



An Automated System for Early Prediction of Miscarriage in the First Trimester Using Machine Learning

Sumayh S. Aljameel¹, Malak Aljabri^{1,2}, Nida Aslam¹, Dorieh M. Alomari^{3,*}, Arwa Alyahya¹, Shaykhah Alfaris¹, Maha Balharith¹, Hiessa Abahussain¹, Dana Boujlea¹ and Eman S. Alsulmi⁴

¹Department of Computer Science, College of Computer Science and Information Technology, Imam Abdulrahman Bin Faisal University, P.O. Box 1982, Dammam, 31441, Saudi Arabia

²Computer Science Department, College of Computer and Information Systems, Umm Al-Qura University, Makkah, 21955, Saudi Arabia

³Department of Computer Engineering, College of Computer Science and Information Technology, Imam Abdulrahman Bin Faisal University, P.O. Box 1982, Dammam, 31441, Saudi Arabia

⁴Department of Obstetrics and Gynecology, College of Medicine, Imam Abdulrahman Bin Faisal University, Dammam, Saudi Arabia

*Corresponding Author: Dorieh M. Alomari. Email: 2180007089@iau.edu.sa

Received: 01 September 2022; Accepted: 08 December 2022

Abstract: Currently, the risk factors of pregnancy loss are increasing and are considered a major challenge because they vary between cases. The early prediction of miscarriage can help pregnant ladies to take the needed care and avoid any danger. Therefore, an intelligent automated solution must be developed to predict the risk factors for pregnancy loss at an early stage to assist with accurate and effective diagnosis. Machine learning (ML)-based decision support systems are increasingly used in the healthcare sector and have achieved notable performance and objectiveness in disease prediction and prognosis. Thus, we developed a model to help obstetricians predict the probability of miscarriage using ML. And support their decisions and expectations about pregnancy status by providing an easy, automated way to predict miscarriage at early stages using ML tools and techniques. Although many published papers proposed similar models, none of them used Saudi clinical data. Our proposed solution used ML classification algorithms to build a miscarriage prediction model. Four classifiers were used in this study: decision tree (DT), random forest (RF), k-nearest neighbor (KNN), and gradient boosting (GB). Accuracy, Precision, Recall, F1-score, and receiver operating characteristic area under the curve (ROC-AUC) were used to evaluate the proposed model. The results showed that GB overperformed the other classifiers with an accuracy of 93.4% and ROC-AUC of 97%. This proposed model can assist in the early identification of at-risk pregnant women to avoid miscarriage in the first trimester and will improve the healthcare sector in Saudi Arabia.

Keywords: Miscarriage; pregnancy; abortion; machine learning; gradient boosting



This work is licensed under a Creative Commons Attribution 4.0 International License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

1 Introduction

Pregnancy loss is a common phenomenon; eight out of ten pregnancy losses occur within the first 20 weeks of the pregnancy (the first trimester), which is known as miscarriage [1]. Pregnancy loss has a great negative impact on the number of new births in the world and on the physical and emotional health of women. Furthermore, it can be a serious problem if no action was taken early. Doctors face some difficulties in predicting miscarriage at the early stages and this prevents them from taking the proper actions to avoid its risk. While the early prediction of miscarriage will help doctors and patients to take the needed care and protect the embryo from being lost [1]. Pregnancy loss occurs for multiple reasons that can be related to both physical and psychological health, making it difficult for doctors to identify the leading cause. According to related works and studies, several indicators can cause miscarriage. These predictors include sociodemographic factors, such as age (becoming pregnant at different ages may affect pregnancy status); occupation (some occupations may require more physical exertion than others); and factors related to a pregnant woman's existing health conditions, such as high or low body mass index (BMI) (obese or underweight), high blood pressure, diabetes, cancer, infertility, and other diseases or conditions. In addition, women can experience health conditions that occur only during pregnancy, such as gestational diabetes, preeclampsia, and eclampsia [2,3]. Numerous studies have investigated pregnancy loss using artificial intelligence (AI) and ML to predict the risk factors of miscarriage by collecting data from pregnant women or women who had experienced pregnancy loss [4]. This study developed ML prediction models to help obstetricians make accurate and timely decisions to avoid the risk of miscarriage. The proposed model in this study will be a great enhancement of the healthcare field and will help doctors to take the needed process to reduce miscarriage possibility. Furthermore, this study examined the impact of routine screening blood tests, including the levels of different types of white blood cells (WBCs), red blood cells (RBCs), and platelets, as well as sociodemographic features, including the age, weight, and height. To achieve the research objectives, we obtained clinical data from King Fahad University Hospital, Khobar, Kingdom of Saudi Arabia (KSA). The data contained the records of patients who had miscarriages or normal deliveries. The proposed model can be used as a tool that can identify early signs of at-risk pregnancy.

