International Journal of Advanced Research and Interdisciplinary Scientific Endeavours, Vol. 3(3), 2025
DOI: 10.61359/11.2206-2549

MaternityAI: Smart Pregnancy Risk Prediction System

Journal Contributors *

Email Correspondence*: editor@ijarise.org

[1] International Journal of Advanced Research and Interdisciplinary Scientific Endeavours.

Abstract:

Early pregnancy risk detection is crucial for maternal healthcare in order to avoid difficulties and this practice supports better health for expecting mothers together with their newborn. Conventional risk assessment techniques mostly rely on clinical knowledge, which can be laborious and subjective. To tackle these challenges, this research presents MaternityAI, a machine learning-based predictive system that categorizes pregnancy risks according to important health factors like age, blood pressure, and blood sugar levels. The suggested methodology analyzes patient data using sophisticated classification algorithms to produce precise and trustworthy risk predictions. Furthermore, feature importance analysis pinpoints important variables that lead to pregnancy difficulties, giving medical practitioners important information. MaternityAI improves maternal risk assessment by incorporating data-driven decision-making, which permits early interventions, lowers unfavorable outcomes, and improves mother care tactics. In the future, the model will be improved using deep learning methods and made available as a web application for clinical usage in real time.

Keywords: Maternal Healthcare, Pregnancy Risk Prediction, Machine Learning, Risk Assessment Classification Algorithms, Feature Importance Analysis, Early Intervention.

1. Introduction

Globally, pregnancy-related problems continue to be a major cause of death for both mothers and infants. Reducing unfavourable outcomes requires early detection of high-risk pregnancies, but present clinical approaches frequently lack objectivity, scalability, and efficiency. Conventional risk assessments depend on healthcare professionals performing manual evaluations, which can be subject to subjective bias, human error, and decision-making delays. Modern data-driven techniques into pregnancy risk assessment became achievable because healthcare data increased along with machine learning (ML) developments. Various medical programs including disease forecasting and patient surveillance alongside tailored therapeutic strategies have demonstrated effective implementation of machine learning technology. Large-scale dataset analysis allows ML algorithms to spot trends that traditional clinical evaluations might miss at first glance. A well-trained prediction model can offer real-time risk classification in the context of maternal healthcare, enabling early intervention and to improve the health of the expecting mothers. Although machine learning applications in healthcare have advanced, there are still difficulties in creating models that are both interpretable and generalizable to various patient groups. Due to insufficient or biased training datasets, many current models have poor generalization, which reduces their dependability in actual clinical contexts. In order to create a reliable and clinically useful pregnancy risk prediction model, several issues must be resolved. In order to categorize risk levels according to important maternal health factors like age,

*The names of the author and co-authors have not been disclosed, as the respective contributors transferred the copyright and publication rights to the journal but did not complete the required formalities in accordance with COPE guidelines and the journal's policies. For any further information regarding this article, readers are kindly requested to contact the Editor-in-Chief of the journal.

Article Timeline: Received: Oct 08, 2025; **Revised:** Nov 05, 2025; **Accepted:** Nov 20, 2025; **Published:** Nov 30, 2025

blood pressure, and blood glucose levels, this study presents MaternityAI, a machine learning-based pregnancy risk prediction system. The suggested approach finds the most important factors influencing pregnancy risks by integrating several classification algorithms and evaluating their performance with key performance indicators such as F1-score, recall, accuracy, and precision. The incorporation of machine learning into maternal health services has the potential to address gaps in accessibility and efficiency, especially in areas with limited resources. In many rural and underserved regions, there is a shortage of medical professionals to perform timely risk evaluations, resulting in delayed identification of high-risk pregnancies. A predictive system such as MaternityAI can function as a decision-support tool, offering automated risk analysis based on easily obtainable health metrics. By utilizing cloud-based or mobile health technologies, these models can enable remote monitoring, allowing for early interventions even in locations with inadequate healthcare infrastructure. This transition towards AI-enhanced maternal care not only improves diagnostic precision but also fosters equitable healthcare access, minimizing disparities in maternal and fetal health outcomes. MaternityAI seeks to help medical professionals make prompt, well-informed decisions by offering a standardised and objective method of evaluating pregnancy risk.

