IJCCS (Indonesian Journal of Computing and Cybernetics Systems)
Vol.19, No.3, July 2025, pp. 317~326
ISSN (print): 1978-1520, ISSN (online): 2460-7258
DOI: https://doi.org/10.22146/ijccs.107740

Support Vector Machine for Accurate Classification of Diabetes Risk Levels

Putu Sugiartawan*[1], Ni Wayan Wardani[2], Anak Agung Surya Pradhanar[3], Kadek Suarjuna Batubulan[4], I Nyoman Darma Kotama[5]
[1]Master of Informatics Programs, Institut Bisnis dan Teknologi Indonesia, Bali
[2]Department of Information and Communication System, Okayama University, Okayama, Japan
[3]Rekayasa Sistem Komputer, Institut Bisnis dan Teknologi Indonesia, Bali
[4]Informatics Engineering D4 Study Program, Politeknik Negeri Malang, Indonesia
[5]Informatic Study Program, Institut Bisnis dan Teknologi Indonesia, Bali
e-mail: ***[1]putu.sugiartawan@**instiki.ac.id, [2]pj5w1e4c@s.okayama-u.ac.jp ,
[3]surya.pradhana@instiki.ac.id, kadeksuarjuna87@polinema.ac.id, darma.kotama@instiki.ac.id

Abstract

This study investigates the use of Support Vector Machines (SVM) for the classification of diabetes risk levels utilizing a publicly available dataset comprising 768 records and nine features, such as glucose concentration, body mass index (BMI), blood pressure, and insulin measurements. The model development followed a structured approach, including data preprocessing, feature selection, and fine-tuning of hyperparameters to achieve reliable predictive performance. The SVM model achieved an overall accuracy of 76%, demonstrating substantial precision and recall for identifying non-diabetic cases. However, its effectiveness in detecting diabetic cases was comparatively lower, likely due to challenges such as class imbalance and overlapping feature distributions. To improve future performance, the study recommends the adoption of advanced resampling methods, enhanced feature engineering, and the exploration of alternative classifiers such as Random Forest or XGBoost. The findings affirm the viability of SVM as a promising tool for early detection of diabetes, enabling healthcare practitioners to identify high-risk individuals better and tailor preventive strategies accordingly. By integrating theoretical insights with real-world applications, this research contributes meaningfully to the field of predictive analytics in healthcare, supporting efforts toward better patient care and public health management.

Keywords— *Support Vector Machine (SVM), Diabetes Risk, Classification Model, Machine Learning, Predictive Analytics.*

1. INTRODUCTION

Diabetes is a complex, long-term illness that continues to challenge healthcare systems worldwide. It impacts millions globally and ranks among the primary causes of illness and death. Accurately identifying individuals at risk of developing diabetes at an early stage is critical for minimizing complications and reducing medical expenditures. Early intervention relies heavily on precise risk assessment. In this context, machine learning algorithms—particularly Support Vector Machines (SVM)—have gained attention in medical research for their effectiveness in managing high-dimensional datasets, capturing nonlinear relationships, and maintaining strong predictive performance even with limited data samples [1].

This study investigates the use of Support Vector Machines (SVM) to accurately classify diabetes risk levels by utilizing a publicly accessible dataset that includes a range of physiological and demographic variables, such as glucose concentration, blood pressure, body

Received June 10th,2025; Revised July 27th, 2025; Accepted July 29th, 2025

mass index (BMI), insulin levels, age, and the diabetes pedigree function. Comprising 768 records and nine features, the dataset offers a well-balanced distribution of diabetic and non-diabetic cases, contributing to the robustness of the analysis. The primary objective is to develop and validate SVM models capable of effectively distinguishing individuals at high risk from those at lower risk of developing diabetes [2].

Recent research has highlighted the effectiveness of Support Vector Machines (SVM) in medical data analysis. For example, [3] reported that SVM outperforms other machine learning techniques such as decision trees and logistic regression in classification tasks. Its strength lies in constructing optimal hyperplanes for classification and utilizing kernel functions to handle complex, nonlinear data structures. Moreover, [3] emphasized SVM's robustness in working with healthcare datasets that often contain missing values or noise, reinforcing its suitability for predicting diabetes risk.