The main contributions of the study are as follows:

- To the authors' knowledge, this study is the first to predict miscarriage using a dataset from Saudi Arabia.
- The proposed model effectively predicted miscarriage at an early stage with a reduced number of features.
- The model is a fully automated solution for predicting miscarriage.
- The dataset is balanced.

The study is organized as follows. Section 2 discusses the previous studies related to the problem addressed in the current study. Section 3 describes the material and methods used in this study. Section 4 describes the experimental setup and reports the results. Finally, Section 5 concludes the paper.

2 Related Work

Mora-Sánchez et al. [4] proposed an ML-based prediction model showing the relationship between recurrent miscarriage and the human leukocyte antigen (HLA) genes according to predicted values. Using a support vector machine (SVM) with a linear kernel as a classifier, the study detected the risk of recurrent miscarriage with an accuracy of 67% and an AUC of 71%. Similarly, Bruno et al. [5] also used an SVM model to classify the risk level of patients with recurrent pregnancy loss (RPL).

Using 43 features, the model obtained a balanced accuracy of $90.24\% \pm 0.36\%$. Furthermore, the researchers developed a model using 18 features that obtained a balanced accuracy of $93.85\% \pm 0.34\%$. However, the dataset contained many missing values for several features, such as age, BMI, activated protein C resistance (APCR), proteins C and S, antithrombin III (AT III), homocysteinemia, and thyroid stimulating hormone (TSH). These values were missing either because their assessment was not prescribed or because the values were within the normal range and thus not included.

Jhee et al. [6] constructed models using ML classifiers to predict late-onset preeclampsia and compared these models with statistical methods. They used six classifiers: logistic regression (LR), DT, naïve Bayes (NB), SVM, RF, and stochastic gradient boosting (SGB). The SGB model had the best performance, with an accuracy of 97.3%, a sensitivity of 60.3%, and a specificity of 99.1%. In addition, the prediction of late-onset preeclampsia using the ML algorithms surpassed that of the statistical methods. However, the study had limitations related to a lack of first-trimester data for most women who participated in the study because they started the antenatal exam after the early second trimester. Most previous studies have shown that women who develop preeclampsia in the second and third trimesters have significant maternal changes, although some reports have reported some changes in the first trimester. A major study limitation was that the number of patients with preeclampsia incidents was smaller than the number of patients in the control group; however, considering the study sample size, the number of patients with preeclampsia was suitable. Moreover, the sample size included in the study was larger than that of previous studies demonstrating the relationship between clinical biomarkers and the development of preeclampsia. An additional limitation concerned the antenatal evaluation; because of differences in patients' symptoms and conditions, the evaluation intervals were diverse.

Another study by Alptekin et al. [7] developed a prediction model for miscarriage based on first-trimester ultrasound findings and maternal characteristics for women with viable single pregnancies. Two models were created using DT to determine genetic abnormalities that could lead to miscarriage. The DT model achieved a sensitivity of 75%, a specificity of 93%, and an AUC of 0.87 ± 0.02 . However, this study examined only embryonic miscarriage (gestational age: 6–9 weeks) and fetal miscarriage (gestational age: 10–20 weeks), excluding preclinical or subclinical miscarriages (at or before a gestational age of 6 weeks).

Miyagi et al. [8] proposed a prediction model for the probability of live birth using ML classifiers based on blastocyst images. Six ML methods were used: LR, NB, nearest neighbors, RF, neural network, and SVM. They used 80 images of blastocysts that led to living birth and 80 images that led to aneuploid miscarriages with fivefold cross-validation for classifying embryos. The study concluded that LR was the best classifier, with an AUC of 0.650 ± 0.04 , a sensitivity of 60%, a specificity of 70%, and an accuracy of 65%.

Another study by Malacova et al. [9] developed a prediction model to support clinical decision-making by easily quantifying stillbirth risk using LR, RF, extreme gradient boosting (XGBoost), and regression tree classifiers. XGBoost outperformed the other classifiers, predicting 45% (95% CI: 43%, 46%) of stillbirths; furthermore, this model predicted 45% (95% CI: 43%, 47%) of stillbirths when pregnancy history was included. The authors noted the limitation of using perinatal records spanning more than three decades; the database changed over time, with more detailed predictor tools becoming available later in the study period. Similarly, Koivu et al. [10] studied risk factors that could be utilized in a clinical setting. The ML classifiers used were LR, deep neural network, and gradient-boosting decision tree. The study used ML models as tools to generate risk prediction models and show the power of improved clinical prediction models. The models were used to predict both early and late

stillbirth. For early stillbirth, both the LR model and the deep neural network model achieved an AUC of approximately 73% to 74%; for late stillbirth, the LR model achieved an AUC of 58% to 61%, and the deep neural network achieved an AUC of 54% to 57%. A limitation of this study was that the data contained observations from multiple years, regions, and hospitals.

