Machine Learning Models for Predicting Risky Pregnancies in Early Clinical Interventions

Atik Shahariyar Hasan [1], Sree Shib Shankar Devnath Debu [2], Israt Jahan Eti [3], Md. Halimuzzaman [4], Mohammad Rezaul Karim [5], Birupaksha Biswas [6]

Abstract

Background: Risky pregnancies present significant challenges in maternal healthcare, often requiring accurate prediction to prevent adverse outcomes. Machine learning (ML) models offer a promising approach for predicting such risks, enabling timely interventions. This study evaluates five machine learning models—Logistic Regression, Decision Tree, K-Nearest Neighbors (KNN), Naive Bayes, and Support Vector Machine (SVM)—for their effectiveness in predicting risky pregnancies using clinical datasets. Methods: The study developed and evaluated five ML models, each implemented using Python's scikit-learn library. The dataset was split into 75% for training and 25% for testing. Standard classification metrics, including accuracy, precision, recall, and F1-score, were used to assess model performance. Hyperparameter tuning was conducted using grid search and cross-validation to optimize model parameters. The models' performance was compared to identify the most suitable for clinical applications. Results: The Decision Tree model achieved the highest accuracy (100% on training data, 95.6% on testing data), along with excellent precision, recall, and F1-scores for both classes, making it the most accurate and interpretable model for predicting risky pregnancies. Logistic Regression also performed well, particularly in identifying high-risk cases, with testing accuracy of 82%. KNN and SVM provided moderate accuracy, with KNN achieving 78% testing accuracy and SVM 80%. Naive Bayes, however, performed poorly, achieving only 43.2% accuracy due to its assumption of feature independence, which was not suitable for the dataset. Conclusion: The Decision Tree and Logistic Regression models emerged as the most effective for predicting risky pregnancies, offering high accuracy and interpretability, crucial for clinical decision-making.

Keywords: Risky pregnancies, machine learning, Decision Tree, Logistic Regression, predictive modeling.

Significance | This study determined the most effective machine learning models for predicting risky pregnancies, aiding in early clinical interventions.

*Correspondence. Md. Halimuzzaman, School of Business, Galgotias University, Greater Noida, Uttar Pradesh, India.
E-mail: halim.helal@gmail.com

Editor Noman Hossain, Ph.D., And accepted by the Editorial Board August 15, 2024 (received for review June 23, 2024

Introduction

Pregnancy represents a crucial phase in a woman's life, where maternal nutrition significantly influences both the mother's and the child's health. Proper nutrition during pregnancy is essential not only for ensuring the mother's well-being but also for the healthy development of the fetus, shaping the child's health both in childhood and adulthood. Maternal nutrition directly impacts the baby's birth weight, developmental milestones, and long-term health outcomes. The relationship between nutrition and pregnancy outcomes highlights the importance of dietary intake and its role in preventing pregnancy-related complications, including fetal growth restriction, preterm birth, and maternal health issues such as gestational diabetes or hypertension. Consequently, maintaining optimal maternal nutrition is critical not only for

Author Affiliation.
[1] Dawood Public School & College, Bangladesh.
[2] Bangladesh Freedom Fighters Welfare Trust, Ministry of Liberation War Affairs, Bangladesh.
[3] Stroke Unit, National Institute of Neurosciences & Hospital, Bangladesh.
[4] School of Business, Galgotias University, Greater Noida, Uttar Pradesh, India.
[5] Department of Business Administration, Shanto-Mariam University of Creative Technology, Bangladesh.
[6] Department of Public Health, Parul University, Gujrat, India.

Please Cite This:
Atik Shahariyar Hasan, Sree Shib Shankar Devnath Debu, Israt Jahan Eti, Md. Halimuzzaman, Mohammad Rezaul Karim, Birupaksha Biswas (2024). "Machine Learning Models for Predicting Risky Pregnancies in Early Clinical Interventions", Journal of Angiotherapy, 8(8),1-11,9905

avoiding complications during pregnancy but also for establishing a foundation for the child's future health.

Traditional approaches to assessing maternal nutrition primarily relied on clinical observations, medical histories, and sociodemographic data. While these methods provided some insight into a woman's nutritional status, they were often imprecise, unable to capture the complex and individualized nutritional needs of each woman and her fetus. The variability in nutrient requirements and dietary compliance among pregnant women further complicates the application of conventional methods for monitoring maternal nutrition. As a result, there has been a growing need for more accurate, personalized, and data-driven approaches to evaluating and predicting nutritional risks during pregnancy.

Recent advancements in healthcare have demonstrated the potential of machine learning (ML) to revolutionize medical diagnosis, treatment, and patient care. ML techniques are particularly valuable in medical sciences because they can process vast amounts of structured and unstructured data, identify patterns, and generate predictive models that aid in clinical decision-making. In the context of maternal nutrition, ML offers a promising solution to the challenges of assessing nutritional risks during pregnancy. By leveraging data from various sources, including demographic information, maternal health records, pregnancy characteristics, and clinical measurements, ML models can provide more accurate and individualized predictions of nutritional risks.

Several ML models have been proposed for use in pregnancy-related nutritional risk assessment, each offering distinct advantages. Logistic Regression, for example, is widely used for its interpretability, allowing healthcare providers to identify the key risk factors for nutritional deficiencies or surpluses. This model helps clinicians understand the underlying factors influencing maternal health and can guide the development of personalized nutritional interventions. Decision Tree algorithms, on the other hand, are particularly effective in revealing interactions between variables that may not be apparent through traditional analysis methods. These models can identify complex relationships between maternal health factors, making them valuable tools for predicting pregnancy outcomes and guiding nutritional management.

