The Intelligent System Review

ISSN: 3106-7840 | 3106-7832
Volume 2, Issue 3, 2025

IDENTIFICATION OF PREGNANCY LOSS RISK FACTORS USING MACHINE LEARNING ALGORITHMS

Ejaz Ahmed[1], Hussain Bux[2], Muhammad Tahir[*3]

*[1,2,*3]Assistant Professor, Department of Computer Science, Faculty of Engineering, Science and Technology (FEST), Iqra University Main Campus, Defence View, Karachi City – 75500, Sindh, Pakistan*

[1]ejazahmad372@gmail.com, [2]hussainahmad23@yahoo.com, [*3]mtahir151@gmail.com

Keywords
Pregnancy loss, Risk factors, Machine learning, KNN, SVM, Decision Tree, Extra Trees Classifier

Article History
Received: 25 July 2025
Accepted: 21 September 2025
Published: 30 September 2025

Corresponding Author: *
Muhammad Tahir

Abstract
Pregnancy loss, or spontaneous abortion, is defined as the loss of a fetus before the 20th week of gestation. According to the American College of Obstetricians and Gynecologists (ACOG), approximately 15–20% of clinically confirmed pregnancies result in pregnancy loss. This study utilized cross-sectional data from the Bureau of Statistics Punjab (BSP) to investigate the risk factors associated with pregnancy loss. Multiple machine learning algorithms—including Logistic Regression (LR), K-Nearest Neighbors (KNN), Linear Discriminant Analysis (LDA), Support Vector Machines (SVM), Naïve Bayes (NB), Regularized Naïve Classifier (RNC), Classification and Regression Trees (CART), Bernoulli Naïve Bayes (BNB), Passive Aggressive, and Extra Trees Classifier (ETC)—were applied to assess predictive performance. Among these, KNN achieved the highest accuracy at 91%, while all other algorithms exceeded 80% accuracy. Feature selection and importance analysis using LR, CART, and ETC identified the number of children ever born and place of delivery as the most influential factors affecting pregnancy loss risk.

INTRODUCTION

The word "pregnancy" evokes images of happy, gurgling babies and a "glistening pregnant woman." It's a word that conjures up images of happiness, hope for the future, unrealized dreams and relationships, and perhaps the next rung on the ladder of life—parenthood. For some, it represents the achievement of a long-held ambition. When a couple loses a pregnancy, there is no doubt that they are devastated [1]. Pregnancy loss is a condition in which a pregnancy is terminated with a negative fetal or neonatal outcome. It causes psychological and emotional distress to the couple as well as the family's close associates. The nature and severity of grief reactions vary depending on the occurrence of an unexpected event [2]. Pregnancy loss including unwanted and pre-empted pregnancies is a main public health subject that contributes to morbidity and mortality among women world- wide [3]. Approximately 4 out of 10 pregnancies are accidental, with half resulting in induced abortions. Almost 15 percent of couples miss one predictable pregnancy and 2 percent lose two. Only 0.34% risk of three or more losses. Within 12 weeks of conception, most miscarriage occurs. The cause of pregnancy loss is hard to assess. By many national organizations, early pregnancy loss is defined as a non-feasible intrauterine pregnancy throughout the first trimester up to 18 weeks from the most recent menstrual period [4]. The previous terminology has involved miscarriages and missed abortions. Early next-trimester pregnancy loss happens after 13 and before 20 weeks of pregnancy [5]. Loss of a pregnancy which is estimated by the American College of Obstetricians and Gynecologists (ACOG) is less than 20+0 weeks the pregnancy is the common procedure of pregnancy loss [6][7][8]. About one-fifth of pregnant women had complications before 5 months gestation, and 50% will end in unplanned abortion. Up to 20 percent of gestations will fail [9]. Death of newborns is considered as pregnancy loss that happens at 5 months of pregnancy and later, or failure at a weight of 0.35 kg or slightly greater, is known as a stillbirth [9]. The word “abortion” has been suggested to be altered to “spontaneous pregnancy loss” because of recognizing the emotive aspects of losing gestation [10][11][12]. Machine learning techniques can deal with structured, semi-structured, and unstructured data [13][14]. The machine learning model takes several correct outputs and learns by comparison between empirical results with the correct outputs to detect mistakes [15][16]. These models were created and used to examine medical datasets from the very beginning [17]. Several factors can impact increasing the ratio of pregnancy loss. In this study, we will analyze different factors like antenatal care (ANC), total children, tetanus injections, use of iron tablets during pregnancy, the complications of low and high blood pressure, diabetes, fever during pregnancy, postnatal care, and different complications during pregnancy that end in a result of pregnancy loss.