Liu et al. [11] developed a prediction model for embryonic development using six ML algorithms: LR, SVM, DT, back-propagation neural network (BNN), XGBoost, and RF. This model was developed to help doctors make more accurate decisions in clinical practice. The RF algorithm produced the best result; it had an accuracy of 97% when it included the fetal heart rate (FHR) feature and 99% when it included the embryo transfer (ETD) feature. Asri et al. [12] created a prediction model with the Apache SPARK Databricks platform and ML through the K-means predictive model algorithm. The experimental results showed that the K-means algorithm performed well, clustering the data into three meaningful clusters. The algorithm predicted miscarriage in 44% of the sample would have a miscarriage, no miscarriage in 21% of the sample, and probable miscarriage in 34% of the sample. A limitation of this study was that better results could be obtained with a higher value of K because of the reduced squared error. However, a higher value of K may not have produced useful and meaningful clusters. Therefore, it was difficult to choose the correct number of clusters, and a result with 100% accuracy was rare because real data are complex. Another study by Mu et al. [13] detected and classified the adverse outcomes of pregnancy before the participants became pregnant. Deep learning (DL) algorithms were applied using a multi-layer neural network (MLP) and DT. The researchers' model provided an accuracy of 89.2%, outperforming two other models, a five-layer NN (85.9% accuracy) and a DT model (79.5% accuracy). Table 1. summarizes the reviewed papers.

Table 1: Summary of the related studies

Reference	Objective	Technique	Dataset	Sample location	Sample Size	Gestational age	Results
[4]	Analyzed HLA haplotypes from couples with either history of successful pregnancy or RM	SVM, linear kernel	Genetic data	Poland	190 samples	–	Accuracy of 0.67, AUC of 0.71
[6]	Classified the level of risk of patients with RPL in different risk classes	SVM	Clinical data with 43 medical features	Italy	734 samples	–	Accuracy of 93.85%
[5]	Developed models to predict late-onset preeclampsia and compared them to statistical methods	LR, DT, NB, SVM, GB, SGB	Clinical data collected from hospital electronic medical records	Korea	11,006 samples	34 weeks	Accuracy of 97.3%
[7]	Predicted miscarriage in women with a viable single pregnancy from first trimester	DT	Ultrasound features and clinical data	Turkey	415 samples	6–10 weeks	AUC of 0.87 ± 0.02
[8]	Predicted the probability of achieving live birth	LR, RF, DT, NB, KNN, ANN	Blastocyst image	Japan	160 images	–	Accuracy of 65%
[9]	Developed a prediction model to easily quantify stillbirth risks	DT, RF, GB	Clinical data with various medical features of the sample	Australia	947,025 livebirths and 5,788 stillbirths	≥20 weeks	Predicted 45% (95% CI: 43%, 46%) of stillbirths

(Continued)

Table 1: Continued

Reference	Objective	Technique	Dataset	Sample location	Sample Size	Gestational age	Results
[10]	Identified unfamiliar risk models that could be utilized in a clinical setting	LR, DNN, GB	Clinical data containing various medical features of the tested sample	United States	364,124 samples		AUC of 0.74
[11]	Provided decision support for doctors who are relatively inexperienced in clinical practice	LR, RF, DNN, DT, SVM	Historical case data with six medical features	China	31,030 samples	First trimester (6–12 weeks)	Accuracy of 99%
[12]	Predicted real-time miscarriage	K-means	Medical data gathered from pregnant women via mobile phone		100,000 samples		Error rate of 10%
[13]	Detected and classified adverse outcomes of pregnancy before becoming pregnant	DNN, DT	Clinical data containing various medical features of the tested sample	China	75,542 samples	–	Accuracy of 89.2%

While another study proposed a software-based solution to enhance the healthcare sector. Bhatti et al. [14] developed an android application for healthcare strengthening that improves the process of communication and collecting data between hospitals and reduces the consumption of time and effort of management staff. To collect the needed data for the proposed system, the authors developed a questionnaire to interview the staff included in using the system. The staff was consisting of both the software engineers who are responsible for the input and processing of the data and the monitors who are responsible for the resulting reports. After collecting the data, it was sent to the centralized server for analysis. The healthcare staff was very satisfied with the proposed application as it helped save time and effort. The system had been used in Pakistan in 24 districts and proved its success in all these areas. However, the proposed system is very recent and needs more time to be generalized across the entire country.