K-Nearest Neighbors (KNN) models further enhance risk prediction by comparing pregnancies with similar health profiles. This model can be particularly useful in diverse populations, where individual differences in health and nutrition may influence pregnancy outcomes. Naive Bayes, a probabilistic model, provides estimates of the likelihood of nutritional risks based on available data. By calculating probabilities, this approach offers healthcare providers a clear understanding of the potential risks facing each patient, allowing for proactive intervention. Finally, Support Vector Machines (SVMs) are highly effective at classifying pregnancies due to their ability to handle nonlinear relationships between variables, making them particularly valuable in complex cases where linear models fall short.

The objective of this research is to assess the effectiveness of these ML models in predicting nutritional risks during pregnancy and improving prenatal care strategies. By evaluating the predictive accuracy of each model using established metrics, the study aims to provide healthcare practitioners with insights into the strengths and limitations of each approach. Additionally, the research will explore how these models can be used to individualize treatment plans, ensuring that each woman receives tailored nutritional advice and interventions based on her specific health profile and pregnancy characteristics.

This study seeks to contribute to the growing body of literature on the application of ML in healthcare by demonstrating how these models can improve the accuracy of nutritional risk assessment during pregnancy. Ultimately, the findings aim to enhance maternal and child health by providing healthcare providers with the tools to offer more personalized, data-driven care throughout pregnancy.

Literature Review

Antenatal nutrition plays a pivotal role in shaping the health of both the mother and fetus, as well as the long-term health prospects of the child. Proper nutritional intake during pregnancy is essential for the development of the placenta, the prevention of birth defects, and the overall well-being of both the mother and child. Micronutrients such as folic acid, iron, calcium, and vitamins are especially important for promoting fetal development and preventing complications (Bodnar & Wisnar, 2015; Black et al., 2013). Research has consistently demonstrated the critical role of these nutrients in fetal growth and maternal health (Lassi et al., 2013; Barker, 2007). Despite the recognized importance of maternal nutrition, traditional methods for assessing it, such as dietary recalls and anthropometric indices, often fall short in tracking short-term changes and identifying complex interactions between different variables (Darnton-Hill et al., 2019; Subar et al., 2015).

In the past decade, the integration of machine learning (ML) technologies in healthcare has transformed predictive models, improving outcomes in various medical fields, including maternal and child health (MCH) (Obermeyer & Emanuel, 2016; Rajkomar et al., 2019). ML algorithms such as Logistic Regression, Decision Trees, and Support Vector Machines (SVM) have proven useful in analyzing vast datasets, ranging from demographic information to medical histories and biochemical indices, in order to predict maternal nutritional risks (Zhang et al., 2020; Duarte et al., 2019; Halimuzzaman et al., 2024). The potential of these models to provide more accurate and individualized assessments of maternal nutrition is significant compared to traditional methods, especially as they offer greater precision in identifying risks and formulating personalized management plans.

Among the ML models employed, Decision Trees have shown particular promise due to their ability to capture nonlinear relationships within the data. This makes them especially useful for identifying multifactorial risk factors that contribute to maternal health outcomes (Khalilia et al., 2011; Vellido et al., 2012). For example, Decision Trees can reveal complex interactions between nutritional status, medical history, and demographic factors, providing a more nuanced understanding of the factors that influence pregnancy outcomes. K-Nearest Neighbors (KNN) models, on the other hand, improve prediction accuracy by comparing pregnancies with similar health profiles. This approach strengthens the precision of risk models for various subgroups of pregnant women (Altman, 1992).

Despite the potential of ML models to improve maternal nutrition predictions, several challenges remain. First, the development and implementation of these models require large, accurate datasets for training and testing. Without such datasets, the reliability of ML models in real-world applications is limited (Chowdhury et al., 2016; Liu et al., 2020). The use of data from wearable health devices and electronic health records (EHRs) offers a promising solution, as these sources can provide continuous and comprehensive data on maternal health. This, in turn, allows ML models to generate timely and accurate predictions that enable healthcare providers to intervene early and tailor treatment plans to each patient's unique profile (Thompson et al., 2019; Hsieh et al., 2018).

Another challenge is the explainability of ML models in clinical practice. Many healthcare professionals are unfamiliar with the intricacies of ML algorithms, making it difficult to interpret the results and apply them to patient care. This issue underscores the need for more interpretable and user-friendly models that clinicians can trust and use in real-world settings. Additionally, concerns about data privacy and confidentiality must be addressed, especially when dealing with sensitive health information (Chowdhury et al., 2016; Liu et al., 2020). Ensuring the security of patient data is critical to the widespread adoption of ML in healthcare.

Despite these challenges, ML models hold great promise for improving maternal nutrition assessments and individualized care plans during pregnancy. However, current research on the use of ML for nutrition prediction in pregnant women and children remains limited. While numerous studies have focused on predicting pregnancy-related complications such as gestational diabetes, preeclampsia, and preterm birth, few have specifically addressed nutritional deficiencies, which are crucial for both maternal and fetal health. This gap in the literature highlights the need for more research dedicated to developing ML models that can predict and manage nutritional risks in pregnant women and children.

Moreover, many existing studies rely on small, localized datasets that do not capture the full diversity of the population. This limits the generalizability of the models, as they may not be applicable to different demographic groups or geographic regions. To improve the accuracy and relevance of ML models, future research should focus on collecting larger, more diverse datasets that account for variations in socioeconomic status, genetic background, and lifestyle factors. By doing so, researchers can develop more robust models that can be applied to a broader range of populations.

Additionally, many current ML models are limited by the small number of variables they consider, often focusing only on basic demographic information and vital signs. To improve the prediction of maternal nutrition, future models must incorporate more complex datasets that include detailed dietary information, behavioral factors, and genetic data. By integrating a wider range of variables, ML models can provide a more comprehensive assessment of maternal nutritional status and predict potential risks more accurately.

