Available online at website: https://jurnal.iaii.or.id/index.php/RESTI

JURNAL RESTI
(Rekayasa Sistem dan Teknologi Informasi)
Vol. 10 No. 4 (2026) 1197 - 1206 e-ISSN: 2580-0760

Predicting High-Risk Pregnancies Using Machine Learning Algorithms and Explainable AI

Sari Puspita[1], Gusrino Yanto[2]*
[1,2]Information System, Faculty of Information Technology and Creative Industries, Universitas Metamedia, Padang, Indonesia

*gusrino@metamedia.ac.id

Abstract

Pregnancies classified as high-risk play a significant role in increasing health complications and deaths among mothers and newborns, making early risk identification essential for timely clinical intervention. Although machine learning has shown promising performance in predicting pregnancy risk, most existing studies rely on binary classification and provide limited model interpretability. This study proposes an interpretable multiclass machine learning framework that predicts pregnancy risk using maternal health records from primary healthcare facilities. A total of 2,553 maternal medical records collected from five community health centers in Koto Tangah District, Padang City, Indonesia, were analyzed. The proposed framework integrates data preprocessing, StandardScaler, Synthetic Minority Over-sampling Technique (SMOTE), GridSearchCV-based hyperparameter optimization, and Shapley Additive exPlanations (SHAP). Four machine learning algorithms, which are Logistic Regression, Decision Tree, Support Vector Machine, and Random Forest, were systematically assessed with the application of Accuracy, Precision, Recall, F1-score, and ROC-AUC. Of the machine learning models considered, Random Forest performed best, achieving 97.26% accuracy, 90.44% precision, 98.06% recall, 93.65% F1-score, and 99.79% ROC-AUC. SHAP analysis identified Heart Rate, Blood Glucose, Diastolic Blood Pressure, and Systolic Blood Pressure as the most influential predictors, while also improving model transparency through feature contribution and interaction analysis. The results indicate that the suggested framework delivers precise and interpretable multiclass pregnancy risk prediction, demonstrating how it can assist with the early detection of pregnancy risks within primary care environments.

Keywords: explainable artificial intelligence; high-risk pregnancy; machine learning; random forest; SHAP

How to Cite: S. . Puspita and G. Y. Yanto, "Predicting High-Risk Pregnancies Using Machine Learning Algorithms and Explainable AI", *J. RESTI (Rekayasa Sist. Teknol. Inf.)* , vol. 10, no. 4, pp. 1197–1206, Aug. 2026, doi: 10.29207/resti.v10i4.7071.
*Permalink/*DOI: https://doi.org/10.29207/resti.v10i4.7071

Received: August 11, 2025
Accepted: August 5, 2026
Available Online: August 27, 2026

This is an open-access article under the CC BY 4.0 License
Published by Ikatan Ahli Informatika Indonesia

1. Introduction

High-risk pregnancies are those affected by certain conditions or circumstances that elevate the likelihood of significant health problems for both the mother and fetus throughout gestational period, childbirth, and after delivery. Such pregnancies demand increased monitoring and specialized care to reduce the chance of life-threatening outcomes [1], [2]. In 2020, pregnancy- and childbirth-related complications claimed the lives of nearly 800 women each day, according to the World Health Organization (WHO). Additionally, an estimated between 20% and 30% of pregnancies are categorized as high-risk, leading to 70-80% of perinatal deaths and illnesses [3], [4]. Maternal deaths are usually caused by complications like severe bleeding, pregnancy-induced hypertension, infections, unsafe abortions, and chronic diseases that worsen during pregnancy [5], [6]. A variety of factors can contribute to an elevated risk, including medical history, advanced age, poor lifestyle habits, socioeconomic background, and ongoing pregnancy complications [7-9]. Given the complexity of pregnancy and the potential for complications, preventive measures are necessary to protect both mothers and fetuses [10].

Ensuring maternal and fetal health throughout pregnancy remains a major challenge because pregnancy-related risks may change dynamically over time [11-13]. Pregnancy risks are dynamic, and a mother's condition, which was initially normal, can suddenly become high risk [14]. Because any risk during pregnancy can endanger both the mother and the fetus, high-risk pregnancies require intensive medical monitoring and treatment throughout the pregnancy and after childbirth.

Earlier research indicates that applying machine learning methods can improve the accuracy of pregnancy risk prediction. Ensemble methods, including bagged trees and quad-ensemble models, have demonstrated effectiveness in identifying high-risk pregnancies [15], [16]. In particular, tree-based algorithms such as Random Forest excel at capturing complex, nonlinear patterns within healthcare data.

However, most existing studies remain limited to binary classification frameworks and rarely incorporate techniques from Explainable Artificial Intelligence (XAI) to improve the clarity of model operations. Consequently, multiclass prediction models based on primary healthcare data remain limited, particularly in Indonesia.

Although previous studies have reported promising predictive performance, several important limitations remain. First, most studies still rely on binary classification approaches that do not adequately represent real-world clinical risk stratification. Second, interpretability methods such as SHAP have not yet been fully integrated into predictive modelling frameworks through comprehensive quantitative analysis, resulting in models that prioritize predictive accuracy over transparency. Third, empirical evidence from primary healthcare data in developing countries, including Indonesia, remains limited, reducing the generalizability and practical applicability of existing models. Therefore, there is a need for a multiclass machine learning framework that not only delivers precise predictions but also offers clear and understandable in supporting clinical decisions.

One of the main achievements in the present work is creating an interpretable multiclass machine learning framework using maternal health records collected from community health centers (Puskesmas) to represent primary healthcare settings. The presented framework incorporates SHAP analysis to clarify the contribution made by each predictor clinical variable and enables a comparative assessment of different machine learning models. This integration enhances model interpretability while maintaining strong predictive performance, thereby demonstrating potential for supporting maternal healthcare applications in Indonesia. In Indonesia, early detection of high-risk pregnancies is a key strategy to reduce maternal mortality and pregnancy-related complications, as emphasized in Law No. 36 of 2009 on health development. Local health authorities, including the Padang City Health Office, prioritize maternal health monitoring as part of efforts to improve public health outcomes. Nevertheless, maternal mortality associated with pregnancy, childbirth, and the postpartum period continues to occur annually in Padang City. According to operational guidelines, approximately 20% of pregnancies are categorized as high-risk, highlighting the need for more effective prediction methods to support timely intervention and referral in primary healthcare settings.

Accordingly, the present study sought to create and examine an interpretable multiclass machine learning system to predict pregnancy risk based on maternal health information collected from community health centers in Koto Tangah District, Padang City. The proposed framework combines StandardScaler, SMOTE, GridSearchCV, and SHAP-based explainability to improve predictive performance while providing transparent clinical interpretation.