3 Material and Methods

This section contains the dataset description and the methodology used to build the proposed model. To train the model, a Saudi clinical dataset was collected and cleaned. This will help in reducing the miscarriage percentage in Saudi Arabia. Several ML classifiers were used. These classifiers were selected according to the findings of the literature review, as the best classifiers were selected to experiment with their performance on our dataset. The evaluation metrics were used to compare the performance of the proposed models. Fig. 1. summarizes the proposed methodology.

3.1 Data Collection

The study was performed using retrospective data from King Fahad University Hospital, Khobar, KSA. The study was approved with institutional review board (IRB) no. UGS-2021-09-057. The dataset contained 23 clinical attributes of 981 pregnant women with normal delivery or miscarriage. All the attributes in the dataset were numeric except for the final attribute. The numbers of cases of normal delivery (529) and miscarriage (425) were almost balanced, as shown in Fig. 2. The data were collected during the first trimester. The average age of the pregnant women with normal delivery was

31.5 years, whereas the average age of those with miscarriages was 32.7 years. The weights of women with normal delivery and miscarriage outcomes were similar (72 kilogram).

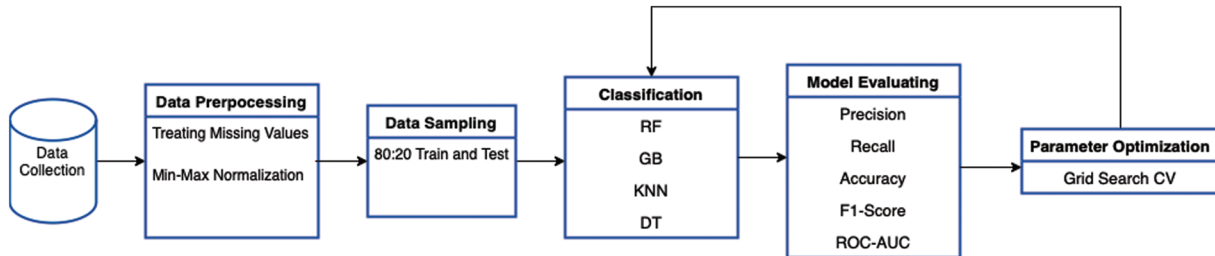


Figure 1: Methodology diagram

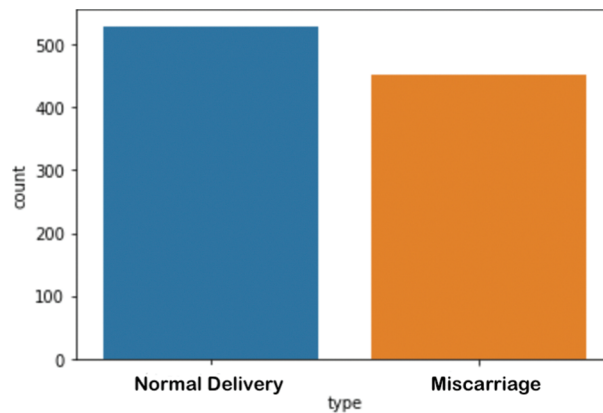


Figure 2: Number of samples per category in the dataset

3.2 Data Preprocessing

Data preprocessing was applied to prepare the dataset for the training of the proposed model. This included cleaning the data to remove noise and handle missing values. Initially, the dataset was normalized. Normalization is a method of scaling numeric data to a specific range [15]. Because the features of the current dataset were all numeric features, but their ranges were very wide, min–max normalization techniques were applied to all features of the dataset so that their ranges were all between 0 and 1. The formula for min–max normalization is represented in Eq. (1).

$$X = \frac{X - X_{min}}{X_{max} - X_{min}} \quad (1)$$

Subsequently, the normalization encoding was applied to the class attributes. After initial preprocessing, the correlations between the features and the attributes were assessed. The correlation of each feature with the target class was key for demonstrating the effect of a feature on the class attribute; some features played important roles in determining the class value, whereas other features had very small effects on the class value. Moreover, features could have positive or negative impacts. Table 2. shows the attribute descriptions and the correlation between each feature and the target class.