2. Literature Survey

By enhancing patient management and pregnancy risk prediction, machine learning has greatly improved maternal healthcare. Numerous studies have investigated different predictive models and approaches to improve the interpretability and accuracy of risk classification. One method for predicting maternal risk is ensemble learning. In order to enhance maternal health risk categorisation, Smith et al. [1] presented the Quad-Ensemble Machine Learning framework (QEML-MHRC), which integrates several classification models. In a similar vein, Zhang and Chen [2] evaluated the efficacy of machine learning models for preeclampsia prediction in detecting high-risk pregnancies using a comprehensive review.

Patel and Mehta [3] created a machine learning model specifically designed for pregnancy risk prediction in rural India in order to address maternal healthcare issues in settings with low resources. Their findings emphasised the significance of AI-based treatments in improving maternal health outcomes. Furthermore, Rao and Sharma [4] proposed a risk prediction model for miscarriage utilising sophisticated classification algorithms, concentrating on immune-related pregnancy problems. One important consideration in healthcare is the interpretability of predictive models. Explainable Boosting Machines (EBMs) were used by Anderson and Grey [5] to forecast maternal and foetal outcomes, offering important information about risk variables for issues such severe maternal morbidity and premature preeclampsia. For foetal health prediction, Singh et al. [6] investigated an ensemble model that combined several hyperparameter-tuned algorithms, showing notable accuracy gains.

Other research has extended the application of machine learning to broader pregnancy-related concerns, such as systematic reviews on pregnancy outcome prediction [7] and precision medicine approaches for maternal healthcare [8]. Predictive models have been utilised in a variety of fields outside of maternal health. A useful manual for creating predictive models with an emphasis on machine learning trends and techniques was presented by Williams et al. [9]. In their analysis of predictive learning analytics developments, Brown and Clark [10] demonstrated how AI models adapt to manage intricate medical data. Kumar et al. [11] showed how machine learning-based predictive models might enhance clinical judgement, especially in situations like early sepsis detection [12]. Predictive analytics has also been investigated for uses outside of maternal health, including generalised forecasting methods for risk assessment [15], educational data mining [14], and mental health risk prediction [13]. Studies examining machine learning applications in maternal and foetal health have also thoroughly examined the integration of machine learning in healthcare [16].

MaternityAI improves pregnancy risk prediction by combining many machines learning algorithms, building on these earlier findings. The solution tackles important issues including data accessibility, model interpretability, and practical clinical implementation that have been noted in earlier studies. MaternityAI seeks to close current gaps and offer a complete, scalable solution for maternal healthcare by utilising predictive analytics.

3. Methodology

From data preprocessing to model evaluation, there are several stages involved in the deployment of MaternityAI: Predicting Pregnancy Risks using Machine Learning. A predictive model development process and optimization strategy with system development steps and optimization techniques are described in this section for effective pregnancy-related risk forecasting.

A. Data Collection and Preprocessing

Medical records, clinical databases, and publicly accessible health datasets were used to curate the dataset for this study. Key indicators of maternal health are included, including.

- Age
- Avoid Body Temperature(F)
- Heart Rate(bpm)
- Blood Glucose (Fasting & HbA1c)
- BMI (kg/m^2)
- Systolic & Diastolic Blood Pressure (mm Hg)

Mean and mode imputation were applied to numerical data in order to guarantee completeness during data cleaning and management of missing values. Z-score analysis and box plots were used for outlier detection in order to find and reduce anomalies. Furthermore, standardisation (Z-score normalisation) was used to apply feature scaling in order to keep all characteristics on the same scale and avoid any one feature having an undue influence on the model. The dataset was separated into three groups for data splitting: 70% for model training, 15% for validation to adjust hyperparameters, and 15% for testing to evaluate the final model's performance.