2. Methods

2.1 Dataset and Research Variables

This study utilized maternal medical records collected from five community health centers in Koto Tangah District, Padang City, Indonesia. The initial dataset consisted of 2,679 records. After data quality assessment, all duplicate entries were excluded, yielding 2,553 distinct records for use in model development and evaluation.

The predictor variables included Age, Systolic Blood Pressure, Diastolic Blood Pressure, Blood Glucose, Body Temperature, and Heart Rate. The target variable was Risk Level, which was categorized into three classes: low, moderate, and high risk. Table 1 provides an in-depth explanation of the variables examined in this research.

Table 1. Variables Employed Within The Research

Variable	Type	Description
Age	Integer	Age of the pregnant woman
Systolic Blood Pressure	Float	Systolic Blood Pressure measurement
Diastolic Blood Pressure	Float	Diastolic Blood Pressure measurement
Blood Glucose	Float	Blood Glucose measurement
Body Temperature	Float	Body Temperature measurement
Heart Rate	Float	Heart Rate measurement
Risk Level	Categorical	Low, Moderate, High

2.2 Research Workflow

The research followed a systematic machine learning workflow which involved cleaning the data, data preprocessing, exploratory data analysis (EDA), train-test splitting, class imbalance handling using the Synthetic Minority Over-sampling Technique (SMOTE), hyperparameter optimization using GridSearchCV, model training, performance evaluation, choosing the top-performing model and interpreting it with Shapley Additive exPlanations (SHAP). An overview of the research process is provided in Figure 1. The workflow was designed to ensure that each stage was performed sequentially and that test-set information was excluded during the model training process or hyperparameter optimization. The training dataset was used for preprocessing parameter estimation, class balancing, cross-validation, and model optimization, while the test set was retained for final performance evaluation.

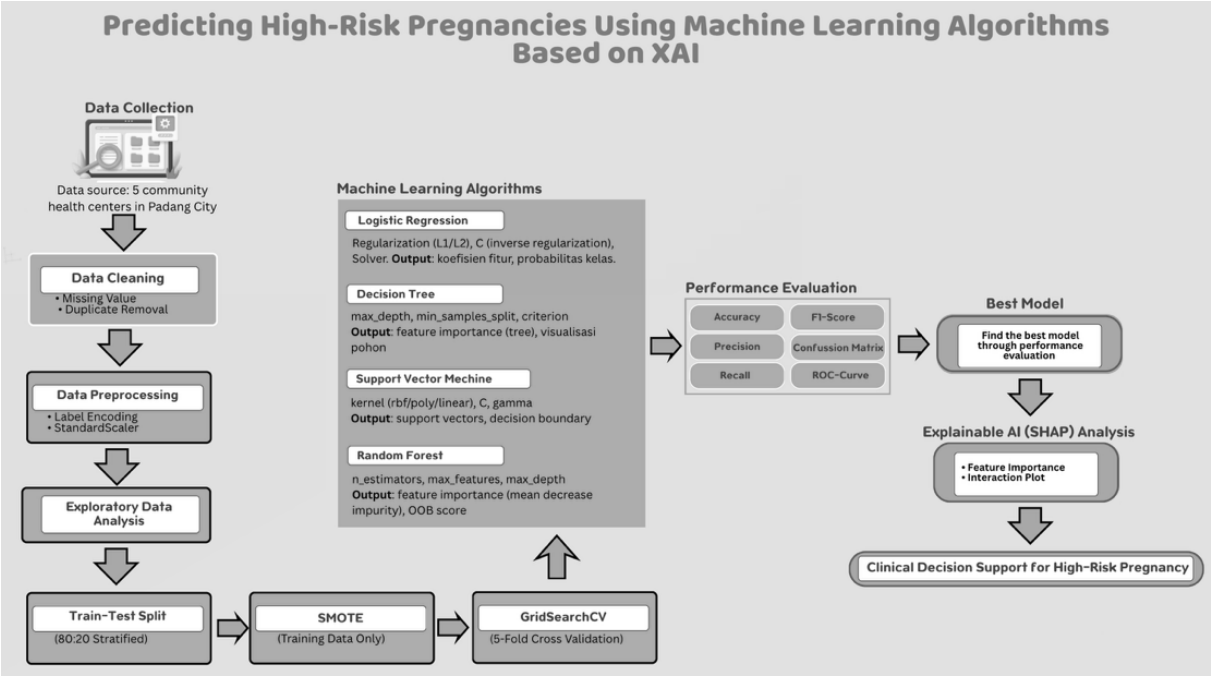


Figure 1. Workflow of the Proposed Machine Learning Framework

2.3 Data Preprocessing

The data preprocessing phase was undertaken to enhance the overall quality of the dataset and ensure its suitability for subsequent machine learning applications. An initial examination of the dataset confirmed that no missing values were present. Duplicate records were removed to ensure that each observation was unique. The categorical target variable was subsequently encoded into numerical labels to facilitate model training.

The dataset was subsequently divided into training and testing subsets to facilitate model development and evaluation before feature scaling and class balancing. Numerical features were standardized with StandardScaler [17], [18]. To mitigate the potential for data leakage, the StandardScaler was fitted exclusively on the training data before being used to transform both the training and test datasets.

2.4 Class Imbalance Handling

The training dataset exhibited an uneven distribution across the three pregnancy risk categories, with the low-risk class having substantially fewer observations than the other groups. Class balancing using SMOTE was performed exclusively on the training set after the dataset had been divided into training and testing subsets to address class imbalance[19], [20]. SMOTE creates new examples for minority classes by leveraging current minority-class instances, thereby increasing their representation during model training.

The test set did not undergo SMOTE processing and maintained its initial class distribution. This procedure was applied to avoid information leakage and guarantee that model performance was assessed using data that remained representative of the original observations [21-23]. Following class balancing, the resulting training set was used to train the models and optimize their hyperparameters.

2.5 Machine Learning Classification Models

This study compared four widely used classification algorithms: Logistic Regression (LR), Decision Tree (DT), Support Vector Machine (SVM), and Random Forest (RF) [24-32]. The selected algorithms were chosen to cover various classification approaches, such as linear models, tree-based methods, margin-based classifiers, and ensemble techniques [33-36].

Following preprocessing and class balancing, each algorithm was trained using the resulting training dataset. The use of multiple classification approaches enabled a comparative evaluation of their predictive capabilities in distinguishing among low-, moderate-, and high-risk pregnancy categories [37], [38].

2.6 Hyperparameter Optimization

GridSearchCV was utilized for hyperparameter tuning of each model using five-fold cross-validation within the data used for model training. The search evaluated multiple combinations of hyperparameter values systematically, with the configuration producing the maximum mean cross-validation accuracy determined to be the optimal setting for each algorithm [39], [40].

The test dataset was not involved in hyperparameter optimization and was reserved solely to assess the optimized models using the independent test set. This approach helps minimize the possibility of overfitting to the test data and offers greater reliability estimate of the models' generalization performance.