Table 2: Attribute description and correlation with the target

#	Attribute category	Attribute	Description	Correlation
1	Demographic	Age	Patient's age	0.086723
2		Height	Patient's height	0.218346
3		Weight	Patient's weight	0.033858
4	Hematological	White blood cells	White blood cell count	-0.176410
5		Red blood cells	Red blood cell count	-0.058829
6		Hemoglobin	Hemoglobin count	0.066638
7		Hematocrit	Hematocrit level	-0.028201
8		MCV	Mean corpuscular volume level	0.028968
9		MCH	Mean corpuscular hemoglobin level	0.083064
10		MCHC	Mean corpuscular hemoglobin concentration	0.315363
11		RDW	Red cell distribution width	-0.137509
12		MPV	Mean platelet volume	-0.200866
13		Platelets	Platelet level	0.234775
14		Eosinophil—Instrument%	Eosinophils—Instrument percentage	0.254486
15		Eosinophil—Instrument Abso.	Eosinophils—Instrument absolute value	0.292272
16		Basophil—Instrument%	Basophils—Instrument percentage	0.384456
17		Basophil—Instrument Abso.	Basophils—Instrument absolute value	-0.276348
18		Lymphocyte—Instrument%	Lymphocytes—Instrument percentage	0.142342
19		Lymphocyte—Instrument Abso.	Lymphocytes—Instrument absolute value	-0.062244
20		Monocyte—Instrument%	Monocytes—Instrument percentage	-0.186689
21		Monocyte—Instrument Abso.	Monocytes—Instrument absolute value	-0.233695
22		Neutrophil granulocyte—Instrument%	Neutrophil granulocytes—Instrument percentage	-0.130288
23		Neutrophil granulocyte—Instrument Abso.	Neutrophil granulocytes—Instrument absolute values	-0.222571

As shown in Table 2. above, most of the features were weakly correlated with the target class. The highest correlation was observed between the basophil percentage and the class value (0.384456).

3.3 Classification Models

Several classification models were applied, namely RF, KNN, GB, and DT. The section below describes the classifiers used in this study.

3.3.1 K-Nearest Neighbors

KNN is a frequently used supervised ML algorithm; it is a nonparametric method that is used for both classification and regression problems. The algorithm determines the new observation classification by calculating the distance metric between the test instance and all the training instances, after which it selects KNN instances [15]. Various distance measure functions can be used to calculate the distance; the most widely used functions are the Euclidean distance, illustrated in Eq. (2); the Manhattan distance, represented in Eq. (3); and the Minkowski metric, shown in Eq. (4):

$$\sqrt{\sum_{i=1}^k (x_i - y_i)^2} \quad (2)$$

$$\sum_{i=1}^k |x_i - y_i| \quad (3)$$

$$\left(\sum_{i=1}^k (|x_i - y_i|^q) \right)^{\frac{1}{q}} \quad (4)$$

where x is the test sample and y is the training sample.

3.3.2 Decision Tree

DT is a supervised ML algorithm that can be used to solve regression and classification problems. It consists of a hierarchical tree structure that is similar to flow charts [15]. The most frequently used method is implementing a top-down greedy search to calculate the entropy see Eq. (5) and information gain (IG) see Eq. (6) for each class:

$$E(T, X) = \sum_{c \in X} P(c) E(c) \quad (5)$$

$$\text{Gain}(T, X) = \text{Entropy}(T) - \text{Entropy}(T, X) \quad (6)$$

3.3.3 Random Forest

RF is a supervised ML algorithm that is applicable in classification and regression problems [15]. The technique is a tree-based ensemble that consists of multiple DTs; each tree represents random variables. The Gini index Eq. (7) is used to determine the branch and spread of nodes on a DT, and Eq. (8) (entropy) is used to select the branches of the nodes as it computes the probability of a certain outcome:

$$\text{Gini} = 1 - \sum_{i=1}^c (P_i)^2 \quad (7)$$

$$\text{Entropy} = \sum_{i=1}^c -p_i * \log_2(p_i) \quad (8)$$

where p_i is the probability of class i .

3.3.4 Gradient Boosting

The gradient boosting (GB) algorithm is an ML algorithm that uses ensembled methods, such as boosting, to build a model with high performance; it can be used for classification and regression problems. The algorithm is based on combining multiple trees with low accuracy rates to create an enhanced classifier with better performance and higher accuracy results than the original trees [15].

3.4 Evaluation Metrics

The evaluation metrics used to evaluate the proposed model were accuracy, precision, recall, F1 score, and ROC-AUC.

A confusion matrix was used to measure the performance of the model by comparing the predicted values with the true values [15,16]; the abbreviations TP, TN, FP, and FN indicate true positive, true negative, false positive, and false negative, respectively.

The accuracy Eq. (9) represents the percentage of the correctly classified instances among all the instances in the testing set:

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \quad (9)$$

The precision Eq. (10) represents the percentage of the correctly classified instances in the positive class among all the positive classifications:

$$Precision = \frac{TP}{TP + FP} \quad (10)$$

The recall Eq. (11) (also known as sensitivity) represents the percentage of the positive instances that were correctly classified among all the true-positive instances:

$$Recall = \frac{TP}{TP + FN} \quad (11)$$

The F score Eq. (12) represents the average of the correctly classified instances in the positive class (precision) and the positive instances that were correctly classified (recall). It is used to evaluate the balance of a model's predictions between the two classes:

$$F\ score = \frac{2 \times Precision \times Recall}{Precision + Recall} \quad (12)$$

The ROC AUC shows the performance of the model at all the classification thresholds. It plots two parameters, TPR and FPR, at different thresholds.