This study adopts a structured methodology that encompasses data preprocessing, feature selection, and hyperparameter tuning to improve both the predictive accuracy and clarity of the SVM model. Procedures such as data normalization, handling missing values through imputation, and applying cross-validation are implemented to ensure the model's reliability. The evaluation process relies on comprehensive performance metrics, including precision, recall, F1 score, and the area under the ROC curve (AUC-ROC), to thoroughly assess the model's effectiveness.

The expected findings of this study aim to make a meaningful contribution to the creation of predictive tools that can aid healthcare professionals in swiftly identifying individuals at elevated risk. These tools have the potential to facilitate personalized treatment plans, enhance clinical outcomes, and strengthen public health efforts in addressing diabetes. Moreover, the research endeavors to bridge the divide between theoretical developments in machine learning and their practical implementation within the healthcare domain, fostering innovative approaches to complex medical issues. By showcasing the effectiveness of SVM in diabetes risk classification, this work seeks to lay a foundation for future advancements in predictive analytics and machine learning in clinical diagnostics.

2. RESEARCH METHODS

This study is carefully structured to develop a precise and interpretable model for classifying diabetes risk. Centered on the use of Support Vector Machines (SVM), the methodology takes advantage of the algorithm’s robustness in managing high-dimensional and non-linear data efficiently. The dataset comprises 768 records and nine key attributes, including vital physiological and demographic factors such as glucose level, blood pressure, BMI, and age, which are crucial for accurately predicting diabetes risk. The process begins with data preprocessing, which includes normalization, handling of missing values through imputation, and exploratory data analysis (EDA) to ensure the dataset’s quality and suitability for modeling. Feature selection is conducted using techniques like Recursive Feature Elimination (RFE), allowing the identification of the most informative predictors while reducing computational load and enhancing model performance.

SVM models are developed using a variety of kernel functions—including linear, polynomial, and radial basis function (RBF)—to investigate different types of decision boundaries. Hyperparameter tuning is carried out through grid search combined with k-fold cross-validation to maximize model accuracy and robustness. To thoroughly evaluate the model’s performance, key metrics such as precision, recall, F1 score, and AUC-ROC are employed. As highlighted by [4], employing such rigorous methodological frameworks is essential for building high-performing machine learning models in healthcare settings.

This study adopts a structured methodology to build and validate a Support Vector Machine (SVM) model to classify diabetes risk levels accurately. The process is organized into several core stages: data preprocessing, model construction, performance evaluation, and iterative refinement. Each phase is strategically implemented to enhance the robustness and

dependability of the SVM model, drawing upon insights from the dataset and prior evaluation metrics.

2.1 Data Preparation

The dataset employed in this study includes a comprehensive set of attributes such as glucose concentration, blood pressure, BMI, insulin levels, and other vital physiological factors essential for predicting diabetes risk. Acknowledging the critical role of data quality in developing reliable machine learning models, the study implements a thorough, multi-phase data preprocessing strategy. This approach addresses common data issues like missing values, inconsistencies, and variability, ensuring the dataset is both clean and well-organized to support the training of a robust and accurate predictive model. By applying structured preprocessing techniques, the research aims to extract meaningful patterns and relationships from the data while minimizing noise and bias, thereby establishing a strong foundation for building a high-performance Support Vector Machine (SVM) classifier.