1 Introduction to Data

We used cross-sectional data from the Bureau of Statistics Punjab (BSP) to investigate the risk factors for pregnancy loss that UNICEF and the Bureau of Statistics Punjab collected (2017-18). This data was gathered from secondary sources. The study’s main goal is to examine various factors such as antenatal care (ANC), total children, tetanus injections, iron tablet use during pregnancy, the complications of low and high blood pressure, diabetes, fever during pregnancy, postnatal care, and various pregnancy complications that result in pregnancy loss.

2 Coding Tools

Jupyter Notebook is an open-source web application for data cleaning and transformation, data visualization, statistical modeling, and machine learning that was launched with the Anaconda software.

3 Data Preprocessing

The dataset has some missing values in specific attributes, which can lead to erroneous prediction results. The model’s accuracy can also be lowered. To overcome the missing values, we use the deletion method in which we delete each row containing the missing value.

4 Aspects of Machine Learning

After observing the data, we can be unable to understand the pattern or extract information from it. Machine learning is used in this case [18]. It may identify patterns or develop prediction models in a variety of ways, including using statistics, probability, absolute conditionality, Boolean logic, and unusual optimization strategies. Machine learning can be classified into two types: supervised learning (using classification) and unsupervised learning, which are determined by the previous data used and its availability. These are just a few of the algorithms that are commonly employed. This study explores the performance of several learning algorithms such as the Linear regression model, logistic regression Classifier, and K nearest neighbor since the classification issue is extremely popular. In the sphere of medical research, however, various methods such as linear discriminant analysis and artificial neural networks have gained prominence. As a result, we’ve decided to do feature selection using all these methods.

5 Supervised Classifiers

- K Nearest Neighbors: The simplest machine learning algorithm involves storing a dataset and predicting the training set’s closest data points as neighbors.
- Linear classification models: These models work by classifying samples based on a decision boundary and fine-tuning parameters and regularizations to achieve the best accuracy.

ISSN: 3106-7840 | 3106-7832
Volume 2, Issue 3, 2025

- Naive Bayes classification algorithms: These models learn parameters by treating each feature separately and collecting simple statistics.
- Decision Tree classifiers: Ensembles of Decision Tree Classifiers are a type of machine learning model that combines several machine learning models to create more powerful models.[19]

5.1 Logistic Regression Algorithm:

The mathematical modeling approach of logistic regression can be used to describe the relationship between several risk factors and a binary or dichotomous outcome. The logistic regression model belongs to the generalized linear regression models family, which was introduced by Needler and Wedderburn (1972) for modeling categorical data. Because linear models do not fit in this situation, these models are an extension of linear models for modeling binary response variables. [20]. Logit transformation is the most suitable for data analysis in case-control studies, owing to its ease of interpretation in terms of the log of odds of success [21]. In a binary regression model, the predicted variable has the value one, where p is the probability of success and 0 is the probability of failure. Multinomial logistic regression is the method for such multinomial predicted variables when the dependent variable has more than two classes. The logistic regression model's mathematical equation is shown below.

$P(y) = \frac{1}{e^{-(\beta_0+\beta_1 x_1+\beta_2 x_1+\cdots+\beta_k x_k)}}$ Where y is the predicted variable and x_1, x_2, . . ., x_k are predictor variables.

5.2 Classification based on K-Nearest Neighbors (KNN)

K Nearest Neighbor (KNN) is a relatively straightforward, simple-to-understand, and very efficient machine learning technique. KNN is used in a wide variety of fields, including banking, healthcare, political science, handwriting recognition, image recognition, and video recognition. Credit rating is the process by which financial institutions forecast their customers' credit worthiness. Banking institutions use risk assessment to determine if a loan is safe or dangerous. Additionally, the KNN algorithm can be utilized to solve classification and regression problems. KNN method based on similarity of features.

5.3 Linear Discriminant Analysis (LDA)

Linear Discriminant Analysis (LDA) is a classification technique that uses a dimensionality reduction technique. It reduces the number of variables in a dataset while recollecting as much information as possible. The linear discriminant analysis aims to find a linear transformation matrix where the ratio of determinants yields the desired result most discriminating features (MDF) [22]. It is based on linear transformations, which can't help with linearly non-separable data classification. Where the frequencies in classes are not equal, Linear Discriminant Analysis easily handles the problem, and their effectiveness has been proven using randomly generated test results. This model maximizes the ratio of between-class variance to within-class variance in any given data set, ensuring optimal separability. To achieve higher classification accuracy than regression models, we chose to implement an LDA algorithm.

LDA aims to improve class separation and draw a decision region between the given classes, not to change the venue. With that, we also used six other classifications of machine learning algorithms which are the decision tree classifier, Gaussian naive Bayes classifier, support vector machines, Bernoulli naive byes classifier, radius neighbor's classifier, and extra tree classifier.

6 Implementation of Algorithms

We implemented all the above-mentioned algorithms to predict pregnancy loss. For predictive analysis, we used k-fold cross-validation for splitting the data into testing and training. After that, we performed feature selection based on different models to find out the factors that matter the most to avoid pregnancy loss.

