004.
25. Putnick, D. L., Bell, E. M., Ghassabian, A., et al., "Maternal antenatal depression's effects on child developmental delays: gestational age, postnatal depressive symptoms, and breastfeeding as mediators", *Journal of Affective Disorders*, 324, pp. 424-432 (2023). DOI: 10.1016/j.jad.2022.12.059.
26. Alipour, Z., Lamyian, M., Hajizadeh, E., "Anxiety and fear of childbirth as predictors of postnatal depression in nulliparous women", *Women and Birth*, 25(3), pp. e37-e43 (2012). DOI: 10.1016/j.wombi.2011.09.002.
27. Cankorur, V. S., Abas, M., Berksun, O., et al., "Social support and the incidence and persistence of depression between antenatal and postnatal examinations in Turkey: a cohort study", *BMJ Open*, 5(4), pp. e006456 (2015). DOI: 10.1136/bmjopen-2014-006456.
28. Tham, E. K., Tan, J., Chong, Y.-S., et al., "Associations between poor subjective prenatal sleep quality and postnatal depression and anxiety symptoms", *Journal of Affective Disorders*, 202, pp. 91-94 (2016). DOI: 10.1016/j.jad.2016.05.028.
29. Javed, F., Gilani, S. O., Latif, S., et al., "Predicting risk of antenatal depression and anxiety using multi-layer perceptrons and support vector machines", *Journal of Personalized Medicine*, 11(3), pp. 199 (2021). DOI: 10.3390/jpm11030199.
30. Bergner, A., Beyer, R., Klapp, B. F., et al., "Pregnancy after early pregnancy loss: a prospective study of anxiety, depressive symptomatology and coping", *Journal of Psychosomatic Obstetrics & Gynecology*, 29(2), pp. 105-113 (2008). DOI: 10.1080/01674820701687521.
31. Zeng, Y., Cui, Y., Li, J., "Prevalence and predictors of antenatal depressive symptoms among Chinese women in their third trimester: a cross-sectional survey", *BMC Psychiatry*, 15, pp. 1-7 (2015). DOI: 10.1186/s12888-015-0452-7.
32. Edwards, B., Galletly, C., Semmler-Booth, T., et al., "Antenatal psychosocial risk factors and depression among women living in socioeconomically disadvantaged suburbs in Adelaide, South Australia", *Australian and New Zealand Journal of Psychiatry*, 42(1), pp. 45-50 (2008). DOI: 10.1080/00048670701732673.
33. Alipour, Z., Kheirabadi, G. R., Eslami, A. A., et al., "Psychological profiles of risk for antenatal depression and anxiety in Iranian sociocultural context", *Journal of Education and Health Promotion*, 7, pp. 1-10 (2018). DOI: 10.4103/jehp.jehp_201_17.
34. Yasuma, N., Narita, Z., Sasaki, N., et al., "Antenatal psychological intervention for universal prevention of antenatal and postnatal depression: a systematic review and meta-analysis", *Journal of Affective Disorders*, 273, pp. 231-239 (2020). DOI: 10.1016/j.jad.2020.04.063.
35. Lewis, A. J., Galbally, M., Gannon, T., et al., "Early life programming as a target for prevention of child and adolescent mental disorders", *BMC Medicine*, 12(1), pp. 1-15 (2014). DOI: 10.1186/1741-7015-12-33.

