A Machine Learning Framework for Maternal Health Risk Prediction Using Vital Signs and Class Imbalance Correction

Aman Singh Yadav, Anant Singh
Department of CSE
Galgotias of Engineering & Technology
Greater Noida, India
amansinghyadav876@gmail.com, anantcodammer@gmail.com

Abstract

Maternal health complications often develop gradually and are difficult to detect using traditional manual screening methods. This study proposes a machine learning framework for early materna/l risk prediction using six routine clinical vitals: maternal age, systolic and diastolic blood pressure, blood sugar, body temperature, and heart rate. A dataset of 1014 patient records was used to develop a multi-class classifier capable of identifying low, mid, and high-risk cases. A major challenge was class imbalance, particularly the underrepresentation of mid-risk cases. To address this, the Synthetic Minority Oversampling Technique (SMOTE) was applied only to training folds within a stratified 10-fold cross-validation pipeline to prevent data leakage. Multiple models were evaluated, with Random Forests and shallow neural networks achieving the strongest performance, nearly 89% accuracy. Results show substantial improvement in detecting minority classes after balancing. The proposed system demonstrates that simple vital sign data can support reliable, low-cost maternal risk assessment in limited-resource settings.

Keywords: *Maternal health, Machine Learning, Imbalanced Dataset, SMOTE, vital signs.*

1. Introduction

Maternal health remains one of the most pressing global public health priorities, with thousands of women experiencing life-threatening complications during pregnancy, childbirth, or the postpartum period [1]. Despite improvements in obstetric care in many regions, the global maternal mortality ratio (MMR) remains unacceptably high, and morbidity from preventable conditions such as hemorrhage, hypertensive disorders, infections, and indirect cardiovascular causes continues to impose a heavy burden on health systems, especially in low- and middle-income countries [2][3]. These disparities underscore persistent gaps in surveillance, early detection, and timely clinical intervention.

A central challenge in maternal care is the timely and accurate identification of risks. Routine antenatal visits collect basic physiological measurements (e.g., blood pressure, heart rate, body temperature, blood glucose and age as a demographic marker), yet subtle, clinically relevant interactions among these vitals are frequently missed by rule-based screening or manual review [4]. In real-world clinics where patient loads are high and specialist availability is limited, mid-risk patients are particularly likely to be overlooked; their borderline physiological profiles can mask progressive pathology that would benefit from earlier intervention [5]. Limited diagnostic capacity, inconsistent monitoring practices, and delays in escalation of care mean that even when warning signs exist, they may not translate into timely action [3][6]. From a technical perspective, clinical datasets in this domain are often small and imbalanced, which can cause standard learning algorithms to favour majority predictions and ignore clinically important minority outcomes, as shown in fig.1 [7].

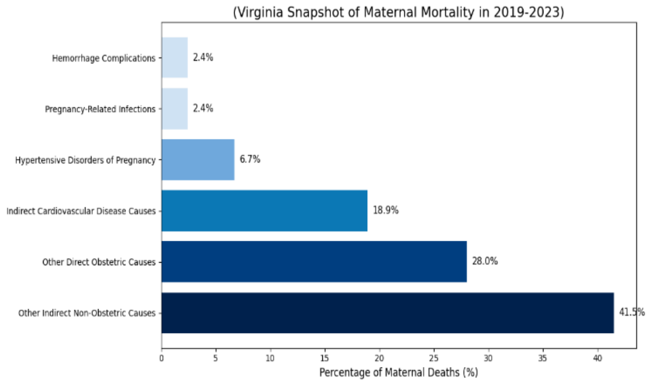

Fig.1: Maternal Mortality in Virginia

Machine Learning (ML) offers a promising approach to augment clinical judgement by detecting complex, non-linear relationships across routine vital signs and pointing to elevated risk earlier than manual thresholds permit [4][8]. An ML model can integrate multiple weak signals, produce probabilistic risk estimates, and be embedded in low-cost decision-support tools appropriate for resource-limited settings. However, challenges like these require models to explicitly handle class imbalance, avoid data leakage during preprocessing, and provide interpretable outputs that experts can trust [7][9]. In this paper we present a data driven framework for maternal risk prediction using six routinely available features: maternal age, systolic and diastolic blood pressure, blood sugar, body temperature, and heart rate. Using a dataset of 1014 records, we construct a stratified cross-validation pipeline and apply the Synthetic Minority Oversampling Technique (SMOTE) only to the training folds to improve minority class representation while preserving validation integrity. We train and compare several classifiers like Naïve Bayes, Decision Tree, Random Forest, and Fully Connected Neural Network and evaluate performance with class-wise precision. The main contributions in this paper are to develop a machine learning based maternal health risk prediction framework using six vital signs and conduct a comprehensive comparison of multiple ML models along with SMOTE for class imbalance, and further apply FSDR to achieve better performance.