6.1 K-Fold Cross-Validation

K-Fold cross-validation is a resampling technique to divide the data into training and testing groups in such a way that every observation can be used as a training and testing set as well. During K-Fold cross-validation, the dataset is randomly shuffled. After that, the dataset is split into k groups. In every iteration, one of the groups is used as a testing set and all remaining sets are considered as the training set. K-fold cross-validation is much more reliable than the conventional method of splitting the data into 70,30 ratios. At k = 100, we performed predive analysis and evaluated the average performance of each machine learning model. After that, we compared the maximum accuracy of each model.

6.1.1 Experimental Results

After data pre-processing, the data set is categorized, and each classifier's findings are validated using a 100-fold cross-validation approach. The data set is divided into training and test set with an 8:2 ratio, meaning that 80% of the data set is randomly chosen as the training sample and the remaining 20% is the testing sample. This division is made after several different combinations have been shown to be effective. The test set is then divided randomly and evenly, and the untrained data is verified. The training set accuracy, test accuracy, and validation accuracy are performance metrics. (**Table 1**) shows the results of each classification at 100 folds cross-validation. For each classification, it was observed that. The accuracy increased or remained constant following feature selection, indicating that the classification algorithm performed better and functioned more efficiently. The highest accuracy observed is 74.56% for many classifiers. Overall, Logistic regression, Linear discriminant analysis, and Gaussian N.B. all performed similarly well, with a 74.56% accuracy.

Table 1: Accuracy of Various Algorithms

Algorithm	Mean of accuracy with 100 folds
Logistic Regression	74.56%
Linear discriminant analysis	74.56%
K-nearest neighbors	74.08%
Decision tree classifier	65.11%
Gaussian N.B.	73.56%
Support vector machines	74.53%
Bernoulli N.B.	69.86%
Passive Aggressive classifier	66.39%
Radius Neighbors classifier	71.38%
Extra tree classifier	65.18%

The accuracy increased or remained constant following feature selection, indicating that the classification algorithm performed better and functioned more efficiently. The highest accuracy observed is 74.56% for many classifiers. Overall, Logistic regression, Linear discriminant analysis, and Gaussian N.B. all performed similarly well, with a 74.56 percent accuracy.

6.2 Algorithm Comparison

In **Figure 1** we have used a boxplot to analyze our experimental data; we used a visual representation of machine learning and statistical techniques using 10 number analyses. The figure compares the performance of ten algorithms. The black circles beyond the box represent outliers that didn't obey the majority distribution principle. Each box range reflects the robustness and stability of the algorithm. The narrower the box, the more stable the model performance is. The Algorithms of LR, KNN, LDA, SVM, NB, RNC, CART, BNB, and Passive produced over 80% accuracy, with the best performance (91%) coming from the KNN algorithm. Compared to these 9 algorithms, the ETC accuracy was relatively lower, 80%.