2.7 Model Performance Evaluation

Each classification model's performance was assessed by calculating accuracy, precision, recall, F1-score, and ROC-AUC metrics [41], [42]. These metrics were used to provide complementary perspectives in terms of the model's ability tomake accurate predictions across the three pregnancy risk categories.

Additionally, confusion matrices were analyzed to investigate errors in classifying the three categories of pregnancy risk [43]. Using the One-vs-Rest (OvR) approach, the ROC curve and corresponding AUC were calculated to assess the model's discriminative performance across the multiple classes [44]. The best-performing model based on the established evaluation metrics was selected for further interpretation using SHAP.

2.8 Model Interpretation Using SHAP

The model achieving the best results was then examined with Shapley Additive exPlanations (SHAP) to enhance understanding of its prediction results [45], [46]. SHAP analysis was performed to quantify the contribution of individual predictor variables to the model output and determine their relative importance in pregnancy risk classification..

Two analyses utilizing SHAP were conducted to help interpret the predicted outcomes generated by the model. SHAP Feature Importance helped determine and rank the predictor variables according to the degree of their influence on the model's predictions. Furthermore, SHAP Interaction Plots were applied to explore how predictor variables interacted with each other during the prediction process [47-49].

3. Results and Discussion

3.1 Results

This section reports results obtained at each stage of the proposed machine learning framework, beginning with data preprocessing and ending with model interpretation using SHAP.

Data preprocessing ensured data quality before machine learning model development. The initial dataset consisted of 2,679 maternal medical records collected from five community health centers in Koto Tangah District, Padang City. An initial review of the data found it to be complete, with no missing entries identified, eliminating the need for imputation. Afterward, 126 duplicate entries were detected and eliminated to reduce redundancy and avoid introducing bias into the model training process. After data cleaning, a total of 2,553 unique records remained and were used throughout all stages of the study.

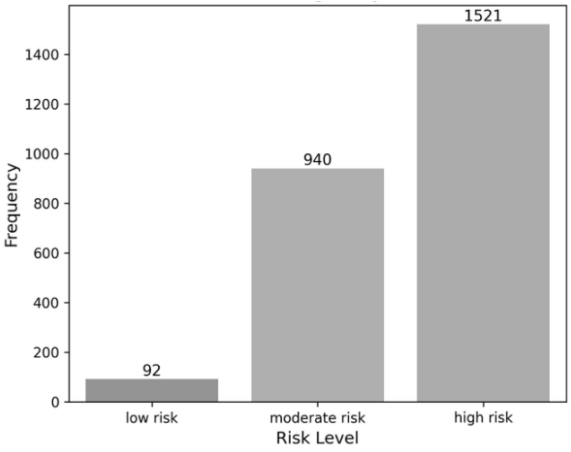


Figure 2. The Distribution of Pregnancy Risk Levels

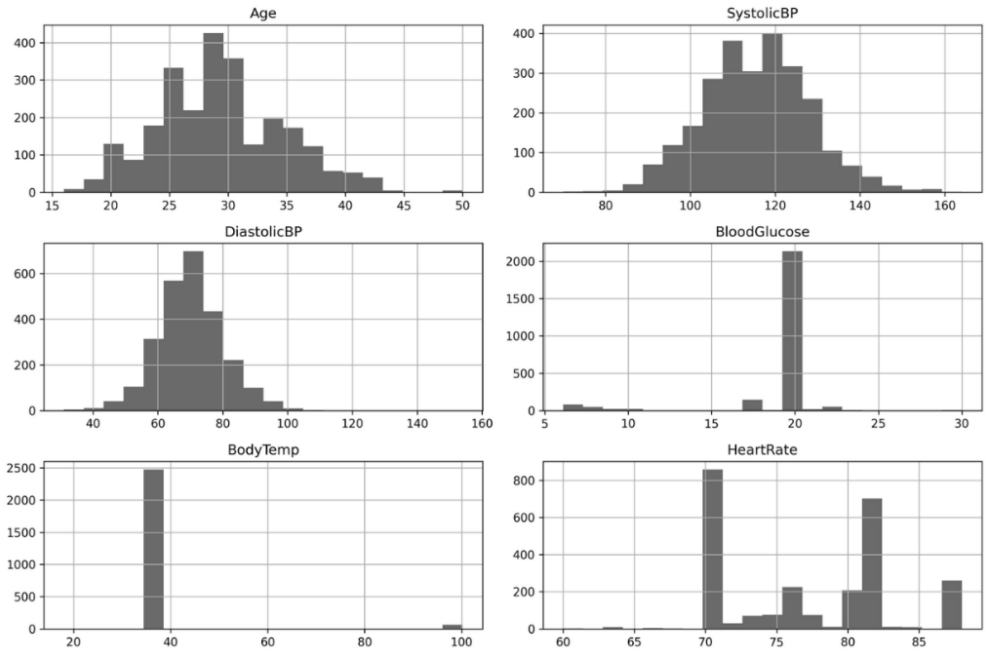


Figure 3. Distribution of the Numerical Variables in the Maternal Health Dataset

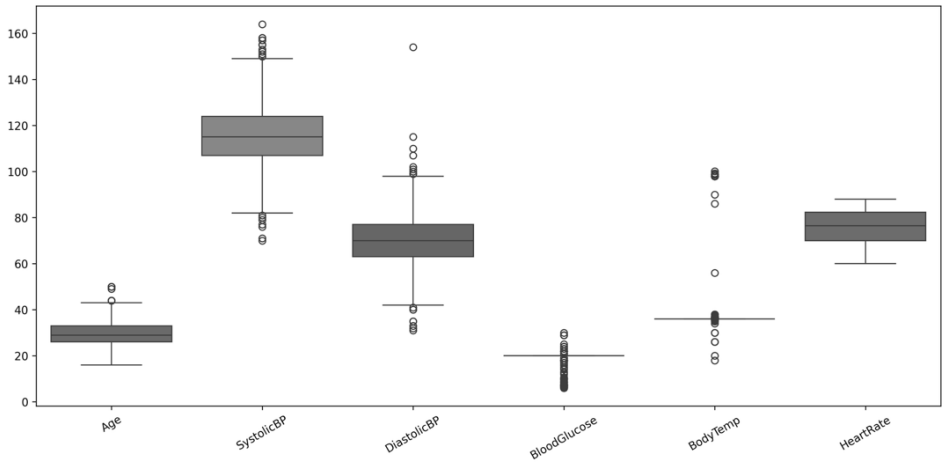


Figure 4. Boxplots of the Numerical Predictor Variables in the Maternal Health Dataset

The processed dataset was further examined using exploratory data analysis (EDA) to identify its key characteristics and assess the distribution of samples among the respective classes. Figure 2 illustrates the distribution of pregnancy risk classes after data cleaning. The dataset remained imbalanced, with the high-risk class comprising the largest proportion of samples and the low-risk class the smallest. This kind of imbalance can skew the learning process in favor of the majority class, which may decrease the model's effectiveness at accurately identifying instances from the minority class.