B. Model Development

Researchers analyzed several machine learning models in order to identify the most suitable classifier for pregnancy risk prediction, including:

- Gradient Boosting Classifier
- Random Forest Classifier
- Logistic Regression
- XGBoost (Final Model)

XGBoost was chosen based on comparable performance because of its excellent accuracy, robust feature selection, and capacity to manage unbalanced data.

We used a number of strategies to remedy the dataset's class imbalance. To enhance the representation of high-risk cases during training, the Synthetic Minority Oversampling Technique (SMOTE) was employed. Furthermore, the `$scale\_pos\_weight$` parameter of XGBoost was adjusted to more effectively penalise minority class misclassification. To maintain the original class distribution during data splitting, stratified sampling was also employed. Together, these techniques enhanced the model's memory for the high-risk group, assisting in the reduction of false negatives in crucial situations.

Key parameters were optimised for better performance using $GridSearchCV$ during model training and hyperparameter tuning. In order to guarantee convergence and avoid overfitting, the learning rate was carefully changed. To manage model complexity and prevent overfitting problems, the maximum depth parameter was adjusted. In order to balance variance and bias, the number of estimators was optimised. In order to improve generalisation and avoid overfitting and guarantee a strong and effective model, L1 and L2 regularisation techniques were also used.

C. Model Evaluation

The model's categorisation accuracy across various pregnancy risk levels is revealed via the confusion matrix:

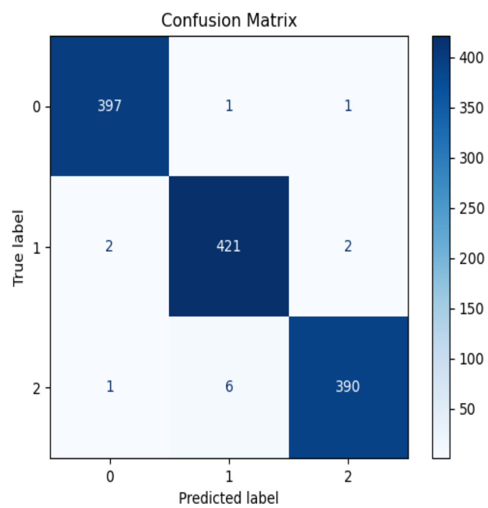

Fig. 1. Confusion Matrix

The model demonstrated high true positive rates, indicating effective classification and strong predictive capability. Additionally, the low number of misclassifications highlights the model's reliability and accuracy in making correct predictions.

To determine which maternal health markers, have a major influence on pregnancy risk classification, feature importance was examined:

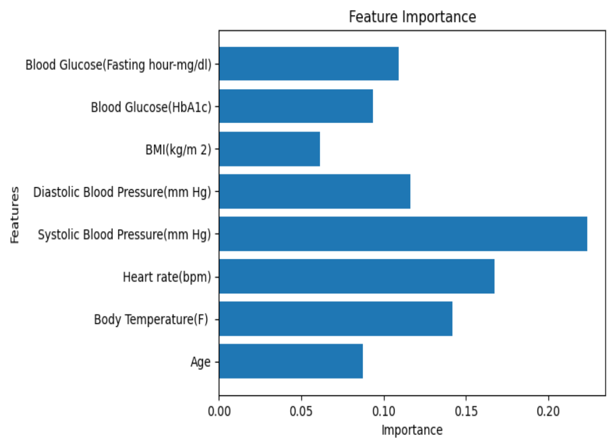

Fig. 2. Feature Importance

Systolic blood pressure was found to be the strongest predictor. Other important factors in risk assessment included heart rate, body temperature and blood glucose levels. While BMI and age had a comparatively lesser influence, they still contributed to the overall risk assessment.