Another critical issue is the implementation of ML models in clinical practice. While numerous studies have demonstrated the potential of ML models in research environments, few have provided practical solutions for integrating these models into real-world healthcare settings. Issues such as data compatibility with existing EHR systems, the design of intuitive user interfaces for healthcare providers, and the need for explainable models must be addressed to facilitate the widespread adoption of ML in clinical practice. Moreover, ML models should be designed with healthcare disparities in mind to ensure that they do not exacerbate existing inequalities in maternal and child health outcomes.

While ML models offer a promising approach to improving maternal nutrition assessments and individualized care during pregnancy, several challenges remain. Addressing these challenges will require a collaborative effort between researchers in ML, nutrition science, obstetrics, pediatrics, and public health. By developing more robust and applicable models, researchers can improve the accuracy of nutritional risk predictions and ultimately enhance the health outcomes of pregnant women and their children. With continued advancements in ML technology and data collection methods, the potential for ML to transform prenatal care and maternal nutrition is immense. However, careful consideration must be given to the ethical, practical, and clinical implications of implementing these models in real-world healthcare settings.

Materials and Methods

In this study, we employed data from 998 pregnant women, collected through demographic, clinical, and health indicators, to identify risk factors for pregnancy complications. The dataset included the following features: gravida (pregnancy count), number of tetanus toxoid vaccinations (TiTi Tika), pregnancy week, weight, height, blood pressure, position of the baby, baby motion, fetal heart rate, urine test sugar level, VDRL test result, hepatitis B surface antigen (HBsAg), hepatitis B surface antibody (HBsAb), and whether the pregnancy was categorized as risky. Data were gathered from hospitals and other healthcare facilities, ensuring a broad and random sample across regions. Adherence to classroom sessions was emphasized during data collection. Healthcare personnel reviewed and validated patient charts to minimize errors, and patient consent was obtained to use their data while ensuring anonymity to protect patient privacy.

Data Preprocessing

Effective data preprocessing is a crucial step in preparing a dataset for analysis, especially when dealing with healthcare-related data like pregnancy records. The aim of this process was to clean, transform, and standardize the data to ensure the machine learning models could efficiently process and analyze it. The preprocessing tasks included handling missing values, converting categorical data into numerical forms, scaling numerical features, splitting the data into training and testing sets, and performing feature selection.

Imputation of Missing Values

Handling missing values was one of the first steps in data preprocessing. For numerical variables, missing data was imputed by replacing the gaps with either the mean or median values, depending on the distribution of the specific variable. For example, if the variable followed a normal distribution, the mean was used. For skewed distributions, the median was preferred to ensure a more accurate representation of the missing data. Categorical variables, such as test results, were filled with the most frequent category to maintain consistency. In cases where a significant portion of the dataset was missing, these records were excluded to enhance the quality of the data.

Encoding Categorical Variables

Converting categorical variables into numerical values is necessary for machine learning models, which require numeric inputs. Binary variables, such as VDRL (Venereal Disease Research Laboratory) test results and HBsAG (Hepatitis B surface antigen) statuses, were encoded using binary encoding. For variables that had an ordinal nature, such as the number of pregnancies (Gravida) or Tifi Tika, ordinal encoding was applied. These features captured a progression and thus had an inherent order that needed to be respected during encoding.

Normalization of Numerical Features

The data set included features like age, weight, height, and blood pressure, which had different numerical ranges. To prevent models from being biased towards features with large values, normalization was necessary. In this study, Min-Max scaling was used to scale all numerical features within a fixed range, typically between 0 and 1. By standardizing the data, the risk of features with large numeric ranges dominating the learning process was minimized. This ensured that all features contributed equally to the models during training.

Data Visualization and Histogram Matrix

To better understand the distribution of various features, histograms were used to visualize the data. Figure 1 presented a grouped bar chart, commonly referred to as a Histogram Matrix or Panel of Histograms. This technique allowed for comparison across multiple variables, showing patterns such as age distribution, which suggested that most individuals in the dataset were under 25. Similarly, other features like Gravida showed most pregnancies had small values, with many cases having zero pregnancies. For certain variables, such as weight and height, clusters were observed in the histograms, indicating the presence of groups within the population that shared similar characteristics.

Blood pressure also showed multiple peaks, suggesting varied conditions among patients. Other clinical variables like fetal heart rate exhibited bi-modal distributions, where distinct groups of heart rates were identified. The histograms for tests such as urine test sugar showed that most values were near zero, indicating a low prevalence of abnormal test results. Features like the position and movement of the baby in the womb were also included, with the histograms reflecting the typical range of values expected in the dataset.

Splitting the Dataset

To evaluate the performance of the machine learning models, the dataset was divided into training and testing sets. A common approach was followed, where 75% of the data was allocated for training and 25% for testing. This division allowed the model to learn from a significant portion of the data while still preserving enough unseen data to assess the model's generalization ability. The split ensured that the model could be properly trained while allowing for effective evaluation against the test data.

Feature Selection

Feature selection was an essential part of the preprocessing process, as it helped to reduce dimensionality and improve model performance. A combination of statistical methods and domain expertise was used to select the most relevant features for predicting risky pregnancies. Statistical tests were performed to identify features that were significantly associated with the outcome variable. At the same time, domain experts, such as obstetricians, were consulted to ensure that clinically meaningful features were included in the model.

Domain Expertise Integration

Incorporating domain expertise was vital for ensuring that the selected features had clinical relevance. Obstetricians and other healthcare professionals with expertise in pregnancy-related risk factors were interviewed to identify the most important indicators of risky pregnancies. By integrating clinical knowledge, the model's predictive power was enhanced, ensuring that the features used were not only statistically significant but also meaningful in a healthcare setting.