Figure 3 presents the histograms of the numerical variables used in this study. Analysis of the processed data identified varying distribution characteristics among the measured variables, namely age, systolic and diastolic blood pressure, blood glucose level, body temperature, and heart rate. Visual inspection of the histograms revealed no obvious outliers or implausible values, indicating that the dataset was suitable for subsequent model development.

Figure 4 displays boxplots for the numerical predictor variables. A number of these variables, especially blood glucose and systolic blood pressure, show indications of possible outliers. These observations were retained because they were considered plausible values rather than measurement errors and therefore could represent genuine clinical variations within the study population.

Figure 5 presents the correlation matrix for the predictor variables included in this research. The strongest correlation was identified between systolic pressure and diastolic pressure (r = 0.61), indicating a moderate positive relationship. This finding is consistent with the physiological relationship between these two cardiovascular measurements.

The remaining variable pairs exhibited relatively weak correlations, providing no evidence of strong pairwise relationships indicative of severe multicollinearity among the predictor variables. Consequently, all predictor variables were retained, as each was considered to provide complementary information for developing the machine learning models.

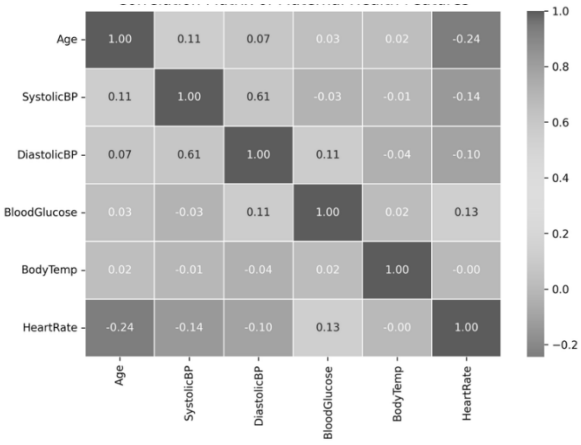


Figure 5. Correlation Matrix of the Numerical Predictor Variables

Class imbalance in the cleaned dataset was resolved by applying SMOTE exclusively to the training set during the model development process. Figure 6 illustrates the change in class distribution in the resampled training data generated using SMOTE.

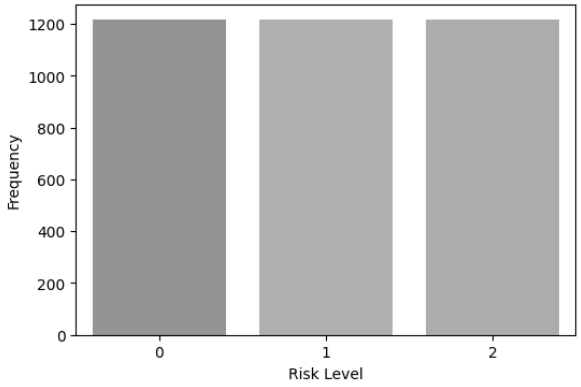


Figure 6. Distribution of Training Samples Across Classes After SMOTE Resampling

Prior to oversampling, the training data showed a clear imbalance, as the low-risk category had significantly fewer samples in comparison to the moderate- and high-risk categories. The use of SMOTE led to a balanced

sample count across all classes by creating artificial instances for the minority class.

Balancing the training data helped reduce majority-class bias and ensured more equitable representation across the three classes during model training. Since SMOTE was restricted to the training set, the original distribution of the test data remained intact, which helped prevent data leakage and maintained an objective evaluation of the model.

The balanced training set was then used to develop four machine learning models. Hyperparameter tuning was carried out through GridSearchCV to identify the best parameter settings for each algorithm.

Four machine learning algorithms- Logistic Regression (LR), Decision Tree (DT), Support Vector Machine (SVM), and Random Forest (RF)- were developed to categorize pregnancy risk into three categories: low, moderate, and high risk. For each algorithm, GridSearchCV evaluated candidate hyperparameter configurations using five-fold cross-validation to determine the optimal setting.

Table 2. Optimized Hyperparameter

Model	Optimal Hyperparameters
Logistic Regression	C = 0.1, solver = lbfgs
Decision Tree	max_depth = 15, min_samples_split = 2
Support Vector Machine	C = 10, kernel = RBF
Random Forest	n_estimators = 200, max_depth = 20, min_samples_split = 2

Table 2 summarizes the optimal hyperparameter configurations obtained for each classification algorithm. The optimization process produced different parameter settings for each model, reflecting the distinct learning characteristics of the algorithms. Random Forest achieved its optimal configuration with 200 decision trees, a maximum tree depth of 20, and a minimum split size of two samples. Meanwhile, the Support Vector Machine achieved its best performance using an RBF kernel with a regularization parameter of C = 10. The classification models were trained with the hyperparameter settings obtained through optimization. An independent test set was used for the final evaluation of the optimized models, providing an assessment of their generalization to previously unseen data. To evaluate the model's ability to generalize to previously unseen data, five metrics—accuracy, precision, recall, F1-score, and ROC-AUC—were calculated for each of the four machine learning models. A comparison of their results is shown in Table 3.

Table 3 demonstrates that Random Forest achieved the most favorable results across the evaluated performance metrics, attaining 97.26% accuracy, 90.44% precision, 98.06% recall, 93.65% F1-score, and a ROC-AUC value of 99.79%. While Decision Tree and Support Vector Machine also yielded strong results, Logistic Regression had the lowest performance among the models assessed. Random Forest's higher predictive performance could be related to its effectiveness in learning intricate relationships and patterns embedded in the clinical dataset.

Table 3. Comparative Analysis of Machine Learning Model Performance

Model	Accuracy	Precision	Recall	F1-score	ROC-AUC
Logistic Regression	0.9335	0.8129	0.9540	0.8590	0.9869
Decision Tree	0.9628	0.8693	0.9055	0.8859	0.9436
Support Vector Machine	0.9609	0.8543	0.9546	0.8921	0.9952
Random Forest	0.9726	0.9044	0.9806	0.9365	0.9979