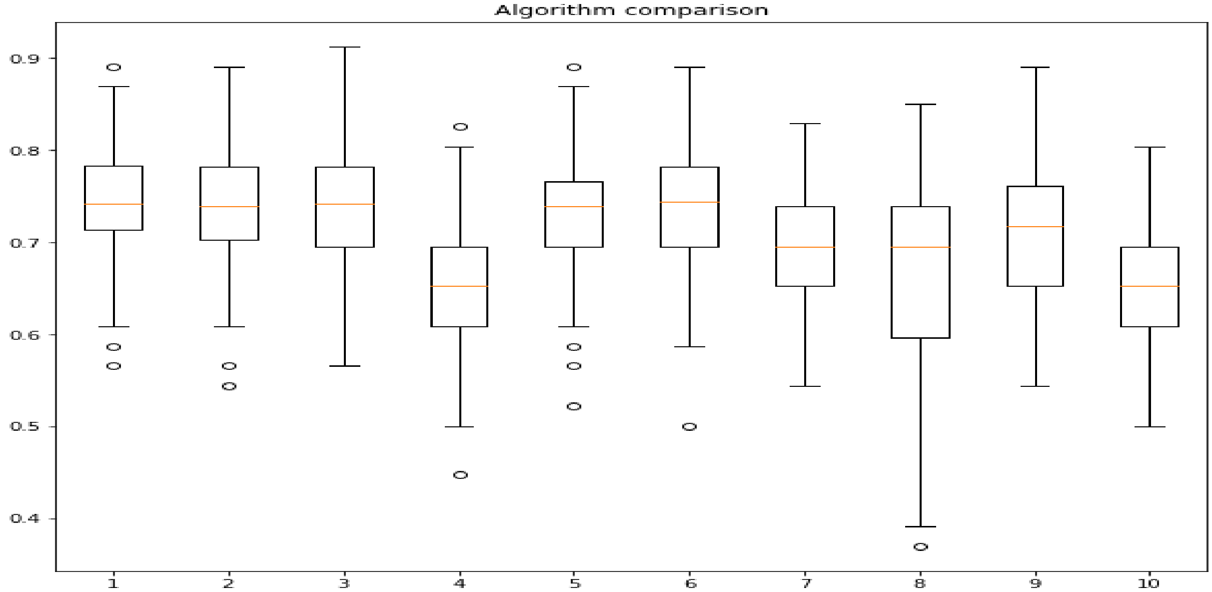

Figure 1: Algorithm Comparison

Visualizing these accuracies allows us to see the differences between them more clearly. We obtained the following accuracies after applying multiple Machine Learning Algorithms to the dataset of pregnancy loss. The highest accuracy of KNN is 91.30% for the pregnancy loss data.

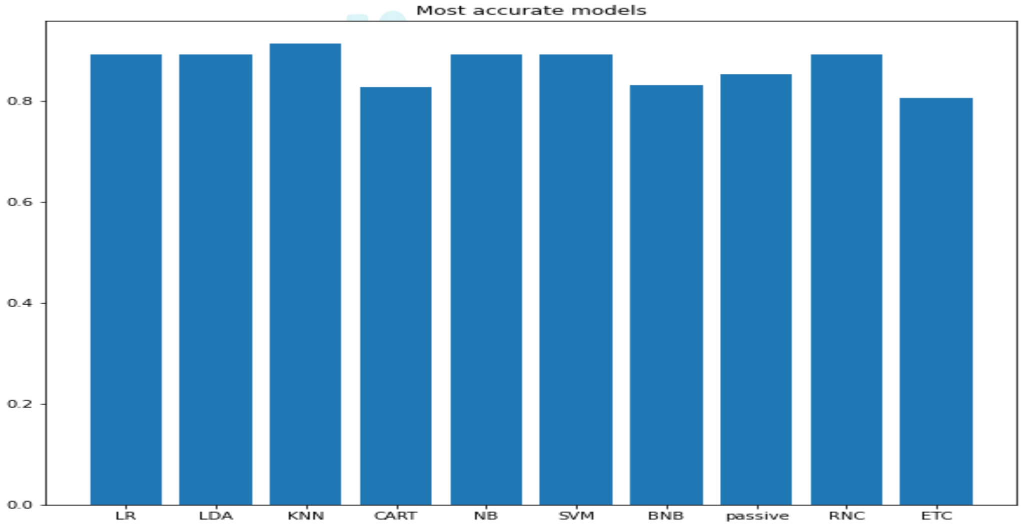

Figure 2: Most Accurate Models

Table 2: Accuracy of Different Algorithms

Algorithm	Maximum accuracy
Logistic Regression	89.13%
Linear discriminant analysis	89.13%
K-nearest neighbors	91.30%
Decision tree classifier	82.60%
Gaussian N.B.	89.13%
Support vector machines	89.13%
Bernoulli N.B.	82.97%
Passive Aggressive classifier	85.10%

The Intelligent System Review
ISSN: 3106-7840 | 3106-7832
Volume 2, Issue 3, 2025

Radius Neighbors classifier	89.13%
Extra tree classifier	80.43%

7 Feature Selection

The most significant qualities or the best collection of parameters are only extracted to maximize the model's performance, which is the main use of feature selection in machine learning. In general, the work of predictive modeling becomes more complex as the number of input factors increases. In the field of medical diagnosis, accuracy is crucial for determining the patient's disease. Experiments have shown that using Feature Selection as a preprocessing approach considerably improves classification accuracy. We also found that when any of the feature selection methods is applied, the classifier's accuracy improves dramatically when compared to the accuracy of the classifier when no feature selection is done.

7.1 Feature Selection using Logistic Regression:

Figure 3 shows the feature selection of 28 variables we identified for logistic regression on pregnancy loss data. This figure is of logistic regression that shows the direct and inverse relation of 28 variables of interest with pregnancy loss. Direct relations show that they are positively correlated with pregnancy loss and inverse shows that they are negatively correlated with pregnancy loss. This figure shows that the top 5 features that include ANC (bp check, urine sample, balanced diet), and place of delivery at home, delivery at RHC are negatively correlated it means that with these features the risk of pregnancy loss decreases in women. Similarly, the top 7 features positively correlated are diabetes, total children, obesity, ANC (ultrasound, blood sample), High blood pressure, and being hospitalized more than 24 during the last pregnancy these are