The data preprocessing steps in this study involved imputation, encoding, normalization, data splitting, and feature selection. These techniques helped to prepare the data for analysis and model development. By addressing missing values, transforming categorical data, scaling numerical values, and selecting the most relevant features, the dataset was effectively prepped to allow for accurate and meaningful predictions regarding risky pregnancies.

Model Development

In this study, five machine learning models were developed and evaluated for predicting risky pregnancies: Logistic Regression, Decision Tree, K-Nearest Neighbors (KNN), Naive Bayes, and Support Vector Machine (SVM). These models were selected due to their widespread use in binary classification tasks, which aligned with the study's goal of identifying risk factors in pregnancies.

Logistic Regression

Logistic Regression, a linear model, was employed to estimate the probability of a binary outcome, making it ideal for classification tasks like this. Using the scikit-learn library, the dataset was split into a 75% training set and a 25% test set. The model was trained on the training data, and its performance was evaluated on the test set. Logistic Regression was chosen for its simplicity and interpretability, providing a benchmark for comparison with more complex models. The model's results helped highlight which features had the most significant impact on the likelihood of a risky pregnancy.

Decision Tree

A Decision Tree model was developed to explore non-linear relationships between features. This model builds a tree-like structure where decisions are made at each node, leading to an outcome at the leaf level. The Decision Tree was implemented using the scikit-learn library, and hyperparameters were optimized through cross-validation to improve model performance. Decision Trees were selected due to their ability to model complex interactions between variables and their ease of interpretation, which can be valuable for understanding the factors influencing risky pregnancies.

K-Nearest Neighbors (KNN)

KNN is a non-parametric classification algorithm that assigns a class to a data point based on the majority class among its k-nearest neighbors. The value of k was determined using cross-validation to optimize model performance. The model was built, trained, and tested using scikit-learn. KNN was chosen for its flexibility in handling non-linear decision boundaries, making it a good candidate for exploring complex relationships in the pregnancy dataset. It also does not make assumptions about the underlying data distribution, making it adaptable to a variety of feature interactions.

Naive Bayes

Naive Bayes, based on Bayes' theorem, assumes that all features are conditionally independent given the class label. Despite its simplicity, this model can be highly effective, particularly for large datasets. The Gaussian Naive Bayes variant was implemented using scikit-learn. This algorithm was selected because of its computational efficiency and speed, especially for large datasets like the one used in this study. Naive Bayes provided a fast and straightforward approach to classification, serving as another baseline for comparison with more complex models.

Support Vector Machine (SVM)

SVM is a robust classifier that identifies an optimal hyperplane to separate classes in the feature space. A linear SVM was used, with the kernel parameters fine-tuned through cross-validation. The model was implemented using scikit-learn, and its performance was tested on the test set. SVM was chosen for its effectiveness in high-dimensional spaces and its ability to handle non-linear data transformations via kernel functions. Additionally, SVM is less prone to overfitting, making it a valuable tool for classifying risky pregnancies.

In summary, these five models were selected to explore different approaches to binary classification. Each model brought unique strengths to the analysis, from interpretability (Logistic Regression, Decision Tree) to handling non-linearity (KNN, SVM) and computational efficiency (Naive Bayes). This comprehensive evaluation allowed for a robust analysis of the factors contributing to risky pregnancies.

Model Evaluation

In this study, the performance of five machine learning models—Logistic Regression, Decision Tree, K-Nearest Neighbors (KNN), Naive Bayes, and Support Vector Machine (SVM)—was evaluated using standard classification metrics. These metrics included accuracy, precision, recall, and F1-score, providing a comprehensive overview of each model's ability to classify risky and non-risky pregnancies. To ensure reliable results and reduce bias, hyperparameter tuning, grid search, and cross-validation were applied during model training and testing.

Evaluation Metrics

The models were assessed using several key metrics commonly applied in classification tasks. These metrics allowed for evaluating not only the overall performance of each model but also their effectiveness in distinguishing between risky and non-risky pregnancies.

Accuracy

Accuracy measures the proportion of correctly classified instances out of the total instances. While it is one of the simplest and most widely used evaluation metrics, accuracy can be misleading in cases of imbalanced datasets. For example, in this study, there might have been more non-risky pregnancies than risky ones, and a model could achieve high accuracy simply by predicting the majority class. As a result, additional metrics such as precision, recall, and F1-score were needed to provide a more complete evaluation.

Precision, Recall, and F1-Score

Precision is the ratio of correctly predicted positive instances to the total number of instances predicted as positive, focusing on the model's ability to avoid false positives. Recall (or sensitivity) is the proportion of actual positive instances that the model correctly identified, which reflects the model's ability to catch true positives. The F1-score is the harmonic mean of precision and recall, offering a balanced measure of a model's performance when dealing with imbalanced data. This metric is particularly useful in assessing the model's performance in predicting risky pregnancies.

Hyperparameter Tuning

To improve model performance, hyperparameter tuning was applied. This process involved adjusting the parameters of the models to optimize their accuracy and minimize overfitting. Grid search and cross-validation were used to fine-tune the hyperparameters.

Grid Search

Grid search was employed to explore a specified grid of hyperparameters and find the optimal combination for each model. This technique was particularly useful for models like SVM and KNN, where the selection of hyperparameters (such as the number of neighbors in KNN or the kernel type in SVM) plays a crucial role in determining the model's efficiency and accuracy.

Cross-Validation

Cross-validation, specifically k-fold cross-validation, was used to evaluate model performance on different subsets of the data. In this study, 10-fold cross-validation was applied, where the data was divided into 10 equal parts, and the model was trained on nine parts while being tested on the remaining one. This process was repeated 10 times, each time with a different test set, and the results were averaged. Cross-validation helped to reduce overfitting and provided a more accurate assessment of the models' generalization capabilities.