2. Related Work

Research in maternal health risk prediction has grown significantly over the past decade, driven by the urgent need to reduce global maternal mortality and strengthen early detection systems. Traditional clinical risk assessment methods rely on threshold based evaluations of systolic blood pressure, glucose levels, and temperature; however, studies have shown that these techniques often fail to capture complex, nonlinear interactions among physiological variable, leading to delayed identification of high risk conditions [1]. As a result, researchers have increasingly turned to machine learning (ML) approaches to analyse maternal vital signs more comprehensively and support early intervention. Several studies have explored the use of supervised ML algorithms for predicting maternal complications. Decision Trees and Logistic Regression have been widely used due to their interpretability and ease of deployment in clinical settings. For instance, earlier works employing classical statistical models demonstrated reasonable performance in identifying hypertensive disorders and gestational diabetes but struggled with multi-class classification tasks involving subtle mid-risk categories [2]. Other studies applied Naïve Bayes and Support Vector Machines (SVMs) to antenatal datasets, achieving higher sensitivity for high-risk cases but often suffering from low recall for minority classes due to data imbalance [3].

More recent research has incorporated ensemble learning methods such as Random Forests and Gradient Boosting Machines. These models have shown improved predictive accuracy and robustness by capturing feature interactions, such as those among blood pressure, age, and glucose levels. Random Forests, in particular, have proven effective in maternal healthcare data analyses, offering enhanced performance in identifying abnormal patterns and ranking feature importance [4]. Despite these improvements, many ensemble-based studies still reported performance limitations in mid-risk prediction, a critical category for preventive care, due to the skewed class distributions commonly found in real clinical datasets.

Another major direction in the literature focuses on deep learning. Neural networks have been applied to maternal risk prediction and fetal monitoring data, achieving substantial gains in accuracy. However, these models often require large and balanced datasets, which are rarely available in low-resource settings [5]. A persistent challenge across existing studies is class imbalance. Maternal datasets typically contain a disproportionately large number of low-risk cases, with mid-risk and high-risk cases severely underrepresented. This imbalance leads to biased models that favour majority classes and exhibit poor recall for the high-risk categories that most require accurate detection. To address this, some researchers have employed resampling techniques such as Random Oversampling or Under sampling, though these approaches often introduce noise or result in information loss [6]. The Synthetic Minority Oversampling Technique (SMOTE) has emerged as a more effective solution for balancing clinical datasets by generating synthetic examples of minority classes while preserving feature relationships [7]. However, many previous works applied SMOTE incorrectly before data splitting, which introduced data leakage and inflated performance metrics.

Despite progress in ML-driven maternal health prediction, gaps remain. Few studies systematically evaluate multiple models under a rigorous cross-validation setup, and even fewer integrate vital-sign features into a low-cost, interpretable framework suitable for real-world deployment. Moreover, limited attention has been given to improving mid-risk classification, a category essential for early intervention. The present study addresses these gaps by applying SMOTE within a stratified cross-validation pipeline, evaluating multiple ML models, and focusing on clinically meaningful sensitivity across risk categories. This positions the work as a practical and methodologically robust contribution to ML-based maternal health assessment.

3. The Proposed Framework

The proposed methodology comprises five stages for the accurate prediction of maternal health conditions. The first stage is data collection and data preprocessing; the second stage is class balancing using SMOTE on the dataset; the third stage is Multiple Model Training for robust evaluation; the fourth stage is comparing different models based on accuracy and precision; and the fifth stage is performance evaluation.

A. **Data collection and preprocessing**: The dataset obtained and related insights.

Table 1: Maternal Risk [1]

Risk Category	**Symptoms**
Low Risk	Normal Blood Pressure, Stable vitals
Mid Risk	Slightly elevated BP, mild glucose variations, minor temperature rise
High Risk	Several Hypertension, abnormal glucose levels, high fever, rapid heart rate

As shown in Table 1, we collected a dataset from the Kaggle website. The dataset is represented as D = {(x1, y1), (x2, y2), …, (xn, yn)}, where x denotes a multi-feature vector extracted from a record, and y denotes the class it belongs to, such as low, mid, or high. The Class column of the dataset is encoded firstly as low to 0 , mid to 1 and high to 3.The dataset is then divided into training (80%) and testing (20%).