4 Experiments and Results

The experiment was carried out using Python 3.6 on the Jupyter notebook platform. The dataset contained two classes with a total of 981 cases; 80% of cases in each class were used for training, and 20% of cases in each class were used for testing. The target prediction had two values: 0 (normal delivery) and 1 (miscarriage). Parameter optimization was used to select the best hyperparameters for each model to obtain optimal performance. In this study, the grid search cross-validation (CV) algorithm was used to identify the optimal parameters for each model. Grid search CV uses all the possible combinations of the available parameters to determine the optimal combination of parameters. Table 3. shows optimal values for each model after applying grid search (CV), and Table 4. presents the results of the proposed models without parameter optimization. Table 5. provides the results of the classifiers after parameter optimization.

Table 3: Optimal parameters of the classifiers after grid search optimization

Classifier	Parameter	Optimal value
RF	Criterion	Entropy
	Max depth	10
	Number estimators	100

(Continued)

Table 3: Continued

Classifier	Parameter	Optimal value
KNN	K-value	3
GB	Loss	Exponential
	Max depth	10
	Max features	Log ₂
	Number estimators	15
DT	Criterion	Gini
	Max depth	15

Table 4: Classifier performance before parameter optimization

Classifier	Precision	Recall	F1 Score	Accuracy	ROC AUC
RF	92%	91%	91%	91.37%	98%
KNN	81%	81%	81%	80.71%	89%
GB	93%	93%	93%	92.89%	98%
DT	89%	89%	89%	88.83%	88%

As shown in [Table 4](#). above, the GB classifier resulted in the best performance, with an accuracy of 92.89%. While the KNN classifier resulted in the lowest performance. After applying grid search CV, the performance of most models improved, as shown in [Table 5](#). and [Fig. 3](#).

Table 5: Classifier performance after parameter optimization

Classifier	Precision	Recall	F1 Score	Accuracy	ROC AUC
RF	92%	92%	92%	91.87%	97%
KNN	82%	82%	82%	81.73%	87%
GB	94%	93%	93%	93.40%	97%
DT	89%	89%	89%	88.83%	88%

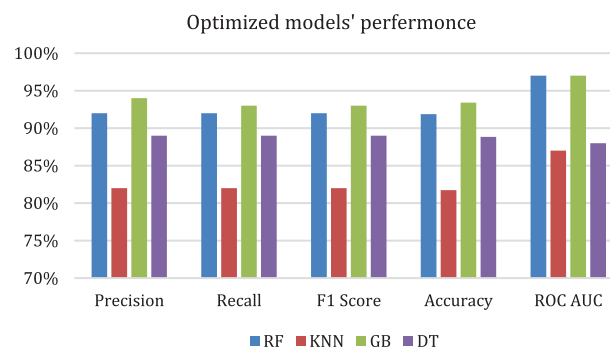
**Figure 3:** Optimized models' performance

Table 5. shows that the performance improved in all models except for the DT model, in which the performance was the same. The GB model resulted in the best performance, with an accuracy of 93.40% and a precision of 94%. Although all the evaluation metrics improved after the optimization, the ROC AUC decreased in all the models; it reached 98% before optimization and 97% after optimization in both the RF and GB models. These experiments demonstrate that the ensemble models outperformed the single classifiers; the RF and GB models resulted in the best performance among all the models. While the KNN model resulted in the lowest performance with an accuracy level of 82%. Even though the DT model resulted in an accuracy level of 89% which is higher than KNN, they both resulted in similar ROC-AUC of 88% and 87% respectively. Fig. 4. and Fig. 5. show the confusion matrices for the RF and GB models.