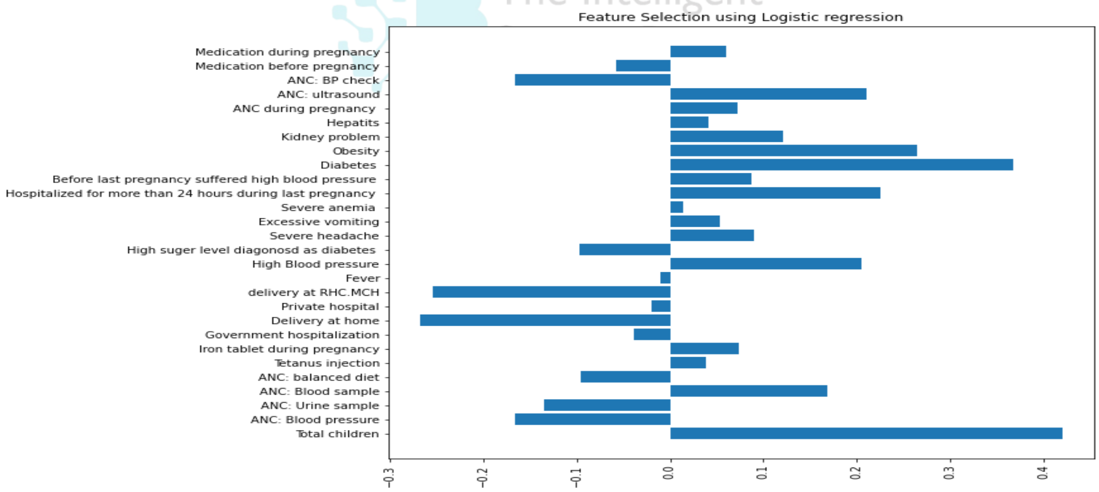

Figure 3: Feature Selection using Logistic Regression

7.2 Feature Importance using Decision Tree Classifier

Figure 4 shows the feature importance of 28 variables using a decision tree classifier. The top 2 significant variables include ANC during pregnancy and total children ever born.

The Intelligent System Review

ISSN: 3106-7840 | 3106-7832

Volume 2, Issue 3, 2025

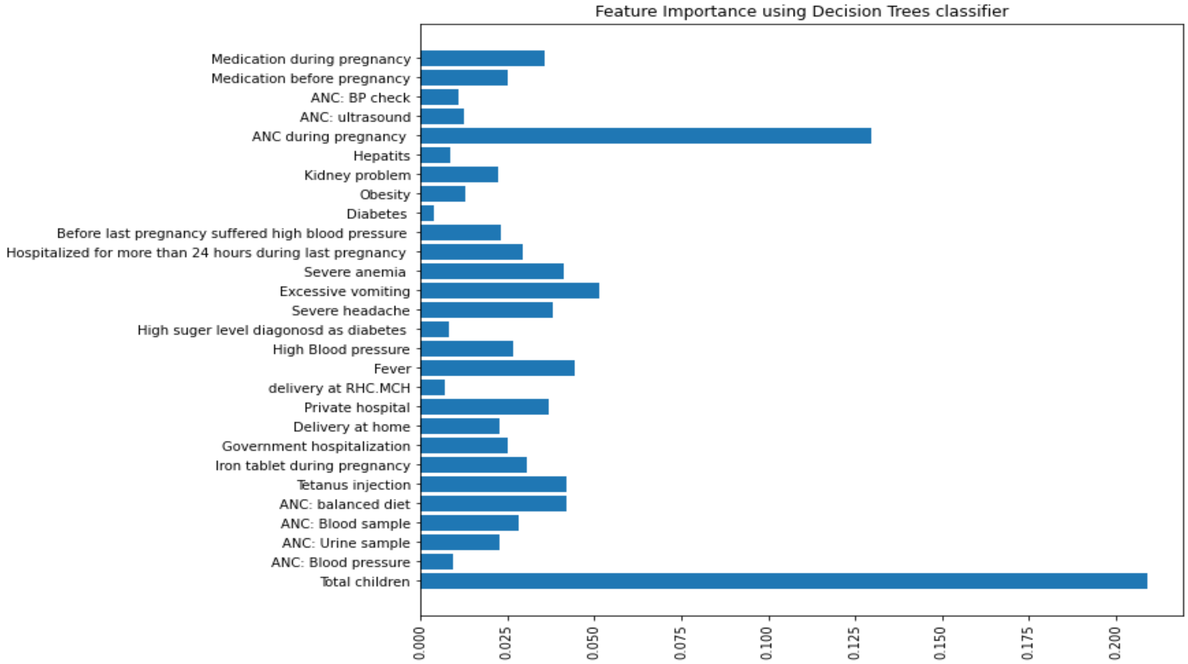

Figure 4: Feature Importance using Decision Tree Classifier

7.3 Feature Importance using Extra Trees classifier

Figure 5 shows the feature importance of 28 variables using the extra trees classifier. The top 2 significant variables include ANC during pregnancy and total children ever born. The important features that are highly increasing the risk of pregnancy loss using the extra trees classifier are the same as the decision tree classifier as ANC during pregnancy and total children ever born. so, these two features highly affect the risk of pregnancy in women.