While aggregate evaluation metrics offer a general indication of model performance, they may not capture errors occurring within individual classes. To provide a more detailed assessment, confusion matrices were examined to determine how accurately each model classified instances into the pregnancy-risk categories. The resulting confusion matrices for Logistic Regression, Decision Tree, Support Vector Machine, and Random Forest are shown in Figure 7. Overall, all four models correctly classified the majority of samples; however, the frequency and distribution of misclassifications varied across algorithms. Compared with the other approaches, Random Forest provided the most accurate classification, successfully classifying the greatest number of cases while minimizing incorrect predictions. Most prediction errors occurred between the moderate and high-risk classes, suggesting these categories share similar clinical characteristics and are therefore harder to distinguish. In contrast, Random Forest correctly classified all low-risk samples, indicating that this class exhibited more distinctive feature patterns. Logistic Regression produced the most classification errors, particularly within the moderate-risk class, whereas the Decision Tree and Support Vector Machine achieved intermediate performance. These findings align with the quantitative evaluation metrics in Table 3, where Random Forest achieved the highest Accuracy, F1-score, and ROC-AUC across the models assessed. The information obtained from confusion matrices is primarily limited to class-level classification errors and does not adequately reflect model discrimination when the decision threshold is varied. Model discriminative performance was further assessed through Receiver Operating Characteristic (ROC) curve analysis and the resulting Area Under the Curve (AUC) values. To examine how effectively the selected classifier distinguishes among the three classes, the One-vs-Rest (OvR) approach was used to generate the ROC curves. The ROC curves for the three pregnancy risk categories are presented in Figure 8. As depicted in Figure 8, the ROC curves for the three pregnancy-risk categories closely approach the upper-left region, reflecting the model's strong discriminative ability across classes. The Random Forest classifier obtained AUC values of 0.998 for high-risk, 0.998 for low-risk, and 0.997 for moderate-risk cases, with an overall ROC-AUC of 0.9979.

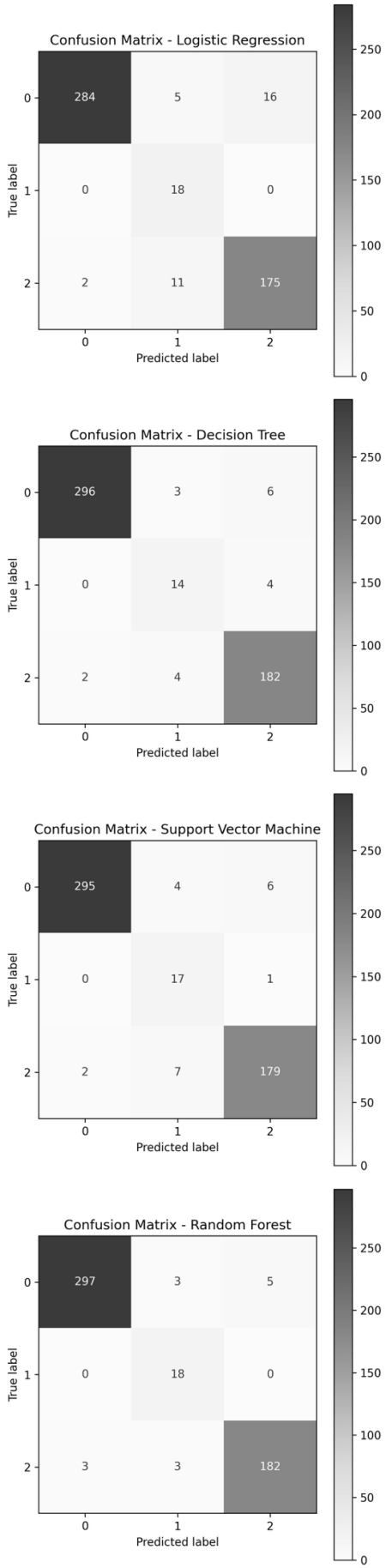


Figure 7. Confusion Matrices of the Four Machine Learning Models

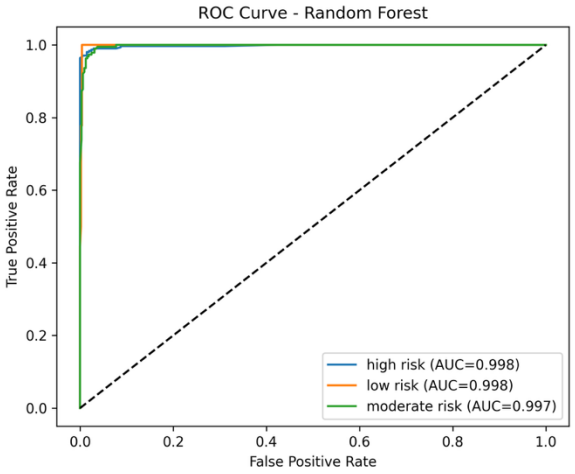


Figure 8. ROC Curves of the Random Forest Model for Multiclass Pregnancy Risk Classification

The analysis results suggest that the classifier maintained consistently high discriminative capability across all three risk categories over a range of classification thresholds.

Given its superior predictive results, Random Forest was chosen for further model interpretation using Explainable Artificial Intelligence (XAI) techniques. SHAP was then used to quantify the influence of individual predictor variables and investigate their contributions to the classification process. Figure 9 illustrates the global feature importance derived from the mean absolute SHAP values. The findings suggest that Heart Rate contributed the most to the model output, followed by Blood Glucose, Diastolic Blood Pressure, Systolic Blood Pressure, Body Temperature, and Age. This ranking indicates that cardiovascular and metabolic measurements played a greater role in determining the model predictions than the remaining variables. Overall, the SHAP analysis improved model interpretability by providing a quantitative measure of each predictor's contribution to the generated predictions.

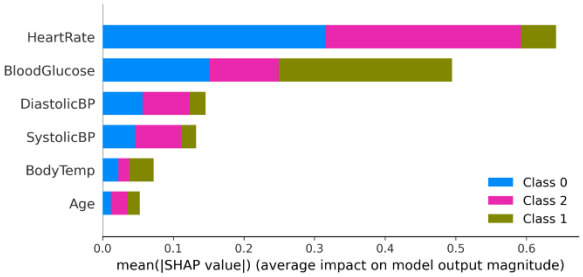


Figure 9. Global SHAP Feature Importance of the Random Forest Model

To further examine the relationships among predictor variables, SHAP interaction analysis was performed, as illustrated in Figure 10. Most SHAP interaction values were concentrated around zero, indicating that the Random Forest model's predictions were primarily influenced by the individual effect of each variable. The most notable interaction was identified between

Systolic Blood Pressure and Diastolic Blood Pressure, suggesting that these variables jointly contributed to distinguishing pregnancy risk levels. Overall, the SHAP findings suggest that the Random Forest model provides both accurate and interpretable predictions.