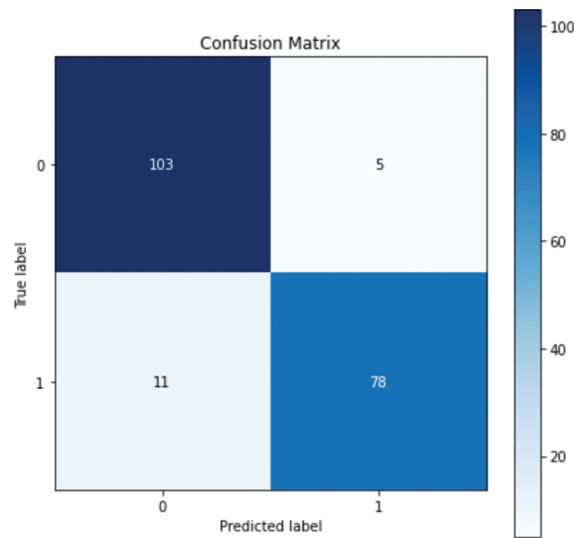


Figure 4: Confusion matrix for RF

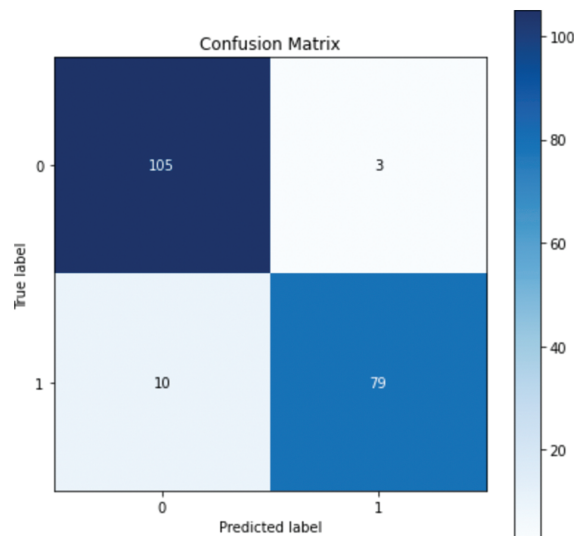


Figure 5: Confusion matrix for GB

The confusion matrix shows that RF misclassified 16 instances, whereas GB misclassified 13 instances; this explains the difference between the accuracy of the two models. For a more detailed comparison, Figs. 6. and 7. show the ROC AUC of the RF and GB models.

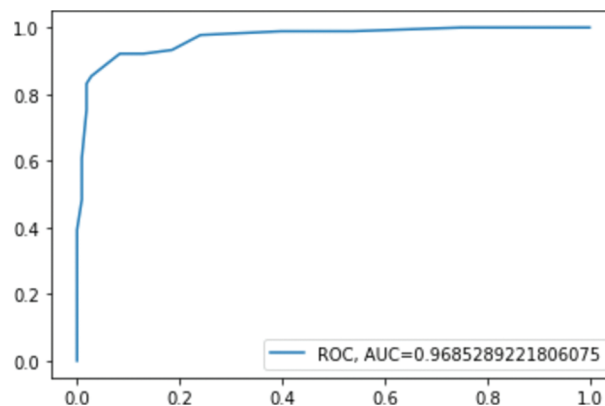


Figure 6: ROC AUC for RF

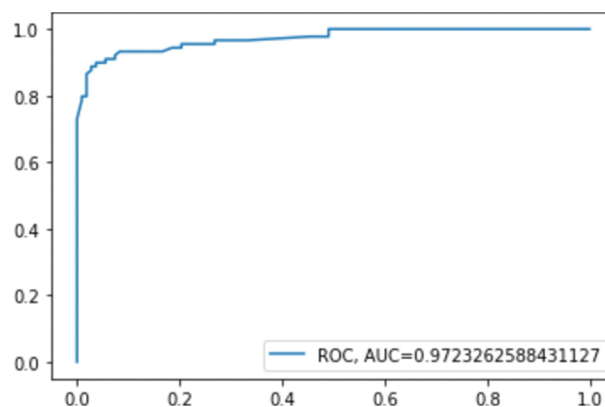


Figure 7: ROC AUC for GB

The proposed model resulted in high performance on the new Saudi clinical dataset. Although the data is new and was collected from a hospital, the quality of the data was high, and the data was balanced. Moreover, the model required a small number of features to predict miscarriage. This will save the time and effort of patients and doctors. The proposed model is fully automated and can be generalized and used in other Saudi hospitals to reduce the percentage of miscarriages in Saudi Arabia.

5 Conclusion

The risk of pregnancy loss is constantly increasing, and many pregnant women are facing a serious risk. Therefore, it is necessary to determine the predictors of these complications, as they could help obstetricians estimate the risk level at the early stages of pregnancy and lower the risk of miscarriage. In this study, several ML algorithms, namely RF, DT, KNN, and GB, were used to predict miscarriage at early stages to improve the healthcare sector and reduce the risk for pregnant ladies. The study used a real dataset, which contained the data of 984 patients at King Fahad University Hospital, Khobar, KSA, including 23 clinical attributes. The results showed that the GB classifier outperformed the other

models with an accuracy of 93.40% and a ROC-AUC of 97%. To the best of the authors' knowledge, this is the first study to use ML and a real dataset from KSA for the early prediction of miscarriage. Collecting the data and preparing it to be used as the main limitation of this work. Although the results achieved by the current study are significant, further studies should be conducted to build on these findings. Specifically, the proposed model must be validated using a large, multicenter dataset. For future work, a larger dataset from multiple Saudi hospitals will be collected and combined to build more accurate models. Also, more advanced techniques such as DL will be used to develop enhanced models.