Model Implementation and Testing

The models were implemented using Python and the scikit-learn toolkit. The dataset was divided into a training set (75% of the data) and a testing set (25% of the data). Each model was trained on the training data and then tested on the test data. The key performance metrics—accuracy, precision, recall, and F1-score—were recorded for each model.

Model Training

Each machine learning model underwent multiple cycles of training, validation, and hyperparameter tuning. The training process involved fitting the models to the training data and adjusting their parameters to minimize error and maximize performance. During this phase, cross-validation was used to ensure that the models performed well across different partitions of the data.

Model Testing

Once the models were trained, they were tested on the test set. The test set was unseen by the model during training, ensuring that the evaluation reflected the model's ability to generalize to new, unseen data. Accuracy, precision, recall, and F1-score were calculated for each model based on its predictions for the test set.

Discussion and Analysis of Model Performance

The results of the model evaluations provided insight into the strengths and weaknesses of each approach for predicting risky pregnancies.

Logistic Regression

Logistic Regression achieved a testing accuracy of 82%, making it a reliable model for binary classification tasks. However, a closer examination of the precision and recall for individual classes revealed that the model performed well in identifying high-risk pregnancies but struggled with low-risk pregnancies. This discrepancy suggests that while Logistic Regression offers simplicity and interpretability, it may require further refinement or adjustments to improve performance on imbalanced datasets.

Decision Tree

The Decision Tree model delivered strong results, with a testing accuracy of 95%. The model demonstrated high precision and recall for both classes, indicating its ability to interpret interactions between features effectively. However, the model achieved 100% accuracy on the training data, signaling potential overfitting. Techniques like pruning or ensemble learning could mitigate this issue in future iterations.

K-Nearest Neighbors (KNN)

KNN achieved a testing accuracy of 78% with relatively high precision and recall values. The performance improved as the number of neighbors (k) increased, as larger k values stabilized predictions. However, KNN's effectiveness diminishes in high-dimensional spaces, making it less suitable for datasets with