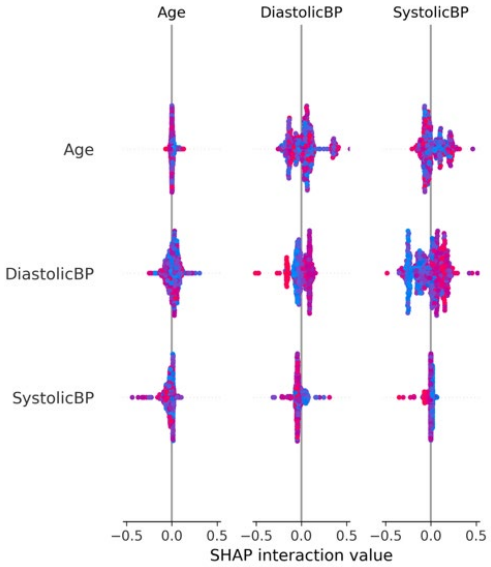


Figure 10. SHAP Interaction Plot of the Random Forest Model

3.2 Discussion

Among the evaluated classifiers, Random Forest emerged as the best-performing model compared with the other classification algorithms, as reflected by its 97.26% accuracy, 93.65% F1-score, and 99.79% ROC-AUC. The observed performance suggests that ensemble learning approaches may be well suited to the accurate classification of multiple pregnancy risk categories using standard maternal health indicators. Compared to Logistic Regression, Decision Tree, and Support Vector Machine, Random Forest demonstrated better generalization, likely because its ensemble mechanism captures complex nonlinear relationships among maternal clinical variables while reducing prediction variance.

The performance advantage of Random Forest is supported by prior research emphasizing the potential of ensemble methods in maternal health prediction. Pavagada and Vemuri [15] demonstrated the applicability of machine learning approaches for detecting maternal health risks, whereas Khadidos et al. [16] reported that ensemble-based techniques can achieve improved predictive results when classifying pregnancy risk levels. Unlike these previous studies, which mainly focused on predictive accuracy, the present study integrates multiclass classification with SHAP-based explainability, enabling the model to provide both accurate predictions and transparent explanations that support clinical decision-making.

Beyond model performance, the proposed framework emphasizes interpretability through SHAP analysis. Heart Rate, Blood Glucose, Diastolic Blood Pressure, and Systolic Blood Pressure were identified as the predictors contributing most strongly to the model's pregnancy risk predictions. These results are clinically meaningful because cardiovascular and metabolic conditions are widely recognized as important indicators of maternal complications during pregnancy. By quantifying the contribution of individual predictors to model outputs, SHAP provides an interpretable basis for understanding individual predictions. Such explanations may help healthcare professionals assess AI-generated predictions and support informed decision-making in clinical settings.

The proposed framework also has practical value for primary healthcare because it was developed using maternal health records collected from community health centers. The proposed model could facilitate early pregnancy risk assessment while assisting healthcare practitioners with referral prioritization and clinical decision-making, especially in primary care settings.

This study extends existing research by combining data preprocessing, hyperparameter optimization, multiclass classification techniques, and Explainable Artificial Intelligence (XAI) within a single predictive framework. The integrated approach achieves high predictive capability while preserving model interpretability, thereby addressing an important limitation of previous studies on pregnancy risk prediction. Nevertheless, validation using independent datasets collected from different healthcare facilities is required to determine whether the model can be generalized to broader clinical settings. Future studies may also investigate advanced boosting approaches, including XGBoost, LightGBM, and CatBoost, to examine their potential to enhance predictive performance and improve the generalizability of the proposed framework.

3.3 Limitations of the Study

The interpretation of these findings requires consideration of several study limitations. One important limitation is the source of the dataset, which was derived from five community health centers in Koto Tangah District. Consequently, the model's applicability may be limited when applied to different populations or healthcare settings. Second, the model was developed using only six clinical variables and did not include potentially relevant predictors such as gestational age, body mass index, medical history, or laboratory test results. Third, the authors did not perform external validation with independent datasets from other healthcare institutions.

Finally, this study compared Random Forest with conventional machine learning algorithms but did not evaluate boosting-based methods such as XGBoost, LightGBM, or CatBoost, which have shown promising performance in clinical prediction tasks. These limitations could be addressed in future studies by incorporating more diverse datasets and additional clinical variables, conducting external validation, and

comparing the model with advanced ensemble learning methods.

4. Conclusions

This research developed an interpretable approach for multiclass pregnancy risk prediction using maternal health data gathered from five community health centers in Koto Tangah District, Padang City, Indonesia. The framework incorporated data preprocessing, feature scaling using StandardScaler, oversampling with SMOTE, hyperparameter optimization using GridSearchCV, and model interpretation through SHAP to enhance prediction transparency. Of the four classification algorithms examined—Logistic Regression, Decision Tree, Support Vector Machine, and Random Forest—Random Forest achieved the strongest overall predictive performance, with a 97.26% accuracy, 90.44% precision, 98.06% recall, 93.65% F1-score, and 99.79% ROC-AUC. Furthermore, SHAP-based interpretation provided greater insight into the model by highlighting Heart Rate, Blood Glucose, Diastolic Blood Pressure, and Systolic Blood Pressure as the most influential predictors and identifying important interactions among the clinical variables.

These results demonstrate the potential of integrating ensemble learning techniques with explainable artificial intelligence to achieve reliable and interpretable pregnancy risk classification. In addition, the integrated predictive approach may serve as a decision-support tool for early risk assessment and referral prioritization, particularly in primary healthcare environments. Future research should evaluate the model's performance with extensive, multi-institutional datasets, incorporate additional maternal clinical and laboratory variables, and benchmark its performance against advanced boosting algorithms such as XGBoost, LightGBM, and CatBoost to further strengthen its generalizability and clinical relevance.

Acknowledgements

The authors wish to thank the Directorate of Research, Technology, and Community Service (DPPM), Ministry of Higher Education, Science and Technology (Kemdiktisaintek), Indonesia, for their financial support via the Research Grant Program (Grant No. 015/LL10/DT.05.00/PL/2025).

References

[1] S. Bhandari, Y. Dwa, M. Maharjan, and M. Bajracharya, "High Risk Pregnancy and its Outcome in a Tertiary Care Hospital: A Descriptive Cross-sectional Study," *Journal of the Nepal Medical Association*, vol. 62, no. 273, pp. 306–310, May 2024, doi: 10.31729/jnma.8561.

[2] R. Sirisomboon *et al.*, "Enhancing the competencies of obstetrical nurses and midwives in high-risk pregnancy management through simulation-based training in Lao people's democratic republic: A pilot study," *Midwifery*, vol. 137, Oct. 2024, doi: 10.1016/j.midw.2024.104132.

[3] R. Arjariya and P. Tiwari, "Prevalence of High Risk Pregnancy: in A Tertiary Care Centre of Sagar Division of M.P," *Int. J. Physiol.*, vol. 9, no. 1, pp. 36–42, Feb. 2021, doi: https://doi.org/10.37506/ijop.v9i1.2609.

[4] D. A. Galgalo, P. N. Nyakundi, Á. Várnagy, and V. Prémusz, "Evaluating the effectiveness of the global position system beaded bracelets to improve the uptake of antenatal care, health facility delivery, and post-natal care services among pastoralist women in Kenya," Jul. 01, 2025, *Elsevier Ltd*. doi: 10.1016/j.healthplace.2025.103484.