36. Kashanian, M., Bozorgzadeh Vosta Kolaei, S., Naderi, Z., et al., "Evaluation of the relationship between serum anti-TPO antibody levels and mood disorders in euthyroid pregnant women", *The Journal of Maternal-Fetal & Neonatal Medicine*, 33(19), pp. 3233-3237 (2020). DOI: 10.1080/14767058.2019.1570120.
37. Bennett, I. M., Marcus, S. C., Palmer, S. C., et al., "Pregnancy-related discontinuation of antidepressants and depression care visits among Medicaid recipients", *Psychiatric Services*, 61(4), pp. 386-391 (2010). DOI: 10.1176/appi.ps.61.4.386.
38. Ashourloo, D., Manafifard, M., Behifar, M., et al., "Wheat yield prediction based on Sentinel-2, regression, and machine learning models in Hamedan, Iran", *Scientia Iranica*, 29(6), pp. 3230-3243 (2022). DOI: 10.24200/sci.2022.57809.5429.
39. Yao, B., Yao, B., Yao, B., et al., "Improved support vector machine regression in multi-step-ahead prediction for tunnel surrounding rock displacement", *Scientia Iranica*, 21(4), pp. 1309-1316 (2014). DOI: 10.24200/sci.2014.28200.
40. Sabahi, H., Vali, M., Shafie, D., "In-hospital mortality prediction model of heart failure patients using imbalanced registry data: a machine learning approach", *Scientia Iranica*, (2023). DOI: 10.24200/sci.2023.61637.7412.
41. Haghbayan, M., Karimi, B., Mozdgir, A., et al., "Increasing the efficacy of umbilical cord blood banking using machine learning algorithms: a case study from Royan cord blood bank", *Scientia Iranica*, (2023). DOI: 10.24200/sci.2023.61068.7132.
42. Tashakkori, R., Mozdgir, A., Karimi, A., et al., "The prediction of NICU admission and identifying influential factors in four different categories leveraging machine learning approaches", *Biomedical Signal Processing and Control*, 90, pp. 105844 (2024). DOI: 10.1016/j.bspc.2023.105844.
43. Shin, D., Lee, K. J., Adeluwa, T., et al., "Machine learning-based predictive modeling of postpartum depression", *Journal of Clinical Medicine*, 9(9), pp. 2899 (2020). DOI: 10.3390/jcm9092899.
44. Asanya, K. C., Kharrat, M., Udom, U. A., et al., "Robust Bayesian approach to logistic regression modeling in small sample size utilizing a weakly informative student’s prior distribution", *Communications in Statistics-Theory and Methods*, 52(2), pp. 283-293 (2023). DOI: 10.1080/03610926.2021.1912767.
45. Ali, M. M., Paul, B. K., Ahmed, K., et al., "Heart disease prediction using supervised machine learning algorithms: performance analysis and comparison", *Computers in Biology and Medicine*, 136, pp. 104672 (2021). DOI: 10.1016/j.compbiomed.2021.104672.
46. Casalicchio, G., Molnar, C., Bischl, B., editors, "Visualizing the feature importance for black box models", *Machine Learning and Knowledge Discovery in Databases: European Conference*, *ECML PKDD 2018*, Dublin, Ireland, September 10–14, 2018, Proceedings, Part I 18, Springer (2019). DOI: 10.1007/978-3-030-10925-7_40.
47. Huynh-Thu, V. A., Saeys, Y., Wehenkel, L., et al., "Statistical interpretation of machine learning-based feature importance scores for biomarker discovery", *Bioinformatics*, 28(13), pp. 1766–1774 (2012). DOI: 10.1093/bioinformatics/bts238.
48. Uddin, S., Haque, I., Lu, H., et al., "Comparative performance analysis of k-nearest neighbour (KNN) algorithm and its different variants for disease prediction", *Scientific Reports*, 12(1), pp. 1–11 (2022). DOI: 10.1038/s41598-022-10358-x.
49. Chicco, D., Jurman, G., "The advantages of the Matthews correlation coefficient (MCC) over F1 score and accuracy in binary classification evaluation", *BMC Genomics*, 21, pp. 1–13 (2020). DOI: 10.1186/s12864-019-6413-7.
50. Hwang, C., Lai, Y., Liu, T., "A new approach for multiple objective decision making", *Computers and Operational Research*, 20(1), pp. 1–9 (1993). DOI: 10.1016/0305-0548(93)90109-V.
51. Greene, R., Devillers, R., Luther, J. E., et al., "GIS-based multiple-criteria decision analysis", *Geography Compass*, 5(6), pp. 412–432 (2011). DOI: 10.1111/j.1749-8198.2011.00431.x.
52. Shafaei, R., Mozdgir, A., "Master surgical scheduling problem with multiple criteria and robust estimation", *Scientia Iranica*, 26(1), pp. 486–502 (2019). DOI: 10.24200/sci.2018.20416.
53. Mozdgir, A., Shafaei, R., "Case mix planning using the technique for order of preference by similarity to ideal solution and robust estimation: a case study", *International Journal of Engineering*, 30(9), pp. 1362–1371 (2017). DOI: 10.5829/idosi.ije.2017.30.09c.09.
54. Zhong, M., Van Zoest, V., Bilal, A. M., et al., "Unimodal vs. multimodal prediction of antenatal depression from smartphone-based survey data in a longitudinal study", *Proceedings of the 2022 International Conference on Multimodal Interaction*, (2022). DOI: 10.1145/3536221.3556605.

55. Payne, J. L., Osborne, L. M., Cox, O., et al., "DNA methylation biomarkers prospectively predict both antenatal and postpartum depression", *Psychiatry Research*, 285, 112711 (2020). DOI: 10.1016/j.psychres.2019.112711.
56. Wang, S., Pathak, J., Zhang, Y., "Using electronic health records and machine learning to predict postpartum depression", *MEDINFO 2019: Health and Wellbeing e-Networks for All*, IOS Press, pp. 888–892 (2019). DOI: 10.3233/SHTI190351.

## Figure and table captions

**Figure Captions:**

**Fig 1.** Methodology steps in the proposed analysis
**Fig 2.** Distribution of antenatal depression prevalence
**Fig 3.** An illustration of the feature importance for each model
**Fig 4.** Heatmap showing correlations among all the features of the dataset
**Fig 5.** A KDE plot is presented for both people with depression and non-depression based on the distribution of features
**Fig 6.** Barplots for both people with depression and non-depression
**Fig 7.** Comparison between the evaluation metrics of the best-chosen models in each feature selection method by topsis ranking

**Table Captions:**

**Table 1**. Illustration of the details among all the features in the study
**Table 2.** The number of missing values in the dataset
**Table 3.** Features' rank in selection methods
**Table 4.** Confusion Matrix
**Table 5**. Result of the seven predictors with different feature selections
**Table 6.** Hyper tunning for the best models in each feature selection method
**Table 7**. The top 12 most contributing features ranked based on their inclusion in the seven ML models

