International Journal of Computer Science Engineering Techniques – Volume 9
Issue 6, November - December - 2025

Women Maternal Risk Prediction using AI

Chandana T.[1], Swati D Mahindrakar [2]

[1]MCA Student, Faculty of Computing and IT, GM University, Davangere, India

[2]Assistant Professor, Faculty of Computing and IT, GM University, Davangere, India

Corresponding Author: chandana03032002@gmail.com

Abstract

maternal mortality and morbidity remain significant global health challenges. Early and accurate identification of at-risk pregnancies is crucial for timely intervention and improved outcomes. This paper explores the application of machine learning (ML) for predicting maternal health risks. We employ a publicly available dataset featuring key physiological and demographic indicators to develop and evaluate three distinct classification models: K-Nearest Neighbours (KNN), Naive Bayes, and Decision Tree. The methodology involves data pre-processing, feature scaling, model training, and performance evaluation. Our results demonstrate that the Decision Tree classifier achieves the highest accuracy in identifying risk levels (low, medium, high). The comparative analysis reveals the Decision Tree's superior capability in handling the dataset's characteristics, offering an interpretable and effective model for clinical decision support. This work underscores the potential of ML to augment traditional risk assessment methods, providing a scalable and data-driven tool for healthcare professionals.

Keywords-*Maternal Health, Risk Prediction, Machine Learning, Decision Tree, K-Nearest Neighbours, NaiveBayes, Predictive Analytics.*

Introduction

Maternal health is a cornerstone of public health and societal well-being. According to the World Health Organization (WHO), approximately 800 women die every day from preventable causes related to pregnancy and childbirth [1]. The vast majority of these deaths occur in low-resource settings, where access to timely and quality healthcare is limited. Key risk factors contributing to adverse maternal outcomes include advanced maternal age, hypertension, diabetes, and abnormal physiological parameters like heart rate and blood sugar levels. Traditional methods for assessing maternal risk often rely on clinical judgment and standardized checklists, which may not capture the complex, non-linear interactions between various risk factors. The advent of digital health records has generated vast amounts of patient data, creating an opportunity for more sophisticated, data-driven approaches. Machine Learning (ML), a subset of artificial intelligence, excels at identifying intricate patterns and relationships within large datasets. By training models on historical patient data, ML can create predictive tools that classify a patient's risk level with high accuracy [2]. Such tools can serve as powerful decision support systems for clinicians, enabling them to prioritize care, allocate resources effectively, and implement preventative measures for high-risk individuals before complications arise.

This paper presents a comparative study of three fundamental ML classification algorithms K-Nearest Neighbours (KNN), Naive Bayes, and Decision Tree—for the task of maternal health risk prediction. Our primary objective is to determine which of these models provides the most accurate and reliable classification on a standard maternal health dataset. We detail our methodology, from data pre-processing to model evaluation, and present our findings, highlighting the superior performance of the Decision Tree algorithm.

Materials and Methods

Our methodology follows a structured workflow designed to ensure robust and reproducible results. It begins with data acquisition and culminates in a comparative performance analysis of the trained models.

A. Dataset Description

The study utilizes the "Maternal Health Risk Data Set" available from the UCI Machine Learning Repository [9]. This dataset contains 1014 instances and 7 attributes. The features include:

Age: Age of the mother in years. SystolicBP: Upper value of Blood Pressure (mmHg).

DiastolicBP: Lower value of Blood Pressure (mmHg).

BS: Blood Sugar levels (mmol/L).

BodyTemp: Body temperature in Fahrenheit.

Heartrate: Resting heart rate in beats per minute.

The target variable, Risk Level, is a categorical feature with three classes: 'low risk', 'mid risk', and 'high risk'.

B. Data Pre-processing Before model training, the dataset was preprocessed to prepare it for the ML algorithms. Data Cleaning: The dataset was checked for missing values. No missing values were found, so imputation was not necessary.

Feature Scaling: Since algorithms like KNN are sensitive to the scale of features (a feature with a larger range can dominate the distance calculation), we applied StandardScaler from the Scikit-learn library. This standardizes features by removing the mean and scaling to unit variance.

C. Classification Workflow

The core of our methodology is the classification process, which is depicted in the flowchart in Fig. 1.