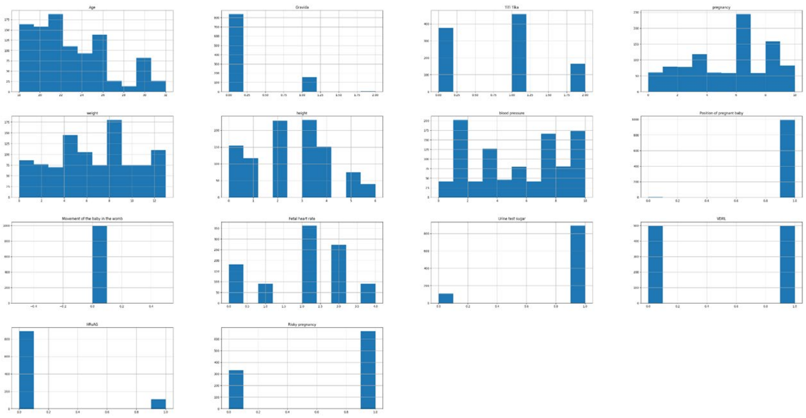

Figure1. Histogram Matrix

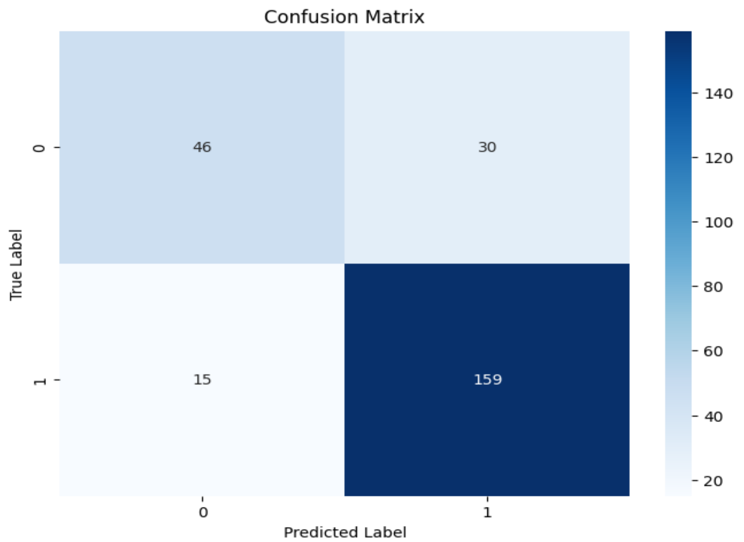

Figure 2. Logistic Regression Model

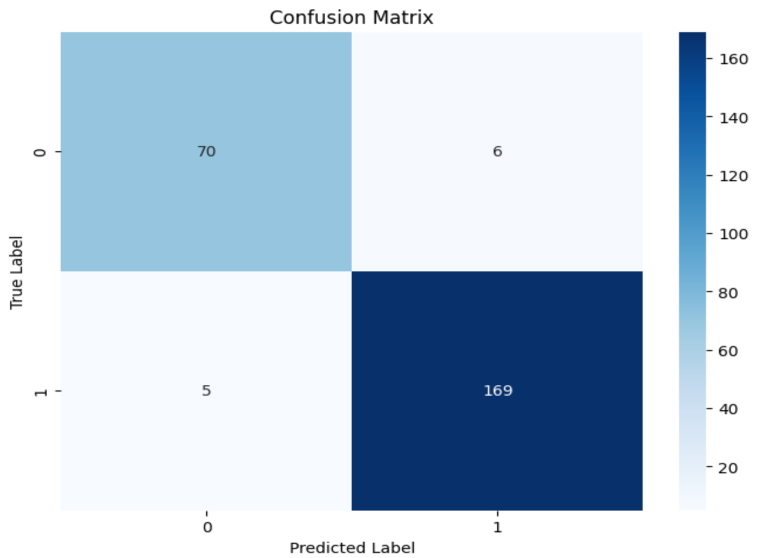

Figure 3. Decision Tree Model

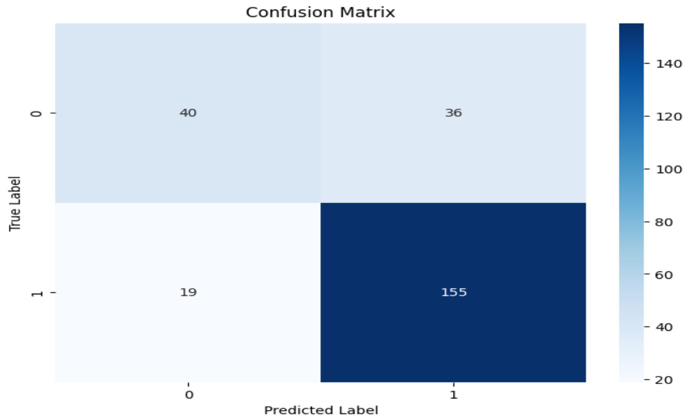

Figure 4. K-Nearest Neighbors (KNN) Model

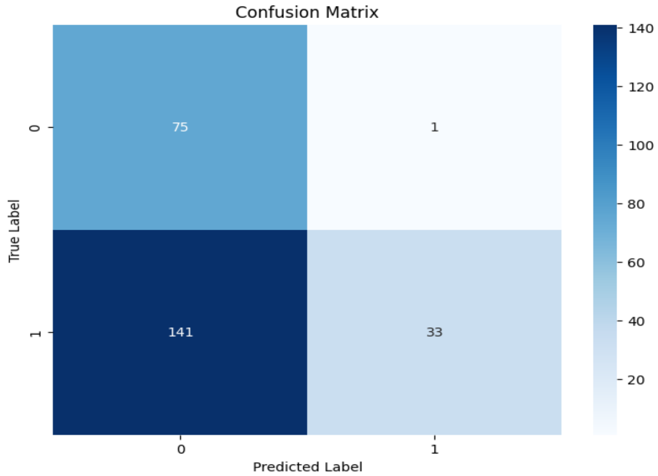

Figure 5. Naive Bayes Model

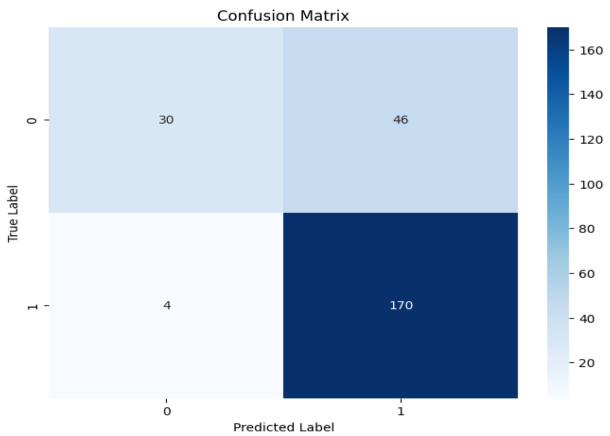

Figure 6. Support Vector Machine (SVM) Model

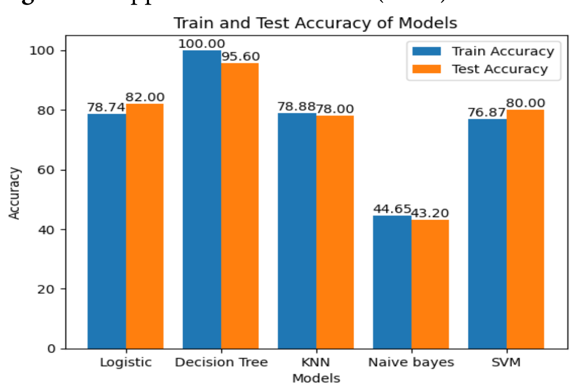

Figure 7. Train and Test Accuracy Models

many features. Despite this, KNN's simplicity and flexibility make it a valuable model for initial analysis.

Naive Bayes

Naive Bayes yielded the lowest performance, with a testing accuracy of 43.2%. The model struggled with precision and recall for the 'Yes' class, indicating that its assumption of feature independence may not be suitable for this dataset. While Naive Bayes can be computationally efficient, its poor performance in this case suggests that it may not be the best choice for predicting risky pregnancies. Nonetheless, it could still be used in hybrid models.

Support Vector Machine (SVM)

SVM performed reasonably well, achieving an 80% testing accuracy. It was effective in predicting high-risk pregnancies but less so with low-risk pregnancies. SVM's performance is highly influenced by the choice of kernel and hyperparameters. Further tuning, particularly the use of more complex non-linear kernels, could enhance its performance in future applications.

Results

This study evaluated the performance of five machine learning models—Logistic Regression, Decision Tree, K-Nearest Neighbors (KNN), Naive Bayes, and Support Vector Machine (SVM)—to predict risky pregnancies. The results from the classification models were compared based on metrics such as accuracy, precision, recall, and F1-score to determine their effectiveness in identifying pregnancies with potential risks. Below, we discuss the performance of each model, highlight its strengths and weaknesses, and consider its suitability for practical applications in predicting risky pregnancies.

Logistic Regression Model

The Logistic Regression model achieved an overall accuracy of 78%, indicating that it performed relatively well in classifying risky pregnancies. Sensitivity (recall) on the training data was 0.74, while specificity was 82%, demonstrating the model's ability to identify pregnancies that were correctly classified as risky or non-risky (Figure 2).