# Figures

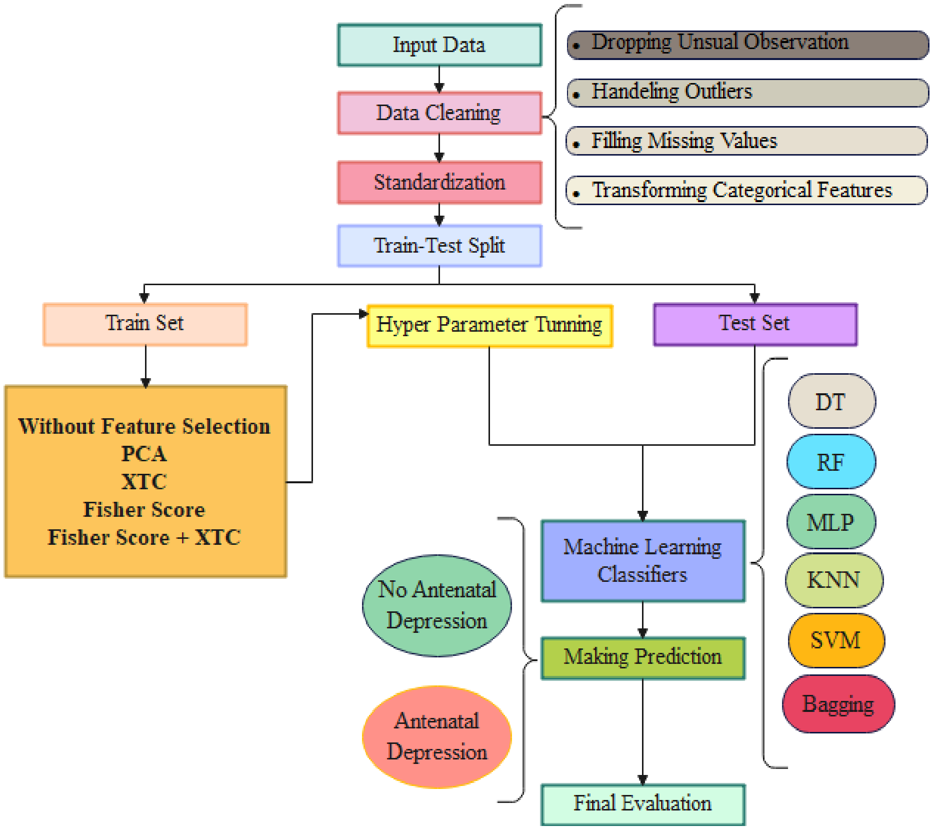



**Fig 1.** Methodology steps in the proposed analysis

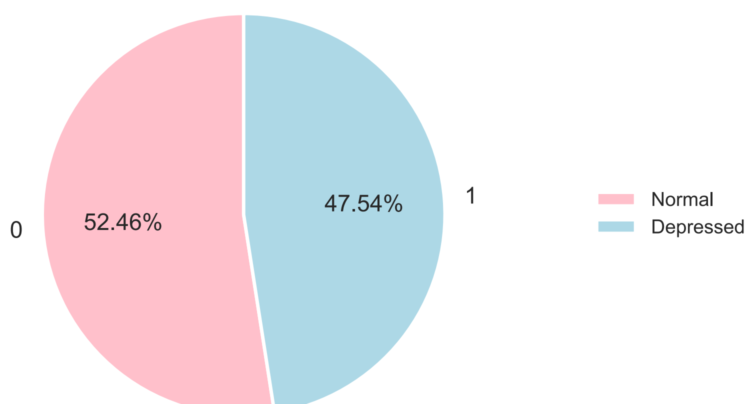
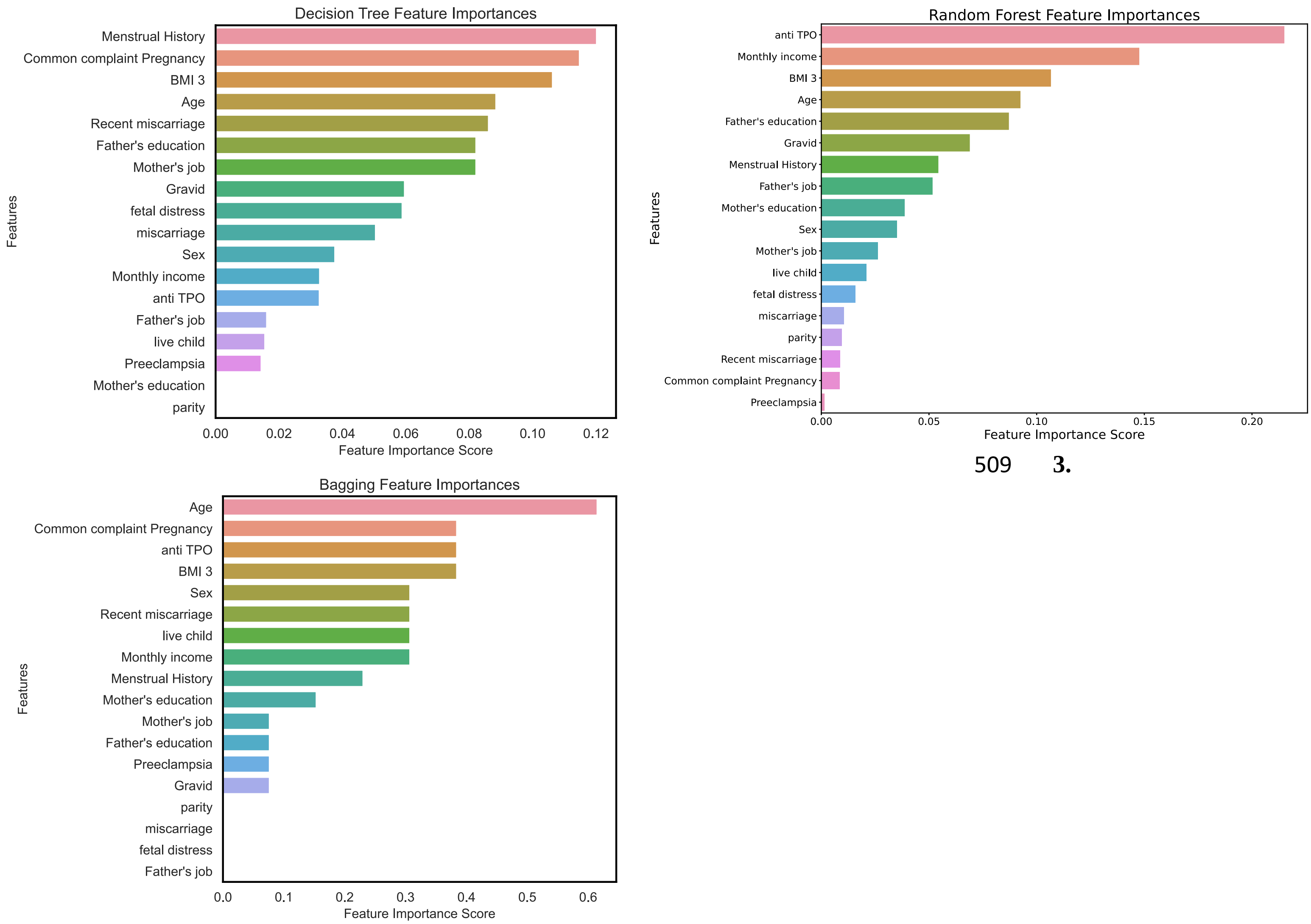