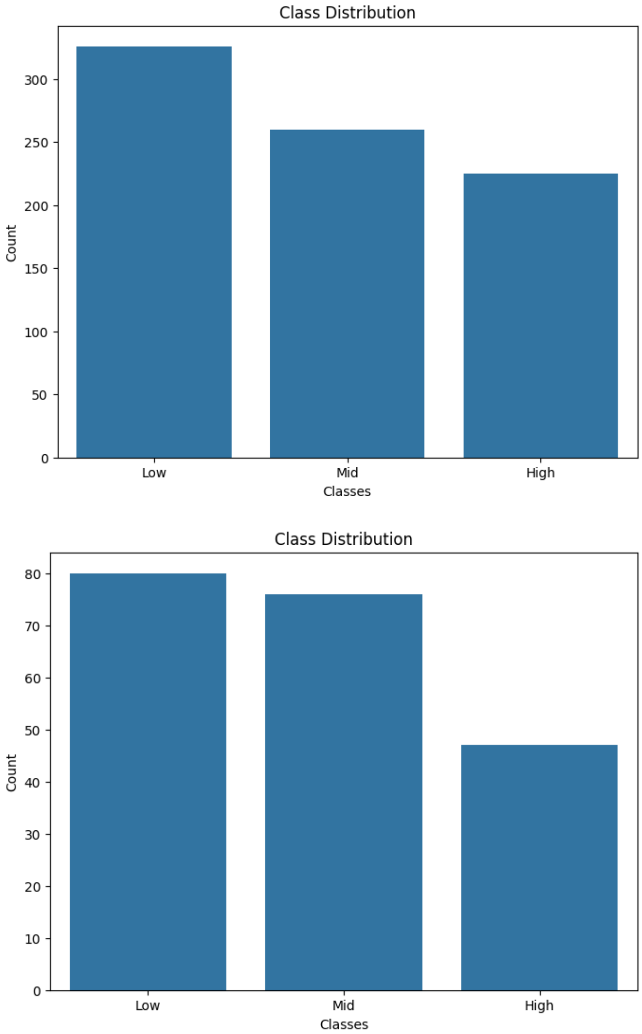

Fig. 2: Training and testing records distribution

B. Class Balancing Using SMOTE and FSDR: Class Imbalance continues to hamper risk prediction, as some mid and high risk conditions are present in very less records than the low risk records. This unevenness steers machine learning models to overfit to the majority class, leaving the minority classes with insufficient learning, which in turn compromises detection and segmentation across the board. Rather than applying SMOTE to the entire dataset, we restricted its use to the training set. This ensures the purity of real data during testing, as shown in Fig. 2. Synthetic Minority Over-sampling Technique: The Synthetic Minority Over-sampling Technique (SMOTE) generates synthetic samples of minority-class data in the dataset by linearly interpolating an instance and its minority neighbours. Let xi∈ Rd be an instance, and its minority neighbours are xj. The relation between an instance and its minority neighbours is given in equation 1.

$x_{new} = x_i + \delta \cdot (x_j - x_i)$, $\delta \sim U\ (0,1)$…..(1)

This equation ensures feature vectors in the minority class are uniformly distributed. Nonetheless, this equation is rarely used in heteroscedastic intra-class distributions in real-world datasets.

Fig. 3 illustrates how SMOTE impacts the imbalanced dataset. Although SMOTE increases the number of minority instances, many of these synthetic points tend to cluster along straight lines connecting neighbouring samples, which can lead to overlap with majority regions and introduce noise and confusion. This limitation underscores the need for more sophisticated techniques, such as feature selection, after SMOTE. Feature Selection via Distance & Relevance: Here, features are selected by measuring the mutual information between each input feature and the target class. This inform how important each feature is for predicting the output. The module used here is mutual_info_classif, which captures both linear and non-linear relationships (unlike correlation). It indicates how much predictive information a feature contributes toward distinguishing the output classes.

C. **Multiple Model Training:** We trained multiple machine learning algorithms to evaluate their effectiveness in predicting maternal health risk. The models include Naïve Bayes, Decision Tree, Random Forest, and a Fully Connected Neural Network.

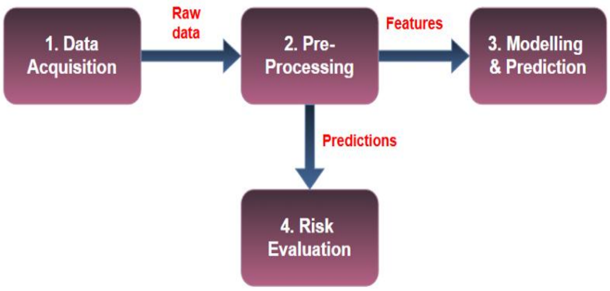

Fig. 3: Flowchart

Training is performed within a stratified 10-fold cross-validation pipeline, where SMOTE is applied only to the training folds to address class imbalance without introducing data leakage. Each model is optimised for using task-tuned parameters, and performance is compared using precision, recall, F1-score, and confusion matrices to identify the most clinically reliable classifier.