Acknowledgement: We would like to acknowledge King Fahd Hospital for providing the dataset needed to perform our experiments and build the proposed model.

Funding Statement: The authors received no specific funding for this study.

Conflicts of Interest: The authors declare that they have no conflicts of interest to report regarding the present study.

References

- [1] S. Prager, E. Micks and V. K. Dalton, "Pregnancy loss (miscarriage): Terminology, risk factors, and etiology," 2021. [Online]. Available: <https://www.medilib.ir/uptodate/show/5439>.
- [2] A. Aslam, S. Perera, M. Watts, D. Kaye, J. Layland *et al.*, "Previous pre-eclampsia, gestational diabetes and hypertension place women at high cardiovascular risk: But do we ask?," *Heart Lung and Circulation*, vol. 30, no. 1, pp. 154–157, 2021.
- [3] C. B. Schmidt, I. Voorhorst, V. H. W. Gaar, A. Keukens, B. J. P. Loon *et al.*, "Diabetes distress is associated with adverse pregnancy outcomes in women with gestational diabetes: A prospective cohort study," *BMC Pregnancy Childbirth*, vol. 19, no. 1, pp. 1–9, 2019.
- [4] A. Mora-Sánchez, D.-I. Aguilar-Salvador and I. Nowak, "Towards a gamete matching platform: Using immunogenetics and artificial intelligence to predict recurrent miscarriage," *NPJ Digital Medicine*, vol. 2, no. 1, pp. 1–6, 2019.
- [5] V. Bruno, M. D'Orazio, C. Ticconi, P. Abundo, S. Riccio *et al.*, "Machine learning (ML) based-method applied in recurrent pregnancy loss (RPL) patients diagnostic work-up: A potential innovation in common clinical practice," *Scientific Reports*, vol. 10, no. 1, pp. 1–12, 2020.
- [6] J. H. Jhee, S. H. Lee, Y. Park, S. E. Lee, Y. A. Kim *et al.*, "Prediction model development of late-onset preeclampsia using machine learning-based methods," *PLoS One*, vol. 14, no. 8, pp. 1–12, 2019.
- [7] H. Alptekin, T. Acar, H. Işık and T. Cengiz, "Ultrasound prediction of spontaneous abortions in live embryos in the first trimester," *Electronic Journal of General Medicine*, vol. 13, no. 4, pp. 86–90, 2016.
- [8] Y. Miyagi, T. Habara, R. Hirata and N. Hayashi, "Feasibility of artificial intelligence for predicting live birth without aneuploidy from a blastocyst image," *Reproductive Medicine and Biology*, vol. 18, no. 2, pp. 204–211, 2019.
- [9] E. Malacova, S. Tippaya, H. D. Bailey, K. Chai, B. M. Farrant *et al.*, "Stillbirth risk prediction using machine learning for a large cohort of births from Western Australia, 1980-2015," *Scientific Reports*, vol. 10, no. 1, pp. 1–8, 2020.
- [10] A. Koivu and M. Sairanen, "Predicting risk of stillbirth and preterm pregnancies with machine learning," *Health Information Science and Systems*, vol. 8, no. 1, pp. 1–12, 2020.
- [11] L. Liu, Y. Jiao, X. Li, Y. Ouyang and D. Shi, "Machine learning algorithms to predict early pregnancy loss after in vitro fertilization-embryo transfer with fetal heart rate as a strong predictor," *Computer Methods and Programs Biomedicine*, vol. 196, no. 3, pp. 1–4, 2020.
- [12] H. Asri, H. Mousannif and H. Al Moatassime, "Big data analytics in healthcare: Case study-miscarriage prediction," *International Journal of Distributed Systems and Technologies*, vol. 10, no. 4, pp. 45–58, 2020.

- [13] Y. Mu, K. Feng, Y. Yang and J. Wang, "Applying deep learning for adverse pregnancy outcome detection with pre-pregnancy health data," *101 Eurotherm Seminar-Transport Phenomena in Multiphase Systems*, vol. 189, no. 1, pp. 1–6, 2018.
- [14] U. A. Bhatti, M. Huang, Y. Zhang and W. Feng, "Research on the smartphone based ehealth systems for strengthening healthcare organization," *International Conference on Smart Health*, vol. 10219 LNCS, no. 1, pp. 91–101, 2017.
- [15] J. Han, J. Pei and M. Kamber, "Data mining: Concepts and techniques." in: *The Morgan Kaufmann Series in Data Management Systems, Data Mining*, 3rd ed., vol. 1. Burlington, Massachusetts, United States: Science Direct, pp. 83–124, 2012.
- [16] A. Tharwat, "Classification assessment methods," *Emerald Insights*, vol. 17, no. 1, pp. 168–192, 2021.