[5] M. Mathieu *et al.*, "Incidence, causes, and risk factors of stillbirth in an Amazonian context: Saint Laurent du Maroni maternity ward 2016–2021," *Eur. J. Obstet. Gynecol. Reprod. Biol. X*, vol. 18, Jun. 2023, doi: 10.1016/j.eurox.2023.100190.

[6] S. Dahiya, S. Anand, V. Rani, S. Gautam, and S. Nanda, "A Study of High-Risk Factors in Ante-natal Women At A Tertiary Care Centre," *Int. J. Reprod. Contracept. Obstet. Gynecol.*, vol. 9, no. 7, p. 2959, Jun. 2020, doi: 10.18203/2320-1770.ijrcog20202740.

[7] O. Kelly *et al.*, "The impact of diabetes mellitus on the development of psychiatric and neurological disorders," Jun. 01, 2024, *Elsevier B.V.* doi: 10.1016/j.dscb.2024.100135.

[8] K. W. Southern *et al.*, "Standards for the care of people with cystic fibrosis; establishing and maintaining health," *Journal of Cystic Fibrosis*, vol. 23, no. 1, pp. 12–28, Jan. 2024, doi: 10.1016/j.jcf.2023.12.002.

[9] S. E. Cuda *et al.*, "Metabolic, behavioral health, and disordered eating comorbidities associated with obesity in pediatric patients: An Obesity Medical Association (OMA) Clinical Practice Statement 2022," *Obesity Pillars*, vol. 3, Sep. 2022, doi: 10.1016/j.obpill.2022.100031.

[10] O. Szewczuk, W. Szewczuk, M. Walędziak, and A. Różańska-Walędziak, "Spontaneous regression of fetal pleural effusion in pregnancy complicated with Herpes simplex infection: Clinical presentation and literature review," Aug. 01, 2024, *Elsevier Ireland Ltd*. doi: 10.1016/j.ejogrb.2024.06.024.

[11] J. A. Joglar *et al.*, "2023 HRS expert consensus statement on the management of arrhythmias during pregnancy," *Heart Rhythm*, vol. 20, no. 10, pp. e175–e264, Oct. 2023, doi: 10.1016/j.hrthm.2023.05.017.

[12] T. Manca, "'Vaccines protect both you and your newborn:' A discourse analysis of risk and uncertainty in information about vaccination in pregnancy," *Soc. Sci. Med.*, vol. 364, Jan. 2025, doi: 10.1016/j.socscimed.2024.117526.

[13] U. Mahadevan *et al.*, "Global Consensus Statement on the Management of Pregnancy in Inflammatory Bowel Disease," *Clinical Gastroenterology and Hepatology*, vol. 23, no. 11, pp. S1–S60, Oct. 2025, doi: 10.1016/j.cgh.2025.04.005.

[14] M. Engelhardt *et al.*, "Exclusion and Othering in maternity care for forced migrant women: a qualitative study with forced migrant mothers and health care professionals in Germany," *SSM - Qualitative Research in Health*, vol. 8, Dec. 2025, doi: 10.1016/j.ssmqr.2025.100614.

[15] K. Pavagada and J. Vemuri, "Prediction of Maternal Health Risk Factors Using Machine Learning Algorithms," in *Procedia Computer Science*, Elsevier B.V., 2025, pp. 2713–2722. doi: 10.1016/j.procs.2025.04.532.

[16] A. O. Khadidos, F. Saleem, S. Selvarajan, Z. Ullah, and A. O. Khadidos, "Ensemble machine learning framework for predicting maternal health risk during pregnancy," *Sci. Rep.*, vol. 14, no. 1, Dec. 2024, doi: 10.1038/s41598-024-71934-x.

[17] K. Mahmud Sujon, R. Binti Hassan, Z. Tusnia Towshi, M. A. Othman, M. Abdus Samad, and K. Choi, "When to Use Standardization and Normalization: Empirical Evidence from Machine Learning Models and XAI," *IEEE Access*, vol. 12, pp. 135300–135314, 2024, doi: 10.1109/ACCESS.2024.3462434.

[18] R. Shetty, S. Gupta, V. Mediratta, S. Rai, and M. Geetha, "Optimizing Machine Learning-Based Ovarian Cancer Prediction Through Normalization Strategies," *IEEE Access*, vol. 13, pp. 128974–128995, 2025, doi: 10.1109/ACCESS.2025.3590871.

[19] I. Aruleba and Y. Sun, "Enhanced credit risk prediction using deep learning and SMOTE-ENN resampling," *Machine Learning with Applications*, vol. 21, p. 100692, Sep. 2025, doi: 10.1016/j.mlwa.2025.100692.

[20] M. Yusoff, Y. Mahmud, P. A. R. Azmi, and M. T. M. Sallehuddin, "The improvement of SMOTE-ENN-XGBoost through Yeo Johnson strategy on Dissolved Gas Analysis dataset," *Energy Reports*, vol. 13, pp. 6281–6290, Jun. 2025, doi: 10.1016/j.egyr.2025.05.013.

[21] R. K. R. Sarker, S. Hafij, M. A. Yasir, M. Assaduzzaman, M. M. H. Shimul, and M. K. Hossain, "Predicting maternal health risk using PCA-enhanced XGBoost and SMOTE-ENN for improved healthcare outcomes," *Intell. Based. Med.*, vol. 12, Jan. 2025, doi: 10.1016/j.ibmed.2025.100300.

[22] S. Matharaarachchi, M. Domaratzki, and S. Muthukumarana, "Enhancing SMOTE for imbalanced data with abnormal minority instances," *Machine Learning with Applications*, vol. 18, p. 100597, Dec. 2024, doi: 10.1016/j.mlwa.2024.100597

[23] D. Dash, M. Kumar, S. Patra, A. Kumar, and A. Ganguly, "Healthcare Fraud Detection Using an Integrated ML Approach with SMOTE," in *Procedia Computer Science*, Elsevier B.V., 2025, pp. 800–810. doi: 10.1016/j.procs.2025.04.312.

[24] Y. Belsti, L. Moran, A. Mousa, H. Teede, and J. Enticott, "Evaluation of machine learning and logistic regression-based gestational diabetes prognostic models," *J. Clin. Epidemiol.*, vol. 187, Nov. 2025, doi: 10.1016/j.jclinepi.2025.111957.

[25] T. Wahyuningsih, D. Manongga, I. Sembiring, and S. Wijono, "Comparison of Effectiveness of Logistic Regression, Naive Bayes, and Random Forest Algorithms in Predicting Student Arguments," in *Procedia Computer Science*, Elsevier B.V., 2024, pp. 349–356. doi: 10.1016/j.procs.2024.03.014.