**Fig 2.** Distribution of antenatal depression prevalence




**3.**

**Fig 3.** An illustration of the feature importance for each mode

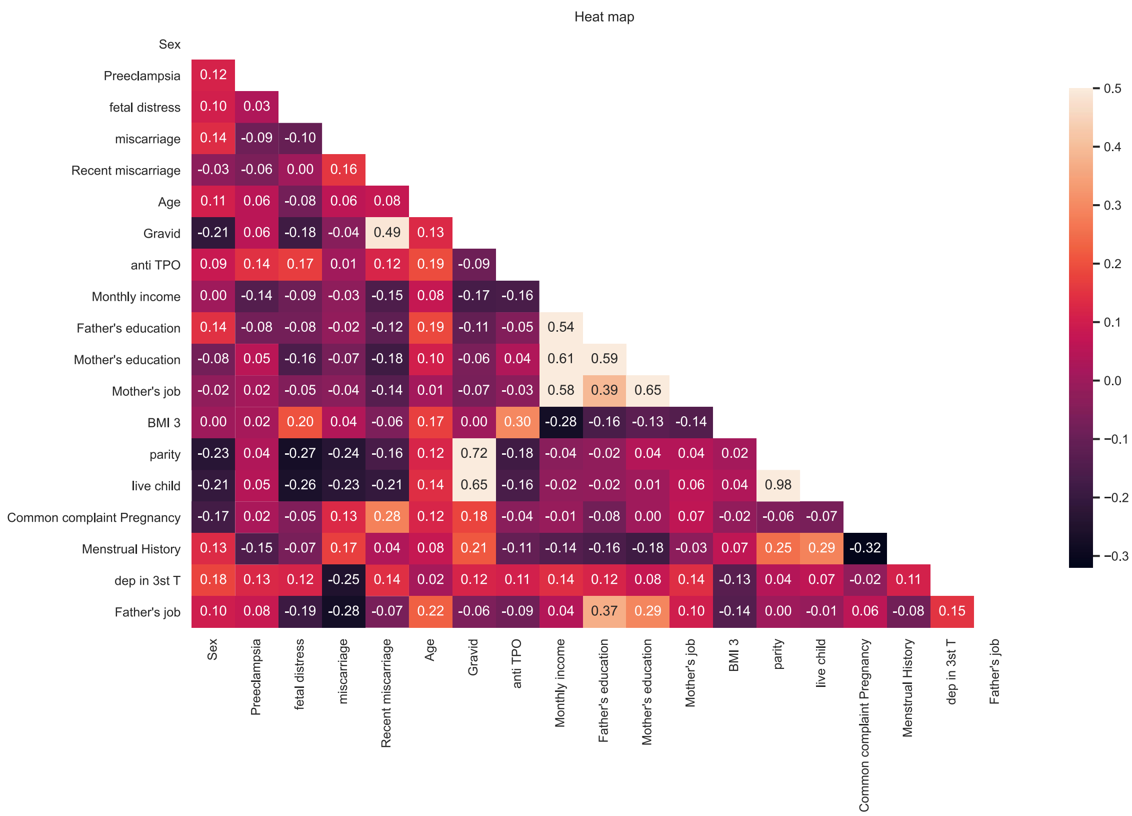



**Fig 4.** Heatmap showing correlations among all the features of the dataset

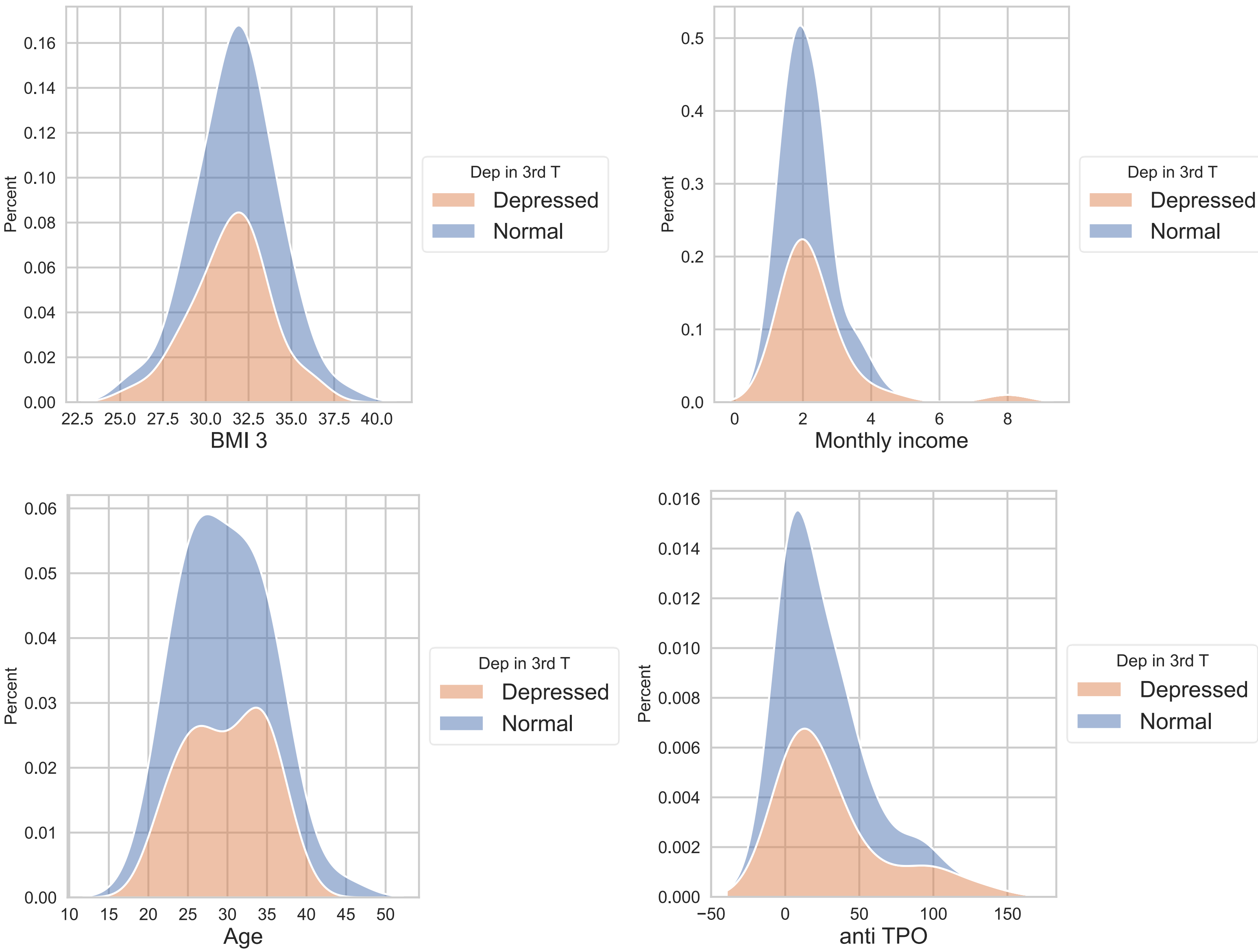



**Fig 5.** A KDE plot is presented for both people with depression and non-depression based on the distribution of features