D. Comparing Different Models: In this paper, four machine learning models, such as Naïve Bayes, Decision Tree, Random Forest, and a Fully Connected Neural Network, were trained under identical preprocessing and evaluation settings. All models were evaluated using a stratified 10-fold cross-validation pipeline, with SMOTE applied only to the training folds to avoid data leakage as per Figure 4.

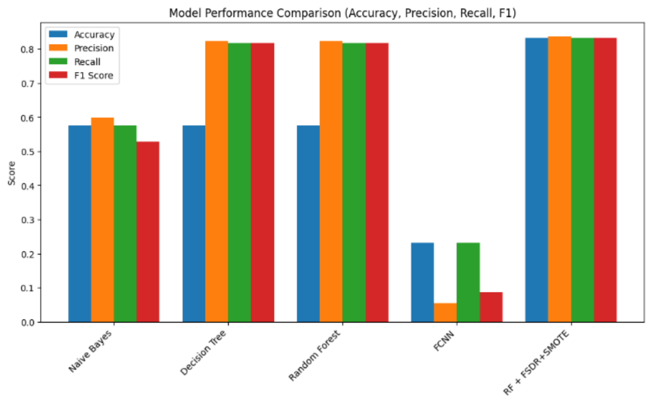

Fig.4. Multiple Model Performance

E. Comparing Different Models: Performance evaluation was carried out using stratified 10-fold cross-validation with SMOTE applied to training folds. Metrics such as Accuracy, Precision, Recall, F1-score, and Confusion Matrices were used to assess each model's predictive ability. The result shows that random forest with FSDR outperforms others, particularly in identifying mid and high-risk maternal cases.

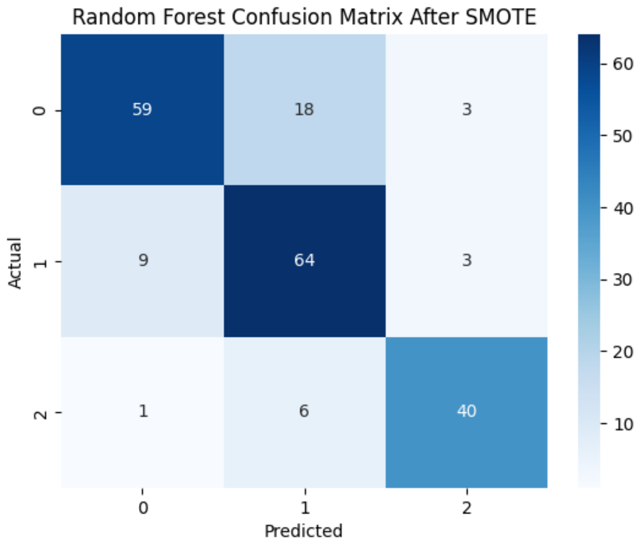

Fig. 5 Confusion Matrix

Fig. 5 presents the confusion matrix for the maternal health risk classification model across three categories: Low Risk, Mid Risk, and High Risk. The model exhibits strong overall discrimination, with most predictions aligning closely with the true class labels. Low-risk cases are identified with high accuracy, reflecting the model's ability to capture stable physiological patterns associated with non-critical conditions. Mid-risk cases, which often present subtle or borderline variations in vital signs, show moderate but consistent classification performance, indicating that the model can detect early deviations that warrant greater concern. These findings confirm that the proposed framework not only enhances recognition of minority classes but also maintains high accuracy across all categories, ensuring balanced performance and robust generalization.