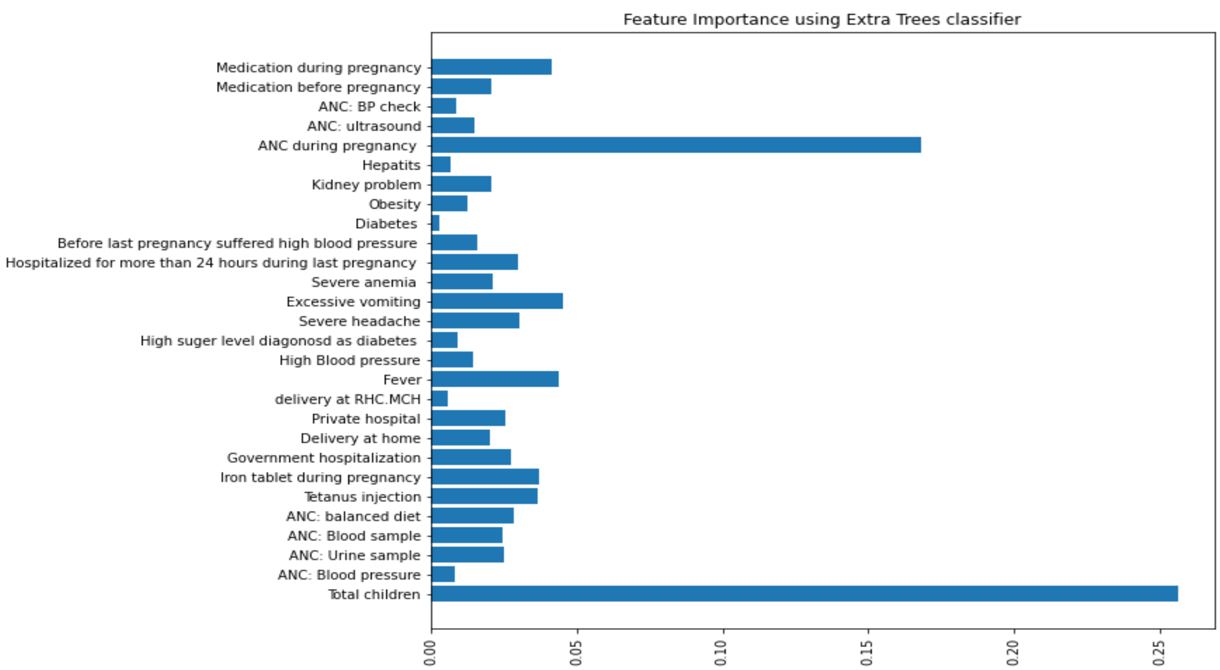

Figure 5: Feature Importance using Extra Tree Classifier

The Intelligent System Review
ISSN: 3106-7840 | 3106-7832
Volume 2, Issue 3, 2025

8 Conclusion

In this study, we have specified the features and their accuracies that must be evaluated for the risk of pregnancy loss using machine learning algorithms. In this study, using the pregnancy loss dataset we check the performance of 10 machine learning algorithms such as LR, KNN, LDA, SVM, NB, RNC, CART, BNB, Passive, ETC. we compare the accuracy result of pregnancy loss data using different machine learning algorithms Logistic Regression, KNN, LDA, SVM, NB, RNC, CART, BNB, Passive, ETC to see their performance. After a comparison of the performance of the models, we found the best accuracy of the model KNN as 91%. Algorithms of LR, KNN, LDA, SVM, NB, RNC, CART, BNB, and Passive produced over 80% accuracy. Feature selection and feature importance of 28 variables were identified using logistic regression, Decision tree classifier, and extra trees classifier the important feature that is highly affecting the risk of pregnancy is the total number of children ever born and place of delivery. In the future, we will try to suggest strategies that will increase the accuracy of pregnancy loss diagnosis. Also, to improve the outcome of this study, we will consider more data files.