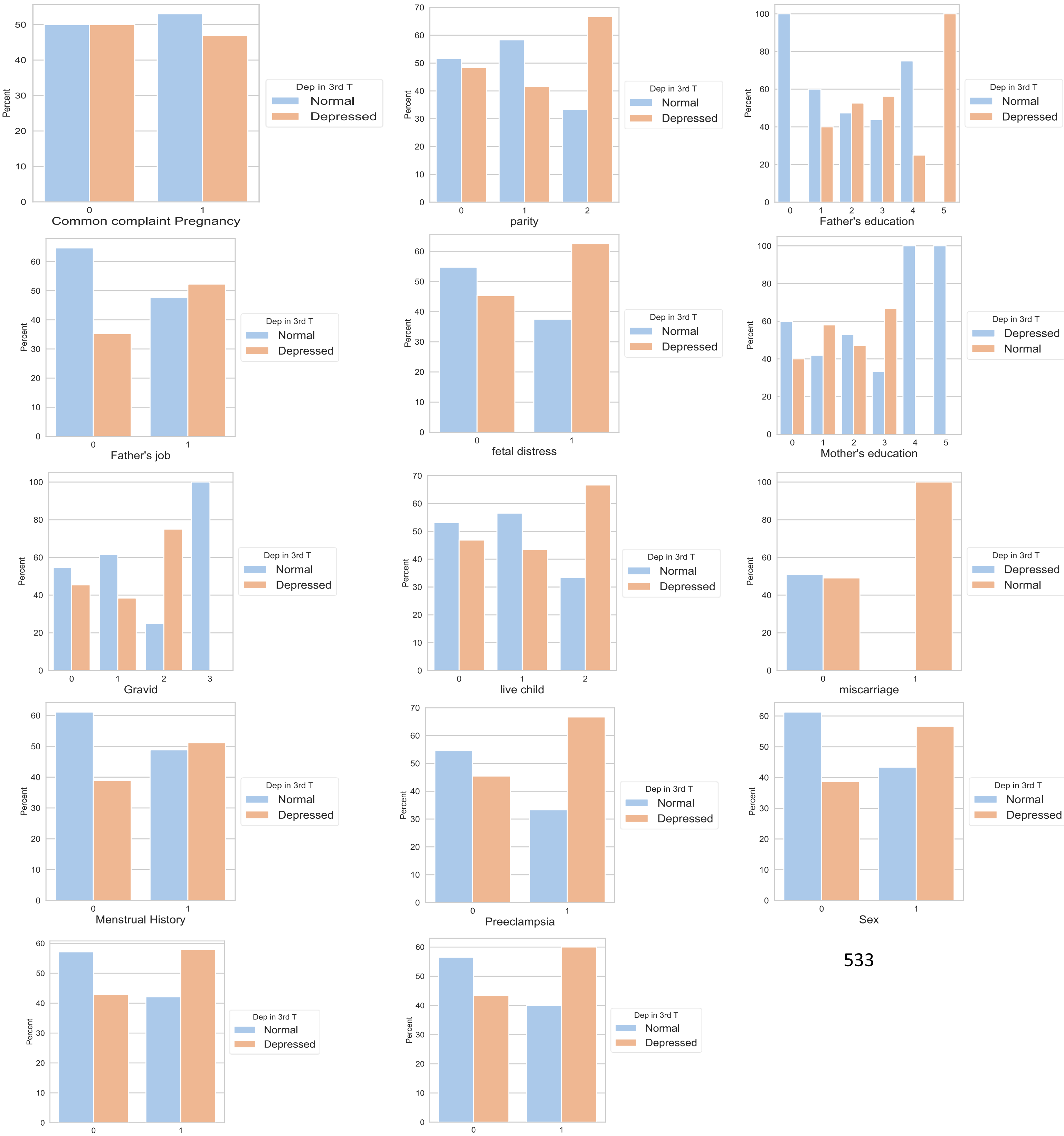



**Fig 6.** Barplots for both people with depression and non-depression

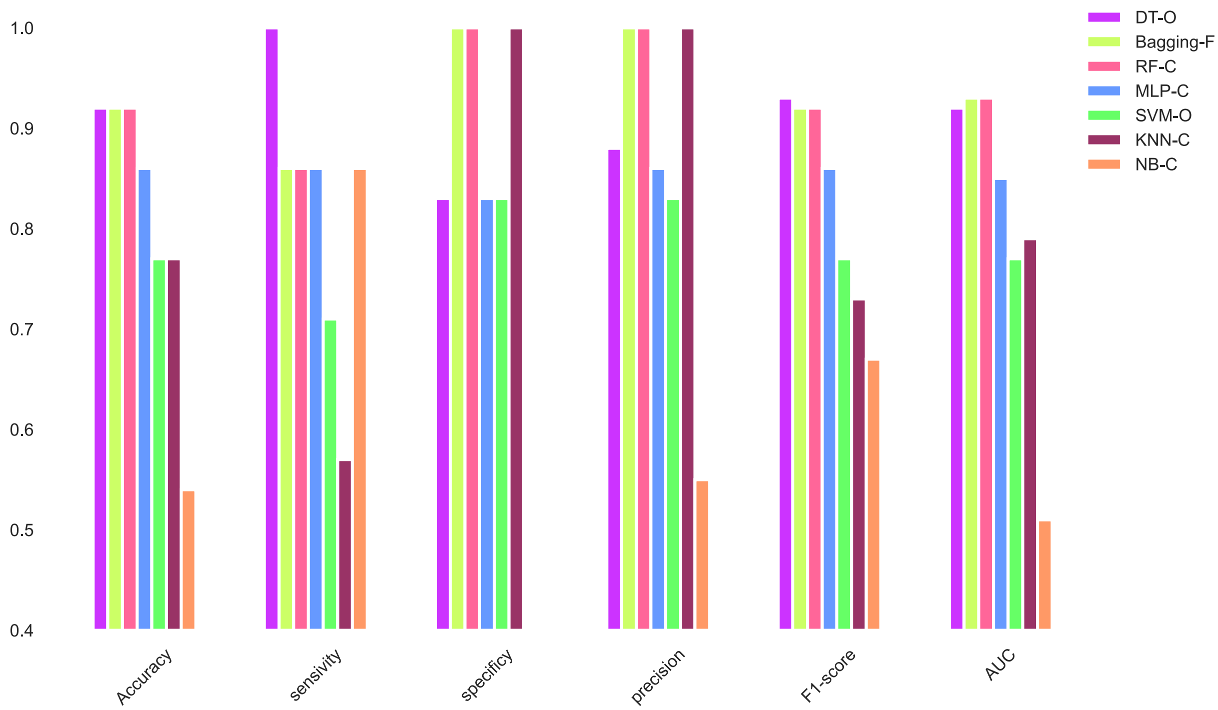



**Fig 7.** Comparison between the evaluation metrics of the best-chosen models in each feature selection method by topsis ranking

## Tables

**Table 1**. Illustration of the details among all the features in the study

| SN | Attribute name | Description | Type |
|---|---|---|---|
| 1 | Sex | Gender of the fetus, male = 1, female = 0 | Nominal |
| 2 | Preeclampsia | If exists 1, else 0. | Nominal |
| 3 | Fetal distress | If exists 1, else 0. | Nominal |
| 4 | Miscarriage | If exists 1, else 0. | Nominal |
| 5 | Recent miscarriage | If exists 1, else 0. | Nominal |
| 6 | Age | Age of the mother at the begging of this questionnaire | Numerical |
| 7 | Gravid | The number pregnancy | Numerical |
| 8 | Anti TPO | Anti-TPO is an autoantibody produced by the immune system. It causes thyroid inflammation and the higher it is, the greater the risk of miscarriage | Numerical |
| 9 | Monthly income | Parents' income | Numerical |
| 10 | Father's education | High school = 0, Diploma = 1, Associate degree = 2, Masters = 3, Bachelor = 4, P.H.D = 5 | Nominal |