Important metrics were computed in order to assess the model:

Table I: Performance Metrics of Original and Optimized Models

Metric	**Original Model**	**Optimized Model**
Accuracy	98.94%	99.10%
Precision	8.94%	99.10%
Recall	98.94%	99.10%
F1 Score	98.94%	99.10%

The model that was optimised demonstrated modest but significant gains in F1-score, recall, accuracy, and precision.

D. Model Deployment and Integration

Deployment considerations were planned once the final model was evaluated to guarantee successful integration. For easy accessibility, the model can be incorporated into a cloud-based healthcare system or delivered using Flask or FastAPI. Because of its scalable design, real-time inference in medical applications is made possible. Furthermore, by offering comprehensible insights, feature importance analysis improves interpretability and qualifies the model for clinical decision assistance.

E. System Architecture

Fig. 3. Architecture Diagram

4. Results and Discussion

A. Model Performance Evaluation

When tested on unseen data, the trained model showed excellent classification accuracy. The optimized model improved **precision, recall, and F1-score**, resulting in an accuracy of **99.10%**, as shown in **Table 1 (Methodology section)**. These findings demonstrate that hyperparameter tuning improved prediction reliability and effectively decreased misclassification errors.

Key observations from model evaluation:

- More high-risk pregnancies are accurately detected thanks to the **optimized model's decreased false negatives.**
- The model's **high precision** reduces needless alarms, making it **dependable for clinical use.**
- Nearly all real high-risk pregnancies are identified thanks to a **99.10% recall**.

B. Confusion Matrix Analysis

The classification performance across various pregnancy risk levels is broken down in detail in the confusion matrix. The model can **accurately classify pregnancy risk levels** with minimal misclassification, as shown by the confusion matrix in **Figure 1(Methodology section).**

The key observations are:

- At every risk level, the model showed excellent classification accuracy.
- With more training data, the few moderate-risk cases that were incorrectly classified as high-risk could be resolved.
- The model's potential for practical clinical applications is further supported by its low misclassification rate.

E-ISSN: 3048-7021 (Online)

C. Feature Importance Analysis

Interpretability depends on knowing which characteristics are most important in predicting pregnancy risk. The most significant predictors of pregnancy risk are blood glucose level, systolic blood pressure and heart rate according to the feature importance analysis displayed in **Figure 2 (Methodology section).**

Key Insights:

- The most significant factor was systolic blood pressure, which is consistent with well-established medical research on pregnancy complications.
- Blood glucose (HbA1c and fasting levels) and heart rate were also significant predictors, confirming their use in evaluations of maternal health.
- BMI and age were comparatively less significant, indicating that biometric indicators are more important in predicting risk than demographic factors.

Relationships between features were also analyzed using the correlation matrix.

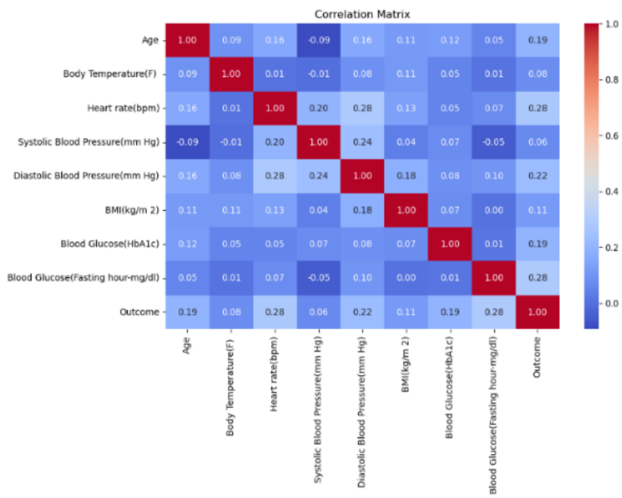

Fig. 4. Correlation Matrix

Blood pressure and heart rate showed a moderate correlation, indicating that they both have an impact on pregnancy risk. A few weaker correlations point to possible feature interactions that might need more research.