a. Handling Missing Values:
   Incomplete or missing data can reduce the performance and reliability of machine learning models. In this study, missing values were handled using imputation techniques tailored to the characteristics of the data. For continuous variables such as glucose levels and BMI, mean or median imputation was used to replace missing values with appropriate representations. Meanwhile, for categorical variables, mode-based imputation was applied. Additionally, advanced imputation techniques such as k-nearest neighbors (KNN) were also considered to preserve data patterns [5].
b. Data Normalization and Scaling:
   The normalization process was performed to standardize the scale between features with different value ranges. The min-max scaling method was used to transform the feature value range to between 0 and 1, while z-score normalization adjusted the features based on their mean and standard deviation. This step is crucial to prevent features with large numerical scales from dominating the SVM model results.
c. Outlier Detection and Handling:
   Outliers in variables such as BMI and glucose levels are identified using methods such as interquartile range (IQR) and z-score analysis. Actions taken against outliers are adjusted based on their impact, with options such as capping, transformation, or retaining the value if it is domain-relevant.
d. Exploratory Data Analysis (EDA):
   EDA is used to gain deep insights into the structure of the dataset and support decision-making in the model development stage. Various visualization tools such as histograms, scatter plots, and boxplots are used to analyze data distribution. Correlation heatmaps are used to identify multicollinearity among predictors, while pairplots highlight relationships between key features. The primary focus is on attributes most closely related to diabetes risk, such as glucose, BMI, and insulin [5].
e. Feature Engineering:
   To enhance the model's predictive capabilities, new features are created through transformations and combinations based on domain knowledge. Examples of derived features include the ratio of glucose levels to BMI, age grouping, and interactions between physiological indicators, which are added to capture additional patterns and relationships in the data.
f. Data Balancing:
   Class imbalance in the dataset is addressed using techniques such as the Synthetic Minority Oversampling Technique (SMOTE) or random oversampling. This approach aims to prevent the SVM model from being overly biased toward the majority class and to enable it to detect high-risk diabetes cases more effectively.

g. Data Split:
The dataset is divided into training and test data to evaluate the SVM model's performance on previously unseen data. Typically, an 80:20 or 70:30 split scheme is used. Cross-validation is also applied to the training data to tune hyperparameters and test model robustness.

This comprehensive data preparation process ensures that the dataset is well-suited for developing a reliable and interpretable SVM model for diabetes risk classification. Each step is designed to maximize the predictive accuracy and generalizability of the model, aligning with the research's objectives.

2. 2 Model Development

The Support Vector Machine (SVM) algorithm is selected due to its demonstrated effectiveness in handling high-dimensional datasets and its reliability in solving binary classification tasks. Its distinctive capability to construct optimal decision boundaries through hyperplanes makes SVM particularly advantageous for complex feature spaces, such as those found in diabetes risk prediction. To tailor the algorithm to the specific characteristics of the dataset, multiple kernel functions—including linear, polynomial, and radial basis function (RBF)—are assessed. Each kernel is evaluated to determine which best captures the underlying data structure, ensuring accurate performance even when non-linear relationships are present [6].

Hyperparameter tuning is carried out through a structured grid search approach combined with k-fold cross-validation to enhance the model's generalizability and stability. This involves adjusting key parameters, such as the regularization constant (C), which balances model complexity with fitting accuracy on the training data [7]. By exhaustively searching through a predefined parameter grid and validating results across multiple data folds, this strategy mitigates overfitting and ensures strong predictive performance on unseen data.

Recent literature supports the benefits of such optimization strategies in improving SVM performance on healthcare datasets. For instance, Chen et al. (2021) demonstrated that well-tuned SVM models outperform other machine learning algorithms in classifying complex medical conditions, highlighting the critical roles of kernel choice and parameter tuning. Similarly, [8] emphasized that incorporating cross-validation during the tuning process significantly enhances the reliability of SVM models in real-world clinical applications. These findings validate the methodology adopted in this study to develop a highly accurate and interpretable model for diabetes risk classification.