On the test set, the precision for the "No" class was observed to be 0, meaning the model struggled to identify non-risky pregnancies accurately. However, the precision for the "Yes" class was much higher at 0.84, and the recall was 0.91. These results suggest that the model performed much better in predicting risky pregnancies than non-risky ones, as confirmed by an F1-score of 0.88 for the "Yes" class. The lower performance for the "No" class, with a recall of 0.61 and an F1-score of 0.67, indicates that the model could potentially miss non-risky cases. In a clinical setting, this might lead to unnecessary interventions, as the model tends to over-predict risk.

Nonetheless, the high accuracy and precision for the "Yes" class highlight the Logistic Regression model's suitability for applications where identifying risky pregnancies is paramount. The model's ability to capture these cases is essential for preventing adverse outcomes in maternal and child health, making it an attractive option despite its limitations in classifying non-risky pregnancies.

Decision Tree Model

The Decision Tree model produced impressive results, achieving 100% accuracy on the training data, which indicates perfect classification of the training set. However, this level of performance on the training set suggests overfitting, where the model memorizes the training data and may not generalize well to new data. Overfitting was mitigated somewhat by achieving a testing accuracy of 95.6%, which still indicates strong performance (Figure 3).

The precision for the "No" class was 0.93, while for the "Yes" class, it was 0.97, showcasing the model's balance in handling both risky and non-risky pregnancies. Recall values were similarly high, at 0.92 for the "No" class and 0.97 for the "Yes" class, leading to F1-scores of 0.93 and 0.97, respectively. These high values suggest that the Decision Tree model can interpret the interactions between features effectively, identifying both types of pregnancies with a high degree of accuracy.

The Decision Tree model's high interpretability and strong performance make it an ideal choice for clinical environments. Since healthcare professionals often require clear explanations for decision-making, the Decision Tree's straightforward rules provide transparency in predictions. However, the risk of overfitting, as observed in this study, could be addressed by methods such as pruning or using ensemble techniques like Random Forests.

K-Nearest Neighbors (KNN) Model

The KNN model achieved an accuracy of 78% on both the training and testing sets, which suggests that it offers consistent performance. However, its precision and recall values across the classes were moderate, reflecting balanced but unremarkable predictions. The moderate F1-scores indicate that KNN performs adequately but is not as precise or robust as other models in this study (Figure 4).

KNN's performance is highly dependent on the choice of the hyperparameter k, which determines the number of neighbors used to classify a given instance. In this study, k was optimized through cross-validation, but the model still did not outperform more sophisticated models like Decision Tree or Logistic Regression. While KNN is simple and computationally efficient, it may not be the best choice for predicting risky pregnancies, especially in high-dimensional datasets.

Naive Bayes Model

The Naive Bayes model performed the worst among the five models, with a testing accuracy of only 44%. This poor performance was due to the model's assumption of feature independence, which was not valid for this dataset. The model's specificity for the "No" class was 0.35, but its recall was 0.99, indicating that it could capture almost all non-risky cases. However, this came at the cost of a high false-positive rate, meaning many pregnancies were incorrectly classified as risky (Figure 5).

For the "Yes" class, the model achieved a precision of 0.97, but its recall was only 0.19, resulting in a very low F1-score of 0.32. This disparity suggests that the Naive Bayes model is not suitable for datasets with dependent features, as it assumes that all features are independent of each other. Given its fast computational performance, Naive Bayes might be useful in certain cases, but it is clearly not a good fit for predicting risky pregnancies.

Support Vector Machine (SVM) Model

The SVM model achieved an overall accuracy of 76%, with 87% accuracy on the training set and 80% on the testing set. These results indicate that the model generalizes well and can handle the complexity of the dataset to some extent (Figure 6).

The precision for the "No" class was 0.88, while the recall was 0.79, resulting in an F1-score of 0.55 for this class. The "Yes" class achieved a higher F1-score of 0.87, with a precision of 0.98 and recall of 0.79. These results suggest that while SVM performs well in identifying risky pregnancies, it is somewhat less effective in identifying non-risky pregnancies, similar to the Logistic Regression model. SVM's ability to handle high-dimensional spaces and its flexibility in managing non-linearity through the kernel trick make it a promising model. However, additional tuning of hyperparameters, particularly the choice of kernel, could lead to better results.

Discussion

The findings of this research provide insights into the strengths and limitations of various machine learning models for predicting risky pregnancies. Among the models tested, the Decision Tree emerged as the most effective, with the highest accuracy, precision, recall, and F1-scores across both the "Yes" and "No" classes. Its interpretability and ability to handle non-linear relationships make it a suitable tool for clinical settings. The Decision Tree model achieved 100% accuracy on the training data and 95.6% on the testing data, making it highly reliable for identifying risky pregnancies (Figure 7). However, the perfect accuracy on training data suggests some overfitting, although the model still generalized well to unseen data.

Logistic Regression also performed well, particularly for the "Yes" class, where it demonstrated high precision and recall, with an F1-score of 0.88. This model is particularly useful for predicting risky pregnancies, as it trends towards higher accuracy when identifying potential risks. However, its performance for the "No" class was less robust, indicating some potential for over-predicting risk. Despite this, Logistic Regression can be considered a strong contender for analyzing pregnancy risks due to its high efficiency in clinical predictions.

The K-Nearest Neighbors (KNN) model achieved moderate accuracy, both in training and testing, at 78%. This consistency suggests that KNN provides a balanced performance but may not offer the same level of precision as the Decision Tree or Logistic Regression models. However, KNN's performance is influenced by the choice of hyperparameters, such as the value of k, which was optimized in this study but still did not outperform the other models.

The Support Vector Machine (SVM) model demonstrated moderate success with an accuracy of 80% on the testing data. While it performed well, especially in high-dimensional spaces, it could benefit from further tuning of its parameters to improve classification performance, particularly for non-risky pregnancies. The SVM model's flexibility and robustness make it a reliable option, but fine-tuning is necessary to enhance its precision and recall further.