D. Comparison with Traditional Approaches

A comparison between the machine learning-based approach and traditional rule-based risk assessment methods was carried out in order to assess the model's impact.

Table 2: Comparison of Traditional Risk Scoring and Machine Learning-Based Prediction

Method	Accuracy	Limitations
Traditional Risk Scoring	~85%	Fixed thresholds, lacks adaptability.

Machine Learning (Ours)	99.10%	Requires robust data preprocessing, sensitive to data quality.

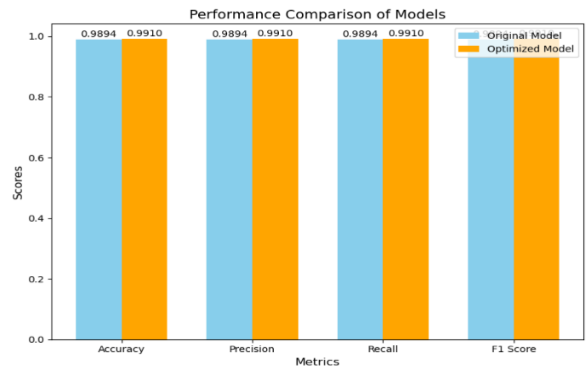

Fig. 5. Performance Comparison

The machine learning model's high recall (99.10%) guarantees that almost all high-risk pregnancies are detected, as shown in Table 1 (Methodology section). This makes it noticeably more effective than conventional rule-based techniques.

Key Advantages of the ML-Based Approach**:**

- **Dynamic Learning:** Unlike conventional models, machine learning algorithms have the ability to adjust to patterns in data as it is received in real time
- **Enhanced Accuracy:** Increases prediction reliability by identifying hidden relationships between risk factors.
- **Scalability**: With regular updates, it can be implemented in various clinical settings.

5. Conclusion and Future Scope

A. Conclusion

The suggested MaternityAI system outperforms conventional risk assessment methods and effectively forecasts pregnancy risk levels using machine learning techniques, with 99.10% accuracy. Through the integration of hyperparameter tweaking and enhanced feature selection, the system exhibits great reliability in recognizing pregnancies at high risk. According to the evaluation results, the optimised model guarantees better early detection of problems by dramatically increasing recall and precision. The model is appropriate for clinical decision assistance because the confusion matrix analysis verified that there were few misclassifications. Additionally, the system's potential as a decision-support tool for medical practitioners was strengthened by feature importance analysis, which identified important maternal health parameters influencing risk prediction. Before being fully used in clinical settings, a number of issues, such as data bias, model interpretability, and regulatory compliance, need to be resolved, even though the current implementation shows strong predictive capability.

B. Future Scope

A number of developments can be investigated in order to advance MaternAI and increase its practicality:

1. Diversity and Expansion of the Dataset: Adding bigger, more varied datasets from various healthcare organisations will improve the model's ability to generalise, lowering biases and enhancing predictions across various populations.
2. Explainable AI (XAI) for Improved Interpretability: By using SHAP values or LIME, medical practitioners can better understand and trust the model's predictions.
3. Real-Time Implementation in Healthcare Systems: The model can be integrated into electronic health record (EHR) systems to give real-time pregnancy risk evaluations, which will automatically provide early warning signals to physicians.
4. Implementation via Mobile & Cloud: By creating a mobile or cloud-based application, patients and healthcare professionals can enter health parameters and obtain risk assessments immediately, increasing accessibility in isolated or underdeveloped locations.
5. Model Interpretability using Explainable AI: Additionally, the model's interpretability can be improved by including Explainable AI (XAI) strategies like SHAP (SHapley Additive Explanations). Healthcare practitioners may better comprehend and have more faith in the system's decision-making process thanks to SHAP, which makes it possible to visualise how each input feature contributes to the model's predictions.