| 11 | Mother's education | High school = 0, Diploma = 1, Associate degree = 2, Masters = 3, Bachelor = 4, P.H.D = 5 | Nominal |
| 12 | Mother's job | If exists 1, else 0. | Nominal |
| 13 | BMI 3 | Body mass index | Numerical |

| | | | |
|---|---|---|---|
| 14 | Parity | Pregnancies that culminated in the delivery after a gestational period of ≥ 6 months, resulting in either a live birth or a stillbirth. | Nominal |
| 15 | live child | Number of mother’s child | Numerical |
| 16 | Common complaint Pregnancy | Common experience during pregnancy include Morning sickness, Constipation, the need for pain management, Chanbes in skin, and Swelling in ankles, feet and fingers. | Nominal |
| 17 | Menstrual History | Regular 1, else 0. | Nominal |
| 18 | Father's job | Employee 1, else 0. | Nominal |
| 19 | Dep in 3st T | If the patient has antenatal depression 1, else 0. | Nominal |

**Table 2.** The number of missing values in the dataset

| Attributes | No. of missing values |
|---|---|
| Common complaint Pregnancy | 19 |
| anti TPO | 6 |
| Weight 3 | 6 |
| Weight 2 | 3 |
| Sex | 2 |
| Delivery age | 1 |
| Menstrual History | 1 |