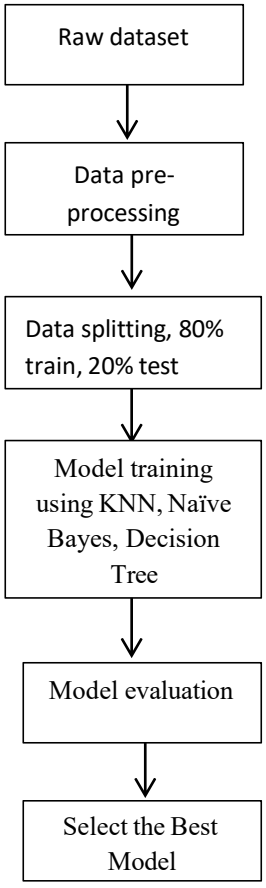

Fig. 1. Flowchart of the Classification Process

D. Machine Learning Algorithms

K-Nearest Neighbours (KNN): KNN is a nonparametric, instance-based learning algorithm. It classifies a new data point based on the majority

class of its 'k' nearest neighbours in the feature space. The "nearness" is typically measured using a distance metric, such as Euclidean distance. For this study, we determined the optimal 'k' value through experimentation.

Naive Bayes: This is a probabilistic classifier based on Bayes' Theorem. It operates on a "naive" assumption of conditional independence between features, meaning it assumes that the presence of one feature does not affect the presence of another, given the class variable. Despite this simplification, it is computationally efficient and performs well in many real-world scenarios.

Decision Tree: A Decision Tree is a supervised learning algorithm that builds a
tree-like model of decisions. It splits the data into smaller subsets based on the values of input features, using criteria like Gini Impurity or Information Gain. The model is highly interpretable, as the decision paths from the root to the leaves can be easily visualized and understood as a set of rules.

E. *Evaluation Metrics*

To evaluate the performance of the classifiers, we used a standard train-test split (80% for training, 20% for testing) and the following metrics:

Accuracy: The ratio of correctly predicted instances to the total instances.

Precision: The ability of the classifier not to label a negative sample as positive.

Recall (Sensitivity): The ability of the classifier to find all the positive samples.

F1-Score: The weighted average of Precision and Recall.

Results and Discussion

The three models were trained and tested using the pre-processed data. The performance of each model was recorded and is summarized in Table I.

TABLE I. PERFORMANCE COMPARISON OF MACHINE LEARNING ALGORITHMS

Algorithm	Accuracy (%)	Precision (%)	Recall (%)	F1Score (%)
K-Nearest Neighbours (KNN)	89.7	89.9	89.7	89.6
Naive Bayes	84.3	85.1	84.3	84.5
Decision Tree	96.6	96.6	96.6	96.6

As evident from Table I, the Decision Tree classifier significantly outperformed both KNN and Naive Bayes across all evaluation metrics, achieving an impressive accuracy of 96.6%.

Discussion:

The superior performance of the Decision Tree can be attributed to several factors. Firstly, Decision Trees are adept at capturing nonlinear relationships and interactions between features, which are prevalent in medical data. For instance, the risk associated with blood pressure might change non-linearly with age. Secondly, the tree-based structure naturally creates a set of explicit rules (e.g., "IF Age > 35 AND Systolic BP > 140 THEN Risk Level = high risk"), which makes the model highly interpretable. This is a critical advantage in a clinical setting, as it allows healthcare providers to understand the reasoning behind a
prediction, fostering trust and facilitating informed decision-making.

The K-Nearest Neighbours model also performed well, with an accuracy of 89.7%. Its performance

relies heavily on the premise that patients with similar physiological profiles will have similar risk levels. However, it can be sensitive to irrelevant features and the curse of dimensionality, which may have slightly hindered its performance compared to the Decision Tree.

The Naive Bayes classifier yielded the lowest accuracy at 84.3%. This is likely due to its core assumption of feature independence. In reality, physiological parameters like SystolicBP, DiastolicBP, and HeartRate are often correlated. The violation of this independence assumption can limit the model's predictive power in complex medical domains.

The implications of these findings are substantial. A highly accurate and interpretable model like the Decision Tree can be integrated into clinical workflows as a screening tool. It can flag high-risk patients for more intensive monitoring or specialized care, optimizing the allocation of healthcare resources and potentially reducing adverse maternal events.

Conclusion

This study successfully demonstrated the application of machine learning for maternal health risk prediction. Through a comparative analysis of K-Nearest Neighbours, Naive Bayes, and Decision Tree classifiers, we established that the Decision Tree model provides the best performance, achieving an accuracy of 96.6%. Its ability to model complex relationships and provide interpretable results makes it an ideal candidate for a clinical decision support system.