9 Discussion

Machine learning algorithms in prediction of pregnancy loss remain a topic of great interest in recent years as scholars have strived to improve maternal health by the use of enhanced predictive modeling approaches. This discussion summarizes the recent findings of the studies on the potential risk factors related to pregnancy losses, especially the effectiveness of machine learning approaches on the identification of these factors. A notable example comes through work in 2025 by Pie, which explores the application of the machine learning algorithm in predicting high-risk pregnancies on expectant mothers. Applying a deeper dataset (N=15,700) in Bangladesh, the researchers applied six various algorithms (multilayer perceptron (MLP), random forest (RF), etc.) to develop predictive models. The MLP algorithm showed better results compared to overall accuracy being 82 percent and amazing 91 percent which was very high in making high-risk predictions. The presented research highlights the possibility of machine learning to contribute to the quick and accurate track of the threats to pregnancy and serve as a helpful decision-support tool to healthcare professionals [23]. Ozer conducted an- other important study in 2022, which investigated risk factors associated with first trimester pregnancy loss in good-quality frozen-thawed embryo transfer (FET) cycles [24]. The analysis of a large sample comprised total 3805 FET cycles, and key risk factors were extracted, which include maternal age and body mass index (BMI), as well as a history of recurrent pregnancy loss. The findings showed that RPL was very significantly related to higher chances of first trimester loss pregnancy, and the odds ratio (OR) was 7.729. This study explains the significance of the need to consider clinical parameters in risk prediction models since machine learning may increase the accuracy of risk estimations by dealing with complex data. Machine learning is found to be useful not only during pregnancy loss but throughout general health problems as well, according to the work of Ma and Liang (2024) [25]. An indicator of interdisciplinarity of machine learning is the reflections in their research concerning the effects of social factors on the recovery of psychiatric patients. Using a combination of analytical methods, the paper has highlighted how machine learning has the potential to facilitate complex relationships between multiple variables, and this model can be equally used in the case of pregnancy loss prediction. Moreover, A study conducted by Nuipian in 2024 examined how to classify the quality of the depression-related messages on the social media platforms with the help of machine learning algorithms.

Using a Twitter data set, the study was used to build predictive models, with accuracy rates between over 99 percent using Decision Trees and Logistic Regression. Despite their focus on mental health, the methodologies utilized in this study can direct other scholars in the creation of predictive models of pregnancy loss as data preprocessing and feature selection should receive special attention [26]. Zhang performed an entire study of the recurrent pregnancy loss

(RPL) and listed a lot of risk factors, including chromosomal abnormalities and autoimmune diseases in 2024. The review has also assessed the usefulness of the predictive categorization models, such as risk scoring systems and genetic screening software. The authors have drawn attention to the progress in machine learning algorithms that promote predictive accuracy, and it is possible that the adopted tools may substantially support the development of a personalized management approach to women with RPL [27]. Besides, Ortiz examined what affects biochemical pregnancy loss in PGT-A (preimplantation genetic testing of aneuploidy) cycles. In this study, 5,892 embryos were studied whose critical variables were uterine alterations, the day of embryo biopsy and mosaicism which were found to significantly raise the risk of biochemical pregnancy losses. Using both classically used statistical analysis and machine learning, the researchers herein explained how the mentioned techniques are useful in revealing complicated associations between numerous variables, contributing to a better comprehension of biochemical pregnancy loss [28]. To conclude, the use of machine learning algorithms in the research of pregnancy loss introduces a potentially fruitful direction that may lead to an increase in accuracy in terms of all the predictions and clinical outcomes. Among the various risk factors addressed in the reviewed literature, there are maternal age, BMI, and chromosomal abnormalities that may be efficiently analyzed on the basis of more complex machine learning methods. Along with the ongoing refinement conducted by researchers there is an increased opportunity to predict personalized interventions and more effective management approaches to the pregnant women who have faced a loss. Further research ought to dwell on the integration of the datasets and methods used, which will make the multifaceted circum- stance of pregnancy loss fully understood.

REFERENCES

[1] H. J. Janssen, M. C. Cuisinier, and K. A. Hoogduin, "A critical review of the concept of pathological grief following pregnancy loss," *OMEGA–J. Death Dying*, vol. 33, no. 1, pp. 21–42, 1996.

[2] N. Brier, "Grief following miscarriage: A comprehensive review of the literature," *J. Women's Health*, vol. 17, no. 3, pp. 451–464, 2008.

[3] R. Zakar, M. Nasrullah, M. Z. Zakar, and H. Ali, "The association of intimate partner violence with unintended pregnancy and pregnancy loss in Pakistan," *Int. J. Gynecol. Obstet.*, vol. 133, no. 1, pp. 26–31, 2016, doi: 10.1016/j.ijgo.2015.09.009.

[4] I. Malasevskaia et al., "A 21st century epidemic—obesity and its impact on pregnancy loss," *Cureus*, vol. 13, no. 1, pp. 1–7, 2021, doi: 10.7759/cureus.12417.

[5] M. C. Magnus, A. J. Wilcox, N.-H. Morken, C. R. Weinberg, and S. E. Håberg, "Role of maternal age and pregnancy history in risk of miscarriage: Prospective register-based study," *BMJ*, vol. 364, 2019.

[6] C. Alves and A. Rapp, "Spontaneous abortion (miscarriage)," *StatPearls* [Internet], 2020.

[7] S. Prager, E. Micks, and V. K. Dalton, "Pregnancy loss (miscarriage): Risk factors, etiology, clinical manifestations, and diagnostic evaluation," *UpToDate*, 2018.

[8] T. I. M. Cundy et al., "Differing causes of pregnancy loss in type 1 and type 2 diabetes," *Diabetes Care*, vol. 30, no. 10, pp. 2603–2607, 2007, doi: 10.2337/dc07-0555.