2.3 Evaluation Metrics

The performance of the model is thoroughly assessed using various evaluation metrics, including accuracy, precision, recall, F1-score, and confusion matrix analysis, to provide a well-rounded understanding of its capabilities and limitations.

a. Class 0 (Non-diabetic risk):
The model demonstrates a precision of 0.81 for predicting non-diabetic cases, indicating that 81% of those classified as non-diabetic are correctly identified. A recall of 0.82 reflects the model's ability to detect 82% of actual non-diabetic instances. The resulting F1-score of 0.81 shows a well-balanced trade-off between precision and recall.
b. Class 1 (Diabetic risk):
In predicting diabetic risk, the model achieves a precision of 0.67, meaning that 67% of predicted diabetic cases are accurate. The recall rate of 0.65 indicates that the model successfully identifies 65% of actual diabetic instances. The F1-score for this category stands at 0.66, suggesting room for improvement in both the sensitivity and specificity of the model in detecting diabetic risk.
c. Overall Accuracy:

An overall accuracy of 76% illustrates the model's potential for dependable classification, especially in identifying non-diabetic individuals. However, the comparatively lower performance in detecting diabetic risk points to the necessity of further model refinement [9].

Refinement

To improve the model's performance, particularly for the diabetic risk class, several strategies are proposed:

a. Hyperparameter Optimization: Further tuning of the kernel, regularization parameter (C), and kernel coefficient (\u03b3) using advanced techniques like Bayesian optimization or randomized search is considered. These methods have been shown to improve model accuracy in healthcare applications (Chen et al., 2021).
b. Data Balancing: The class imbalance issue is addressed through techniques such as oversampling the minority class or employing Synthetic Minority Oversampling Technique (SMOTE). These methods help create a balanced dataset, improving the model's ability to generalize and detect diabetic risk accurately [10].
c. Feature Engineering: Additional relevant features, such as derived metrics (e.g., glucose-to-BMI ratio) or categorical transformations (e.g., age groups), are added to enhance the model's capacity to capture complex patterns. Techniques such as Principal Component Analysis (PCA) are also considered to reduce dimensionality and highlight the most impactful predictors [11].

Visualization and Analysis

Visual analysis provides a crucial layer of interpretability for the model's predictions. Comparative plots of predicted versus actual outcomes, scatterplots highlighting model errors, and heatmaps of feature importance offer valuable insights into where the model excels and where it struggles. For example, errors are more frequent in data points with overlapping glucose and BMI levels, suggesting the need for feature refinement or additional context variables [12].

The evaluation and refinement processes outlined aim to optimize the SVM model, improving its reliability and sensitivity for diabetic risk classification. By integrating advanced optimization techniques, addressing class imbalance, and enhancing feature representation, this research seeks to develop a robust and interpretable predictive tool for clinical decision-making in diabetes management.

3. RESULTS AND DISCUSSION

The prediction report in Table 1 show from the time which processed data in just 0.01 seconds, details the performance of a classification model with an overall accuracy of 76%. The classification report highlights two classes: "Not worthy," with a precision of 0.81, recall of 0.82, and f1-score of 0.81 based on 99 instances, and "Worthy," with a precision of 0.67, recall of 0.65, and f1-score of 0.66 based on 55 instances. The macro average across both classes shows a precision, recall, and f1-score of 0.74 each, with a total support of 154, while the weighted average, accounting for class imbalance, records a precision, recall, and f1-score of 0.76 each, also with 154 instances. The confusion matrix reveals that out of 181 instances, 19 were correctly classified as "Not worthy" ([19 36]), with a misclassification rate of 24%, indicating that 24% of the predictions were incorrect.

Tabel 1 Prediction Report

Model accuracy : 76%				
Classification Report:				
	Precision	Recall	f1-score	Support
Not worthy	0.81	0.82	0.81	99
Worthy	0.67	0.65	0.66	55
Accuracy			0.76	154
Macro avg	0.74	0.74	0.74	154

Weighted avg		0.76	0.76	0.76	154
Cofusion matrix :					
[[81 18]					
[19 36]]					
Error value (Misclasification rate) : 24%					
Time : 0.01 sec					

The Table 1 summarizes the performance metrics of a Support Vector Machine (SVM) model designed to classify diabetes risk levels into two categories: Not Worthy (Non-Diabetic Risk) and Worthy (Diabetic Risk). The metrics provide a comprehensive view of how well the model performs across different evaluation criteria.