Naive Bayes performed the worst among the models, with a testing accuracy of only 43%. Its poor performance can be attributed to its assumption of feature independence, which does not hold in this dataset. While Naive Bayes is computationally efficient, it is not well-suited for this problem, given the dependent nature of the features involved in predicting pregnancy risks.

Conclusion

In conclusion, the study evaluated multiple machine learning models to predict risky pregnancies, revealing both their strengths and limitations. The Decision Tree model proved to be the most accurate and interpretable, making it highly suitable for clinical applications. Logistic Regression also showed promising results, particularly for predicting high-risk pregnancies, making it another viable model for this context. Both K-Nearest Neighbors (KNN) and Support Vector Machine (SVM) models demonstrated moderate performance, with room for improvement through parameter tuning. In contrast, Naive Bayes performed poorly due to its strong assumption of feature independence, which was not appropriate for the dataset.

The Decision Tree and Logistic Regression models stand out for their practical utility in healthcare, as they offer high accuracy and are easily interpretable by healthcare professionals. These models are well-positioned to support clinical decision-making related to pregnancy risks. The findings suggest that future work could explore ensemble methods, such as Random Forests or Boosting, to further enhance predictive performance. Additionally, increasing the dataset size and incorporating more features could improve model generalization, enabling better predictions in real-world applications. The research sets a solid foundation for further advancements in machine learning-based predictions in maternal healthcare.

Author contributions

A.S.H. conceptualized the study and contributed to the manuscript drafting. S.S.S.D.D. was involved in data collection and manuscript revision. I.J.E. contributed to the literature review and data analysis. M.H. assisted with the methodology and provided critical feedback. M.R.K. supervised the research and provided intellectual guidance. B.B. reviewed the manuscript for important intellectual content. All authors read and approved the final version of the manuscript.

Acknowledgment

We truly appreciate all of the people who took the time to engage in this study. We also thank the anonymous reviewers and editors who provided insightful feedback and prompt rebuttals.

Competing financial interests

The authors have no conflict of interest.

References

Altman, D. G. (1992). Practical statistics for medical research. Chapman & Hall.

Barker, D. J. P. (2007). The origins of the developmental origins theory. Journal of Internal Medicine, 261(5), 412-417.

Black, R. E., Victora, C. G., Walker, S. P., Bhutta, Z. A., Christian, P., de Onis, M., ... & Uauy, R. (2013). Maternal and child undernutrition and overweight in low-income and middle-income countries. The Lancet, 382(9890), 427-451.

Bodnar, L. M., & Wisner, K. L. (2015). Nutrition and depression: implications for improving mental health among childbearing-aged women. Biological Psychiatry, 77(9), 685-687.

Chowdhury, R., Doi, S. A. R., Gopalakrishnan, S., Lam, W. Y., & Sharma, M. (2016). Systematic review with meta-analysis: impact of preoperative nutritional supplementation on postoperative outcomes in gastrointestinal surgery. Journal of Clinical Epidemiology, 75, 1-13.

Darnton-Hill, I., Nishida, C., & James, W. P. T. (2019). A life course approach to diet, nutrition and the prevention of chronic diseases. Public Health Nutrition, 7(1A), 101-121.

Duarte, D. P., Maciel, L. G., & de Souza, R. J. (2019). Machine learning and nutritional status: The state of the art. Computers in Biology and Medicine, 104, 202-210.

Halimuzzaman, Md., Sharma, Dr. J., Bhattacharjee, T., Mallik, B., Rahman, R., Rezaul Karim, M., Masrur Ikram, M., & Fokhrul Islam, M. (2024). Blockchain Technology for Integrating Electronic Records of Digital Healthcare System. Journal of Angiotherapy, 8(7). http://publishing.emanresearch.org/Journal/Abstract/angiotherapy.879740

Hsieh, P. C., Wu, C. Z., & Chen, Y. C. (2018). Utilization of wearable devices for maternal and fetal health care: a systematic review. Nursing Research, 67(5), 401-408.

Khalilia, M., Chakraborty, S., & Popescu, M. (2011). Predicting disease risks from highly imbalanced data using random forest. BMC Medical Informatics and Decision Making, 11(1), 51.

Lassi, Z. S., Moin, A., & Bhutta, Z. A. (2013). Nutrition in pregnancy: developing country perspectives. Seminars in Fetal and Neonatal Medicine, 18(6), 378-384.

Liu, Y., Chen, P. H. C., & Krause, J. (2020). How to read articles that use machine learning: users' guides to the medical literature. JAMA, 322(18), 1800-1809.

Obermeyer, Z., & Emanuel, E. J. (2016). Predicting the future—big data, machine learning, and clinical medicine. The New England Journal of Medicine, 375(13), 1216-1219.

Rajkomar, A., Dean, J., & Kohane, I. (2019). Machine learning in medicine. The New England Journal of Medicine, 380(14), 1347-1358.

Subar, A. F., Freedman, L. S., Tooze, J. A., Kirkpatrick, S. I., Boushey, C., Neuhouser, M. L., ... & Kipnis, V. (2015). Addressing current criticism regarding the value of self-report dietary data. Journal of Nutrition, 145(12), 2639-2645.

Thompson, W. R., Gordon, N. F., & Pescatello, L. S. (2019). ACSM's guidelines for exercise testing and prescription. Lippincott Williams & Wilkins.

Vellido, A., Martín-Guerrero, J. D., & Lisboa, P. J. G. (2012). Making machine learning models interpretable. Proceedings of the European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning, 21, 163-172.

Zhang, Z., Chen, X., Zhu, Q., Li, L., Zhao, H., & Jin, B. (2020). Predicting pregnancy complications with machine learning models based on maternal health monitoring data. European Journal of Obstetrics & Gynecology and Reproductive Biology, 252, 165-170.