[9] S. G. Gabbe, "A story of two miracles: The impact of the discovery of insulin on pregnancy in women with diabetes mellitus," *Obstet. Gynecol.*, vol. 79, no. 2, pp. 295–299, 1992.

ISSN: 3106-7840 | 3106-7832
Volume 2, Issue 3, 2025

[10] E. V. Nosova, G. O'Malley, E. Dassau, and C. J. Levy, "Leveraging technology for the treatment of type 1 diabetes in pregnancy: A review of past, current, and future therapeutic tools," *J. Diabetes*, vol. 12, no. 10, pp. 714–732, 2020.

[11] R. Nambiar, R. Bhardwaj, A. Sethi, and R. Vargheese, "A look at challenges and opportunities of big data analytics in healthcare," in *Proc. IEEE Int. Conf. Big Data*, 2013, pp. 17–22.

[12] B. Schultz, "Big data in big companies," *Baylor Bus. Rev.*, vol. 32, no. 1, pp. 20–21, 2013.

[13] R. Bhardwaj, A. R. Nambiar, and D. Dutta, "A study of machine learning in healthcare," in *Proc. IEEE 41st Annu. Comput. Softw. Appl. Conf.* (COMPSAC), vol. 2, 2017, pp. 236–241.

[14] K. R. Dalal, "Analysing the implementation of machine learning in healthcare," in *Proc. Int. Conf. Electron. Sustain. Commun. Syst. (ICESC)*, 2020, pp. 133–137.

[15] I. Kononenko, "Machine learning for medical diagnosis: History, state of the art and perspective," *Artif. Intell. Med.*, vol. 23, no. 1, pp. 89–109, 2001, doi: 10.1016/S0933-3657(01)00077-X.

[16] J. D. Murray, *Mathematical Biology I: An Introduction*. New York, NY, USA: Springer-Verlag, 2001.

[17] P. Pongsumpun, "Transmission model for dengue disease with and without the effect of the extrinsic incubation period," *KMITL Sci. Tech. J.*, vol. 6, pp. 74–82, 2006.

[18] A. Dey, "Machine learning algorithms: A review," *Int. J. Comput. Sci. Inf. Technol.*, vol. 7, no. 3, pp. 1174–1179, 2016.

[19] A. P. Bhuvaneswari, R. P. Sam, and C. S. Bindu, "EPMD: Effective prediction model for diseases by reducing the dimensions and applying clustering with different machine learning algorithms," 2016.

[20] G. Molenberghs, G. Verbeke, C. G. B. Demétrio, and A. M. C. Vieira, "A family of generalized linear models for repeated measures with normal and conjugate random effects," *Stat. Sci.*, vol. 25, no. 3, pp. 325–347, 2010.

[21] R. B. Patel et al., "Demographic and clinical predictors of mortality from highly pathogenic avian influenza A (H5N1) virus infection," *PLoS One*, vol. 9, no. 3, p. e91630, 2014.

[22] M. Strzelecki, P. Szczypinski, A. Materka, and A. Klepaczko, "A software tool for automatic classification and segmentation of 2D/3D medical images," *Nucl. Instrum. Methods Phys. Res. A*, vol. 702, pp. 137–140, 2013.

[23] X. Pi, J. Wang, L. Chu, G. Zhang, and W. Zhang, "Prediction of high-risk pregnancy based on machine learning algorithms," *Sci. Rep.*, vol. 15, no. 1, p. 15561, 2025.

[24] G. Ozer, A. Akca, B. Yuksel, I. Duzguner, A. C. Pehlivanli, and S. Kahraman, "Prediction of risk factors for first trimester pregnancy loss using machine learning algorithms," *J. Assist. Reprod. Genet.*, vol. 40, no. 2, pp. 279–288, 2023.

[25] H. Ma and L. Liang, "Machine learning-based impact analysis of social factors in psychiatric patients," in *Proc. 4th Int. Conf. Big Data, Artif. Intell. Risk Manag.*, Nov. 2024, pp. 308–313.

[26] V. Nuipian, S. Hanumas, and K. Plangklang, "Constructing depression prediction model using machine learning algorithms," in *Proc. 8th Int. Conf. Nat. Lang. Process. Inf. Retrieval*, Dec. 2024, pp. 413–418.

[27] X. Zhang, J. Gao, L. Yang, X. Feng, and X. Yuan, "Recurrent pregnancy loss: Risk factors and predictive modeling approaches," *J. Matern.-Fetal Neonatal Med.*, vol. 38, no. 1, p. 2440043, 2025.

[28] J. A. Ortiz et al., "Factors affecting biochemical pregnancy: Machine learning-assisted identification," *Hum. Reprod.*, vol. 37, suppl. 1, p. deac104-025, 2022

.