**Table 3.** Features’ rank in selection methods

| Features | Fisher Score | XTC Score |
|---|---|---|
| Sex | 14 | 0.044298 |
| Preeclampsia | 17 | 0.027560 |
| Fetal distress | 0 | 0.082824 |
| Miscarriage | 13 | 0.039890 |
| Recent miscarriage | 6 | 0.022794 |
| Age | 9 | 0.076264 |
| Gravid | 5 | 0.039810 |
| Anti TPO | 16 | 0.086461 |
| Monthly income | 11 | 0.075798 |
| Father's education | 15 | 0.098887 |
| Mother's education | 8 | 0.043405 |
| Mother's job | 1 | 0.036786 |
| BMI 3 | 10 | 0.090735 |
| Parity | 7 | 0.024571 |
| live child | 3 | 0.025127 |
| Common complaint Pregnancy | 4 | 0.011660 |
| Menstrual History | 2 | 0.090311 |
| Father's job | - | 0.082818 |

**Table 4.** Confusion Matrix

| | | Predicted | |
|---|---|---|---|
| | | **No Antenatal Depression** | **Antenatal Depression** |
| **Actual** | **No Antenatal Depression** | TN | FP |
| | **Antenatal Depression** | FN | TP |

**Table 5**. Result of the seven predictors with different feature selections

| | Classifier Model | Accuracy | Sensitivity | Specificity | Precision | F1 Score | AUC | FPR | FNR | Topsis Rank |
|---|---|---|---|---|---|---|---|---|---|---|
| **Original Data** | *DT-o | 92% | 100% | 83% | 88% | 93% | 91.67% | 16.67% | 0% | 1 |
| | Bagging-o | 92% | 86% | 100% | 100% | 92% | 82.86% | 0% | 14.29% | 3 |
| | RF-o | 77% | 71% | 83% | 83% | 77% | 77.38% | 16.67% | 28.57% | 10 |
| | MLP-o | 85% | 71% | 100% | 100% | 83% | 85.71% | 0% | 28.57% | 7 |
| | *SVM-o | 77% | 71% | 83.33% | 83% | 77% | 77.38% | 16.67% | 28.57% | 10 |
| | KNN-o | 62% | 86% | 33% | 60% | 71% | 59.52% | 66.67% | 14.28% | 17 |
| | NB-o | 54% | 71% | 33% | 56% | 63% | 52.38% | 66.67% | 28.57% | 19 |
| **XTC + Fisher** | DT-c | 92% | 86% | 100% | 100% | 92% | 92.86 % | 0% | 14.29% | 2 |
| | Bagging-c | 85% | 100% | 66.67% | 78% | 88% | 83.33% | 33.33% | 0% | 6 |
| | *RF-c | 92% | 86% | 100% | 100% | 92% | 92.86% | 0% | 14.29% | 2 |
| | *MLP-c | 86% | 86% | 83.33% | 86% | 86% | 84.52% | 16.67% | 14.29% | 4 |
| | SVM-c | 69% | 67% | 83% | 80% | 67% | 70.24% | 16.67% | 42.86% | 14 |
| | *KNN-c | 77% | 57% | 100% | 100% | 73% | 78.57% | 0% | 42.86% | 11 |
| | *NB-c | 54% | 86% | 16.67% | 55% | 67% | 51.19% | 83.33% | 14.29% | 18 |
| **XTC** | DT-x | 86% | 86% | 83.33% | 86% | 86% | 84.52% | 16.67% | 14.29% | 4 |
| | Bagging-x | 85% | 86% | 83.33% | 86% | 86% | 84.52% | 16.67% | 14.29% | 5 |
| | RF-x | 69% | 71% | 66.67% | 71% | 71% | 69.05% | 33.33% | 28.57% | 12 |
| | MLP-x | 85% | 71% | 100% | 100% | 83% | 85.71% | 0% | 28.57% | 7 |
| | SVM-x | 62% | 43% | 83.33% | 75% | 55% | 63.10% | 16.67% | 57.14% | 16 |
| | KNN-x | 75% | 86% | 50% | 67% | 75% | 67.86% | 50% | 14.29% | 13 |
| | NB-x | 54% | 86% | 16.67% | 55% | 67% | 51.20% | 83.33% | 14.29% | 18 |
| **Fisher Score** | DT-f | 77% | 71% | 83.33% | 83% | 77% | 77.38% | 16.67% | 28.57% | 10 |
| | *Bagging-f | 92% | 86% | 100% | 100% | 92% | 92.86% | 0% | 14.29% | 2 |
| | RF-f | 85% | 71% | 100% | 100% | 83% | 85.71% | 0% | 28.57% | 7 |
| | MLP-f | 85% | 86% | 83.33% | 86% | 86% | 84.52% | 16.67% | 14.29% | 5 |
| | SVM-f | 77% | 71% | 83.33% | 83% | 77% | 77.38% | 16.67% | 28.57% | 10 |
| | KNN-f | 77% | 57% | 100% | 100% | 73% | 78.57% | 0% | 42.86% | 11 |
| | NB-f | 54% | 86% | 16.67% | 55% | 67% | 51.19% | 83.33% | 14.29% | 18 |
| **PCA** | DT-p | 92% | 86% | 100% | 1% | 92% | 92.86% | 0% | 14.29% | 8 |
| | Bagging-p | 85% | 100% | 66.67% | 78% | 88% | 83.33% | 33.33% | 0% | 6 |
| | RF-p | 77% | 86% | 66.67% | 75% | 80% | 76.19% | 33.33% | 14.29% | 9 |
| | MLP-p | 77% | 57% | 100% | 100% | 73% | 78.57% | 0% | 42.86% | 11 |
| | SVM-p | 54% | 57% | 50% | 57% | 57% | 53.57% | 50% | 42.86% | 20 |
| | KNN-p | 69% | 86% | 50% | 67% | 75% | 67.86% | 50% | 14.29% | 15 |
| | NB-p | 54% | 57% | 50% | 57% | 57% | 53.57% | 50% | 42.86% | 20 |