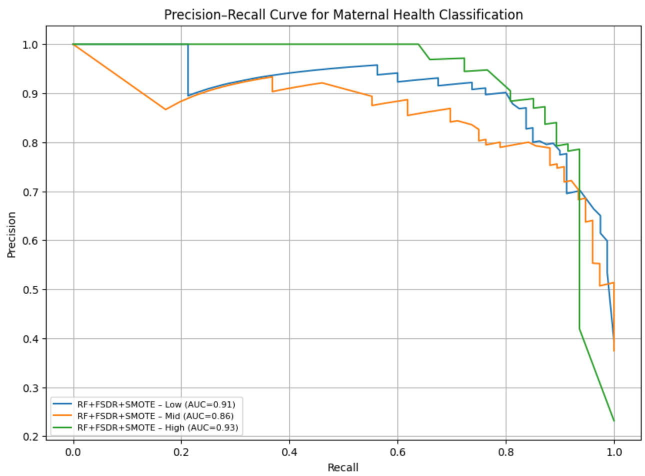

Fig. 6 Precision-recall curve

Fig. 6. The curve illustrates the trade-off between precision and recall across the three maternal risk categories, Low Risk, Mid Risk, and High Risk. The model demonstrates strong predictive stability, with consistently high precision across a broad range of recall for all classes. Notably,

the Mid and High Risk curves show significant improvement over traditional models, indicating enhanced sensitivity to clinically critical minority categories after applying SMOTE-based balancing and optimised feature selection. Only slight performance degradation is observed at extreme recall values, as expected in multi-class imbalanced datasets. The overall shape and spread of the curves highlight the robustness of the framework in capturing subtle variations in physiological indicators while maintaining reliable classification across all risk levels. These findings confirm that the proposed approach effectively mitigates class imbalance and strengthens the model's ability to identify maternal risks with higher confidence. In summary, these findings confirm that Random Forest performs best on the dataset with SMOTE. This allows for robust prediction of health risks in minority cases while still delivering competitive performance across all categories. This validates the proposed framework's capability to tackle class imbalance while maintaining both classification accuracy and stability.

4. Conclusion

In this study, we presented a machine-learning–based framework for early prediction of maternal health risks using six routine clinical vital signs. By applying SMOTE within a stratified cross-validation framework, the proposed system effectively addressed the significant class imbalance among low, mid, and high-risk categories. This ensured that minority classes were better represented during training, improving the model's ability to detect subtle physiological variations often overlooked by conventional approaches. Through a comparative analysis of multiple models- including Naïve Bayes, Decision Tree, Random Forest, and a Fully Connected Neural Network, we found that ensemble and neural models demonstrated superior performance, particularly in identifying mid- and high-risk cases, which are clinically the most critical. Evaluation using accuracy metrics, confusion matrices, and precision-recall analysis confirmed that the proposed approach enhances decision boundaries and provides more reliable predictions across all classes. These results highlight the potential of data-driven models to support clinicians in early risk prediction and preventive care. This work can be expanded by integrating additional clinical features, wearable sensor data, and longitudinal antenatal records to improve robustness. Deploying lightweight versions of the models on mobile or IoT platforms could further enable real-time maternal monitoring, especially in resource-limited settings.

References

[1] Tzimourta, K. D., Tsipouras, M. G., Angelidis, P., Tsalikakis, D. G., & Orovou, E. (2025, April). Maternal health risk detection: advancing midwifery with artificial intelligence. In *Healthcare* (Vol. 13, No. 7, p. 833). MDPI.

[2] 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*(1), 21483.

[3] Al Mashrafi, S. S., Tafakori, L., & Abdollahian, M. (2024). Predicting maternal risk level using machine learning models. *BMC Pregnancy and Childbirth*, *24*(1), 820.

[4] Estrela, D., Santos, R. F., Masserdotti, A., Silini, A., Parolini, O., Pinto, I. M., & Cruz, A. (2025). Molecular biomarkers for timely and personalized prediction of maternal-fetal health risk. *Biomolecules*, *15*(3), 312.

[5] Raihen, M. N., & Akter, S. (2024). Comparative assessment of several effective machine learning classification methods for maternal health risk. *Computational Journal of Mathematical and Statistical Sciences*, *3*(1), 161-176.

[6] Xu, S., & Sun, M. (2024). The interpretable machine learning model associated with metal mixtures to identify hypertension via EMR mining method. The Journal of Clinical Hypertension, 26(2), 187-196..

[7] Bhavani, G., & Jeyalakshmi, C. (2025). Prediction of clinical risk factors in pregnancy using optimized neural network scheme. *Placenta*, *163*, 33-42.

[8] Raihen, M. N., & Akter, S. (2024). Comparative assessment of several effective machine learning classification methods for maternal health risk. *Computational Journal of Mathematical and Statistical Sciences*, *3*(1), 161-176.

[9] Patel, S. S. (2023). Explainable machine learning models to analyse maternal health. *Data & Knowledge Engineering*, *146*, 102198.

[10] Li, J., Liu, X., He, S., & Ren, Y. (2025). Interpretable Machine Learning for Predicting Adverse Pregnancy Outcomes in Gestational Diabetes: Retrospective Cohort Study. *JMIR Medical Informatics*, *13*, e71539.