[26] Y. Gao, Y. Wang, L. Tian, X. Hong, C. Xue, and D. Li, "Evolving adaptive and interpretable decision trees for cooperative submarine search," *Defence Technology*, Feb. 2025, doi: 10.1016/j.dt.2025.02.007.

[27] R. Zhao *et al.*, "Decision tree-based parameter identification and state estimation: Application to Reactor Operation Digital Twin," *Nuclear Engineering and Technology*, vol. 57, Jul. 2025, doi: 10.1016/j.net.2025.103527.

[28] M. Piazza, A. Spinelli, F. Maggioni, M. Bedoni, and E. Messina, "A robust support vector machine approach for Raman data classification," *Decision Analytics Journal*, vol. 16, Sep. 2025, doi: 10.1016/j.dajour.2025.100595.

[29] M. P. Raghunath *et al.*, "PCA and PSO based optimized support vector machine for efficient intrusion detection in internet of things," *Measurement: Sensors*, vol. 37, Feb. 2025, doi: 10.1016/j.measen.2024.101806.

[30] W. Li, S. Chen, L. Lin, and L. Chen, "Random-forest-based task pricing model and task-accomplished model for crowdsourced emergency information acquisition," *Systems and Soft Computing*, vol. 7, Dec. 2025, doi: 10.1016/j.sasc.2025.200235.

[31] M. García-Torres, F. Saucedo, F. Divina, and S. Gómez-Guerrero, "RFMSU: A multivariate symmetrical uncertainty-based random forest," *Pattern Recognit.*, vol. 169, Jan. 2026, doi: 10.1016/j.patcog.2025.111939.

[32] D. C. E. Saputra, K. Sunat, T. Ratnaningsih, A. Ma'arif, and I. Suwarno, "SmartForest: Optimizing anemia diagnosis with enhanced random forest and XGBoost integration," *Results in Engineering*, vol. 27, Sep. 2025, doi: 10.1016/j.rineng.2025.106442.

[33] E.-B. Rachid, S. Abderrahim, A. Hafid, and R. Souad, "Predicting water potability using a machine learning approach," *Environmental Challenges*, vol. 19, p. 101131, Jun. 2025, doi: 10.1016/j.envc.2025.101131.

[34] F. L. P. de Souza, M. A. Dias, T. D. Setiyono, S. Campos, L. S. Shiratsuchi, and H. Tao, "Identification of soybean planting gaps using machine learning," *Smart Agricultural Technology*, vol. 10, Mar. 2025, doi: 10.1016/j.atech.2025.100779.

[35] T. Pamutha, W. Promthong, and S. Pahlawan, "Diabetes Disease Prediction Using Machine Learning Classification Algorithms," vol. 15, no. 4, 2025, doi: https://doi.org/10.18517/ijaseit.15.4.20457.

[36] H. Elbatanouny, H. Tawfik, T. Khater, and A. Gorbenko, "An interpretable machine learning model to predict hospitalizations," *Clinical eHealth*, vol. 8, pp. 53–65, Dec. 2025, doi: 10.1016/j.ceh.2025.03.004.

[37] M. Ali, "Classification of imbalanced travel mode choice dataset with SMOTE and prediction using interpretable machine learning," *Sustainable Futures*, vol. 10, Dec. 2025, doi: 10.1016/j.sftr.2025.101119.

[38] M. Shamshuzzoha, T. T. B. Audry, M. J. Alam, Z. A. Bhuiyan, M. Motaharul Islam, and M. M. Hassan, "A novel framework for seasonal affective disorder detection: Comprehensive machine learning analysis using multimodal social media data and SMOTE," *Acta Psychol. (Amst).*, vol. 256, Jun. 2025, doi: 10.1016/j.actpsy.2025.105005.

[39] P. Supakar, M. Sarkar, and B. Sarkar, "Forecasting grocery item sales using gradient boosting models: A study of GridSearchCV, RandomizedSearchCV, and optuna optimization approaches," *Ain Shams Engineering Journal*, vol. 17, no. 3, Mar. 2026, doi: 10.1016/j.asej.2026.104012.

[40] S. Sadaghiyanfam, H. Kamberaj, and Y. Isler, "Enhanced prediction of ionic liquid toxicity using a meta-ensemble learning framework with data augmentation," *Artificial Intelligence Chemistry*, vol. 3, no. 1, Jun. 2025, doi: 10.1016/j.aichem.2025.100087.

[41] O. Rainio, J. Teuho, and R. Klén, "Evaluation metrics and statistical tests for machine learning," *Sci. Rep.*, vol. 14, no. 1, Dec. 2024, doi: 10.1038/s41598-024-56706-x.

[42] J. Opitz, "A Closer Look at Classification Evaluation Metrics and a Critical Reflection of Common Evaluation Practice", doi: 10.1162/tacl.

[43] K. Gajowniczek and T. Ząbkowski, "ImbTreeAUC: An R package for building classification trees using the area under the ROC curve (AUC) on imbalanced datasets," *SoftwareX*, vol. 15, Jul. 2021, doi: 10.1016/j.softx.2021.100755.

[44] G. Niveditha, B. U. Maheswari, and R. P. De Prado, "Elevating Maternal Healthcare: Synergy of Cardiotocography, Machine Learning Models and Interpretive Analysis," in *Procedia Computer Science*, Elsevier B.V., 2025, pp. 2787–2797. doi: 10.1016/j.procs.2025.04.539.

[45] A. M. Salih *et al.*, "A Perspective on Explainable Artificial Intelligence Methods: SHAP and LIME," *Advanced Intelligent Systems*, vol. 7, no. 1, Jan. 2025, doi: 10.1002/aisy.202400304.

[46] Z. Sadeghi *et al.*, "A review of Explainable Artificial Intelligence in healthcare," *Computers and Electrical Engineering*, vol. 118, Aug. 2024, doi: 10.1016/j.compeleceng.2024.109370.

[47] V. Vimbi, N. Shaffi, and M. Mahmud, "Interpreting artificial intelligence models: a systematic review on the application of LIME and SHAP in Alzheimer's disease detection," Dec. 01, 2024, *Springer Science and Business Media Deutschland GmbH*. doi: 10.1186/s40708-024-00222-1.

[48] O. O. Oladimeji, H. Ayaz, I. McLoughlin, and S. Unnikrishnan, "Mutual information-based radiomic feature selection with SHAP explainability for breast cancer diagnosis," *Results in Engineering*, vol. 24, Dec. 2024, doi: 10.1016/j.rineng.2024.103071.

[49] M. Herrero-Tudela, R. Romero-Oraá, R. Hornero, G. C. Gutiérrez Tobal, M. I. López, and M. García, "An explainable deep-learning model reveals clinical clues in diabetic retinopathy through SHAP," *Biomed. Signal Process. Control*, vol. 102, Apr. 2025, doi: 10.1016/j.bspc.2024.107328.