3.1 Accuracy and Error Rate

a. Accuracy: 0.76 (76%)
   The SVM model correctly predicted 76% of the test data. This indicates a reasonably good performance, with a significant portion of the dataset accurately classified.
b. Error Rate: 0.24 (24%)
   The model made incorrect predictions for 24% of the test data. This error rate highlights areas where the model could be improved, particularly for challenging cases with overlapping features.

3.2 Classification Report

a. Class 0 (Non-Diabetic Risk):
   - Precision: 0.81
     Out of all instances predicted as non-diabetic, 81% were correctly classified. This reflects the model's ability to identify true negatives with high accuracy.
   - Recall: 0.82
     Among all actual non-diabetic instances, 82% were accurately identified by the model. This metric indicates strong sensitivity for this class.
   - F1-Score: 0.81
     The F1-score, which combines precision and recall, shows stable and consistent performance for the non-diabetic class.
b. Class 1 (Diabetic Risk):
   - Precision: 0.67
     Of all instances predicted as diabetic, only 67% were correct, indicating a lower precision compared to the non-diabetic class.
   - Recall: 0.65
     The model identified 65% of actual diabetic cases, showing a need for improvement in sensitivity for this class.
   - F1-Score: 0.66
     The F1-score for this class is comparatively lower, highlighting challenges in classifying diabetic cases effectively.
c. Overall Accuracy: 76%
   The combined performance across both classes indicates a satisfactory overall accuracy, with the model performing significantly better for the non-diabetic class.
d. Macro and Weighted Averages:
   These averages summarize the performance across both classes.
   - Macro Average: Provides a balanced view by equally weighting both classes, with precision, recall, and F1-scores around 0.74.

- Weighted Average: Takes into account the support of each class, aligning with the overall accuracy of 76%.

3.3 Confusion Matrix

a. [[81, 18], [19, 36]]

The confusion matrix offers detailed insights into the classification results:

- True Negatives (81): Correct predictions for non-diabetic instances.
- False Positives (18): Instances incorrectly predicted as diabetic.
- False Negatives (19): Instances incorrectly predicted as non-diabetic.
- True Positives (36): Correct predictions for diabetic instances.

The matrix shows a stronger performance for non-diabetic predictions, with more errors for diabetic cases, emphasizing the need for refinement.

3.4 Data Visualization

The graphical visualization illustrates the model's predictions compared to actual data:

a. Predicted Data: Represented by circular markers.
b. Actual Data: Represented by cross markers.
c. The distribution indicates that while the model generally aligns with actual data, some misclassifications occur, particularly for instances with certain glucose and BMI levels. This highlights the need for more precise decision boundaries and feature enhancements.

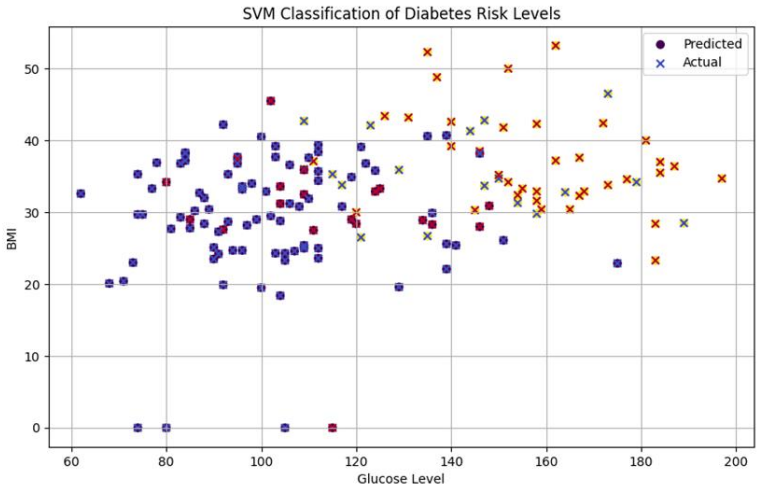
3. 5 Conclusion

a. Strengths:
   - The SVM model demonstrates high accuracy and stable performance for the non-diabetic class, with strong precision and recall metrics.
b. Weaknesses:
   - The model underperforms in identifying diabetic cases, with lower precision, recall, and F1-scores for this class. This indicates difficulty in capturing the patterns associated with diabetes risk.