While promising, this work has limitations. The model was trained on a specific dataset and its generalizability should be tested on larger, more diverse populations. Future work should focus on:

Exploring more advanced ensemble models like Random Forest and Gradient Boosting, which often build upon the strengths of Decision Trees.

Incorporating a wider range of features, including lifestyle factors, socioeconomic data, and past medical history.

Developing and deploying a user-friendly application to make this predictive tool accessible to healthcare professionals in real world settings.

Ultimately, machine learning stands as a powerful ally in the global effort to improve maternal health outcomes, offering data-driven insights to protect the well-being of mothers everywhere.

Conflicts of Interest

There’s no personal stake involved in putting out this study.

Funding Statement

I didn't get any particular financial support to carry out this study.

Acknowledgement

I would like to sincerely thank my guide **Mrs. Swathi D Mahindrakar** mam for their constant support, guidance, and encouragement throughout this project. Their valuable suggestions and patience helped me learn and complete my work successfully.

I am also grateful to the Dean, **Dr. Shweta Marigoudar** mam and all the faculty members of the Faculty of Computing and Information Technology (FCIT) for providing the facilities, knowledge, and motivation that made this project possible.

Lastly, I want to express my heartfelt thanks to my family and friends for always being there with their

love, support, and encouragement at every stage of this journey.

References

[1] World Health Organization, "Maternal mortality",22-February-2023. https://www.who.int/news-room/factsheets/detail/maternal-mortality

[2] A. S. Panicker and A. P. S. Kumar, "Maternal Risk Prediction Using Machine Learning and Deep Learning," in 2023 7th International Conference on Trends in Electronics and Informatics (ICOEI), 2023.
Link: https://ieeexplore.ieee.org/abstract/document/10126027

[3] A. Akbulut, A. E. Ertugrul, and V. Topcu, "A new maternal health risk dataset for machine learning and its effective utilization," Health Informatics Journal, vol. 29, no. 2, 2023.
Link: https://journals.sagepub.com/doi/full/10.1177/14604582231172545

[4] A. E. Johnson, T. J. Pollard, L. Shen, H. Li-wei, M. Feng, and R. G. Mark, "MIMICIII, a freely accessible critical care database," Scientific Data, vol. 3, 2016. (Often used for such studies, this paper describes the database).
Link: https://www.nature.com/articles/sdata201635

[5] M. R. Devi, S. K. B. S. S, N. R. M, and P. N, "Maternal Health-Risk Prediction using Decision Tree and SVM Algorithms," in 2022 International Conference on Advancements in Electrical, Electronics, Communication, Computing and Automation (ICAECA), 2022. Link: https://ieeexplore.ieee.org/abstract/document/9952219

[6] P. S, S. Malathy, R. S. Sabeenian, and M. P. Ram, "Web Based Application for Maternal Health Care using Random Forest," in 2022 6th International Conference on Devices, Circuits and Systems (ICDCS), 2022.
Link: https://ieeexplore.ieee.org/abstract/document/9788647

[7] Y. Guan, A. M. H. Al-Juboori, A. K. Singh, and M. A. Al-Rubaie, "Predicting fetal distress from cardiotocography recordings using machine learning," Patterns, vol. 4, no. 8, 2023.
Link: https://www.cell.com/patterns/fulltext/S26663899(23)00150-5

[8] M. Goyal, D. Singh, A. K. Singh, and A. K. S. Bhati, "Machine Learning-Based Approach for Maternal and Fetal Health Prediction," in 2023 5th International Conference on Inventive Research in Computing Applications (ICIRCA), 2023.
Link: https://ieeexplore.ieee.org/abstract/document/10221356

[9] S. Ahmed, A. E. E. Emon, S. T. R. Shah, and V. Topcu, "Maternal Health Risk Data Set," UCI Machine Learning Repository, 2020.
Link: https://archive.ics.uci.edu/dataset/863/maternal+health+risk+data+set

[10] H. Koivu, S. Sairanen, M. Gissler, P. Stefanovic, and P. K. Heinonen, "Machine learning reveals complex risk factor patterns for severe maternal morbidity," ActaObstetricia et GynecologicaScandinavica, vol. 101, no. 6, pp. 691-699, 2022.
Link: https://obgyn.onlinelibrary.wiley.com/doi/full/10.1111/aogs.14364

[11] S. D. M. Swathi, "Significance of Early Disease Detection in Arecanut using Convolutional Neural Network," *JNRID International Open Access Journal Peer Review*, 2024.