C. Final Remarks

By offering early, precise, and data-driven pregnancy risk forecasts, the MaternityAI system has the potential to completely transform maternal healthcare. Medical professionals can reduce maternal and newborn problems by using machine learning in clinical settings to help them intervene promptly. In order to establish AI-assisted pregnancy risk prediction as a common practice in obstetric care, future research will concentrate on enhancing model interpretability, growing datasets, and permitting real-time deployment.

6. References

[1] Khadidos, A. O., Saleem, F., Selvarajan, S., Ullah, Z., & Khadidos, A. O. (2024). Ensemble machine learning framework for predicting maternal health risk during pregnancy. Scientific Reports, 14, 21483.

[2] Ranjbar, A., Montazeri, F., Ghamsari, S. R., Mehronoush, V., Roozbeh, N., & Darsareh, F. (2024). Machine learning models for predicting preeclampsia: A systematic review. BMC Pregnancy and Childbirth, 24, 6.

[3] Zhou, X., & Li, L. (2021). A machine learning-based risk prediction model during pregnancy in rural India. Journal of Healthcare Engineering.

[4] Wu, Y., Yu, X., Li, M., Zhu, J., Yue, J., Wang, Y., Man, Y., Zhou, C., Tong, R., & Wu, X. (2024). Risk prediction model based on machine learning for miscarriage in immune-abnormal pregnancies. Frontiers in Pharmacology.

[5] Bosschieter, T. M., Xu, Z., Lan, H., Lengerich, B. J., Nori, H., Painter, I., Souter, V., & Caruana, R. (2023). Interpretable predictive models to understand risk factors for maternal and fetal outcomes. Journal of Healthcare Informatics Research.

[6] Talukder, M. S. H., & Akter, S. (2023). An improved model ensembled of different hyper-parameter tuned machine learning algorithms for fetal health prediction. International Journal of Information Technology.

[7] Islam, M. N., Mahmud, S. N. M. T., & Khan, N. I. (2022). Machine learning to predict pregnancy outcomes: A systematic review. BMC Pregnancy and Childbirth.

[8] Eloranta, S., & Borman, M. (2022). Predictive models for clinical decision making: Deep dives in precision psychiatry and outcome predictive after lymphoma. Journal of Internal Medicine.

[9] Gupta, R., Sharma, A., & Alam, T. (2023, November). Building predictive models with machine learning.

[10] Bokonda, L., Khadija, Q. T., & Souisii, N. (2021). Predictive analysis using machine learning: Review of trends and methods. In Proceedings of the International Conference on Artificial Intelligence and Smart Systems (ICAIS).

[11] Stahl, D. S. (2023). New horizons in prediction modelling using machine learning in clinical practice. Age and Ageing.

[12] Zhai, Y., Zhang, Y., Chu, Z., Geng, B., Almaawali, M., Fulmer, R., Lin, Y.-W. D., Xu, Z., & Daniels, A. D. (2024). Machine learning predictive models to guide prevention and early intervention of anxiety and depressive disorders. Journal of Counseling & Development.

[13] Dr., X. (2020). Predictive models with machine learning algorithms: A case study in educational data mining. Open Science Framework (OSF).

[14] Sgir, N., Adadi, A., & Lahmer, M. (2023). Recent advances in predictive learning analytics: A decade systematic review. Education and Information Technologies.

[15] Calvert, J., Nemati, S., Scherpf, M., & Marghertia, R. (2023). Application of machine learning predictive models for early detection of sepsis. Critical Care.

[16] Mennickent, D., Rodriguez, A., Opazo, M., Riedel, C. A., Castro, E., Eriz-Salinas, A., Appel-Rubio, J., Aguayo, C., Damiano, A. E., Guzman-Guttierez, E., & Araya, J. (2023). Machine learning applied in maternal and fetal health: A narrative review. Frontiers in Endocrinology.

7.Conflict of Interest

The authors declare that there are no conflicts of interest regarding the publication of this article.

8.Funding

No external funding was received to support or conduct this study.