* The best result for each model according to its feature selection method

**Table 6.** Hyper tunning for the best models in each feature selection method

| DT-O | Value | DT-C | Value | RF-C | Value | DT-X | Value | Bagging-F | Value | Bagging-P | Value |
|---|---|---|---|---|---|---|---|---|---|---|---|
| **criterion** | entropy | **criterion** | entropy | **criterion** | entropy | **criterion** | entropy | **criterion** | entropy | **criterion** | entropy |
| **splitter** | random | **splitter** | random | **n_estimators** | 10 | **splitter** | random | **splitter** | random | **splitter** | random |
| **max_depth** | 9 | **max_depth** | 14 | **max_depth** | 8 | **max_depth** | 7 | **max_depth** | 14 | **max_depth** | 5 |
| **max_features** | sqrt | **max_features** | sqrt | **max_features** | log2 | **max_features** | auto | **max_features** | sqrt | **max_features** | sqrt |
| **random_state** | 101 | **random_state** | 1 | **random_state** | 101 | **random_state** | 356 | **random_state** | 1 | **random_state** | 1 |
| - | - | - | - | - | - | - | - | **n_estimators** | 13 | **n_estimators** | 19 |
| - | - | - | - | - | - | - | - | **random_state** | 1 | **random_state** | 1 |

**Table 7**. The top 12 most contributing features ranked based on their inclusion in the seven ML models

| Features | Frequency | DT Rank | Bagging Rank | RF Rank | MLP Rank | SVM Rank | KNN Rank | NB Rank |
|---|---|---|---|---|---|---|---|---|
| **Sex** | 7 | 11 | 6 | 10 | 10 | 3 | 4 | 10 |
| **Mother's education** | 6 | - | 10 | 9 | 3 | 6 | 9 | 4 |
| **Mother's job** | 6 | 7 | 11 | 11 | 9 | - | 6 | 4 |
| **BMI 3** | 6 | 3 | 12 | 3 | 8 | 2 | - | - |
| **Common complaint Pregnancy** | 6 | 2 | 4 | - | 5 | 8 | 2 | 8 |
| **Menstrual History** | 6 | 1 | 8 | 7 | 7 | 7 | - | 7 |
| **Father's education** | 5 | 6 | - | 5 | 1 | 10 | 10 | - |
| **Gravid** | 5 | 8 | - | 6 | 2 | - | 7 | 12 |
| **Preeclampsia** | 4 | - | - | - | 6 | 11 | 1 | 6 |
| **live child** | 4 | - | - | 12 | - | 9 | 5 | 3 |
| **Miscarriage** | 4 | 10 | 1 | - | - | - | 12 | 1 |
| **Recent miscarriage** | 4 | 5 | 5 | - | 4 | 1 | - | - |
| **Age** | 4 | 4 | 2 | 4 | - | - | 11 | - |
| **Anti TPO** | 4 | - | - | 1 | - | 5 | 3 | 11 |
| **Monthly income** | 3 | 12 | - | - | - | 12 | 8 | - |
| **Fetal distress** | 3 | 9 | 7 | - | 11 | - | - | - |
| **Parity** | 3 | - | 3 | - | - | 4 | - | 9 |
| **Father's job** | 2 | - | - | 8 | - | - | - | 2 |

## Brief technical biography of each author

## Atena Karimi

Atena Karimi is a Scientific Researcher at the Barbod AI Lab in Tehran, Iran. She holds a B.Sc. in Industrial Engineering from Kharazmi University. She is set to pursue a Master's degree in Data Science at the University of Luxembourg. Atena's research spans data science, data analysis, data-driven decision-making, data mining, decision science, machine learning, and artificial intelligence.

She has conducted advanced research on the applications of data science and machine learning in healthcare, focusing on improving patient treatments, diagnoses, and outcomes. Atena is actively involved in professional communities, including the Iranian Institute of Industrial Engineering, the

Scientific Association of Industrial Engineering at Kharazmi University, and the Barbod Artificial Intelligence Lab Journal Club.Currently, she is leading research projects aimed at leveraging data science and machine learning to enhance healthcare outcomes.

**Ashkan Mozdgir**

Ashkan Mozdgir is assistant professor, member of industrial engineering in the Faculty of Engineering, at Kharazmi University. He has a Ph.D. in Industrial Engineering (2018) from KNTU, Iran. He is author and coauthor of more than 50 papers. His research interests include data science and machine learning in healthcare, planning and scheduling on production and healthcare, supply chain design and management.

**Reza Tashakkori**

Reza Tashakkori is a Scientific Researcher at the Barbod AI Lab in Tehran, Iran. He holds a B.Sc. in Industrial Engineering from Kharazmi University. His research interests include data-driven analysis, decision science for big data, artificial intelligence applications, and analytics in healthcare.

He has conducted advanced research exploring the applications of data science and artificial intelligence in healthcare, focusing on enhancing treatments and diagnosing patients. He is a member of the Iranian Institute of Industrial Engineering and the Scientific Association of Kharazmi University. Currently, he is leading a research project exploring the applications of data science and artificial intelligence in healthcare.

**Abolfazl Bagheri Tofighi**

Abolfazl Bagheri Tofighi received a bachelor's degree in industrial engineering from the Sharif University of Technology and a master's degree in industrial engineering (specializing in healthcare systems) from Amirkabir University of Technology in Tehran, Iran. Abolfazl is a Scientific Researcher at the Barbod AI Lab in Tehran, Iran. He is passionate about using machine learning techniques and expert systems to solve critical health challenges.

**Shabnam Bozorgzadeh**

Shabnam Bozorgzadeh is a specialist in obstetrics and gynecology. she is graduated from Iran University of Medical Sciences. She has done a lot of research work in the field of the effect of various factors on women's health in different periods of life, especially during pregnancy.