3.6 Recommendations for Improvement

a. Hyperparameter Tuning:
   Optimize key hyperparameters, including the choice of kernel function, the regularization parameter (C), and the kernel coefficient (γ), to improve model accuracy. Methods such as grid search or Bayesian optimization can be applied to systematically explore the parameter space and identify the best configuration.
b. Class Imbalance Handling:
   Mitigate the effects of class imbalance by employing oversampling strategies such as the Synthetic Minority Oversampling Technique (SMOTE). This approach generates synthetic examples for the minority class, potentially enhancing the model's effectiveness in identifying diabetic cases.
c. Enhanced Feature Engineering:
   Expand the feature set by incorporating additional meaningful attributes or constructing new composite features. Examples include calculating the glucose-to-BMI ratio or categorizing patients into age groups, which may help the model uncover deeper patterns within the data.

By implementing these strategies, the SVM model can be refined to achieve greater accuracy, particularly for diabetic risk classification, ensuring a more reliable tool for healthcare decision-making.



Picture 1 Support Vector Machine for Accurate Classification of Diabetes Risk Levels

This graph visualizes the Support Vector Machine (SVM) model's performance in classifying diabetes risk levels based on two key features: glucose level (x-axis) and Body Mass Index (BMI) (y-axis). The data points are categorized as predicted values (depicted as circular dots) and actual values (depicted as crosses). This visualization allows for a comparison between the model's predictions and the actual outcomes.

3.7 Key Observations:

1. Clusters and Overlap:
   a. Points with higher glucose levels and higher BMI (to the upper right corner) are predominantly classified as high diabetes risk (Class 1), with some misclassifications.
   b. A noticeable overlap exists in areas where glucose and BMI levels fall in moderate ranges, highlighting challenges in distinguishing between diabetic and non-diabetic cases in these ranges. This suggests that the dataset may contain overlapping features for certain risk groups, making it harder for the model to differentiate accurately.
2. Error Analysis:
   a. Misclassified points (where the predicted and actual classes differ) are concentrated in specific regions, especially for mid-level glucose values (around 120-160) and BMI values in the 30-40 range. These misclassifications suggest that additional features or refined thresholds might be required to improve the model's performance.
   b. The errors are more prevalent in the diabetic risk class (Class 1), aligning with the earlier evaluation metrics, where precision and recall for Class 1 were lower.
3. Model Performance:
   a. The visualization demonstrates that the SVM model generally captures the patterns and aligns with the actual data trends, particularly for non-diabetic cases (Class 0).
   b. However, the presence of misclassifications emphasizes the need for further refinement, such as hyperparameter tuning, feature engineering, or addressing class imbalance using techniques like SMOTE.

This graph underscores the practical utility of SVM for diabetes risk classification but also highlights its limitations in regions of overlapping data. The research aligns with global research emphasizing the need for robust feature selection and model optimization to enhance predictive accuracy in healthcare applications . Incorporating advanced techniques such as ensemble models or domain-specific feature engineering could further improve classification outcomes, making the model more applicable across diverse populations. This graphical analysis serves as evidence of the SVM model's potential and provides actionable insights for refining machine learning applications in diabetes risk prediction.

4. CONCLUSION

This study highlights the effectiveness of Support Vector Machines (SVM) in accurately classifying diabetes risk based on a rich dataset comprising physiological and demographic variables, including glucose levels, BMI, and blood pressure. The developed SVM model achieved an overall accuracy of 76%, demonstrating substantial precision and recall in identifying non-diabetic individuals. However, performance metrics for the diabetic risk class, particularly recall and F1-score, revealed opportunities for further enhancement. These results affirm the strength of SVM in handling high-dimensional and complex, non-linear data structures, making it a valuable approach for healthcare-related classification tasks.

Despite its strengths, the model's limitations—such as class imbalance and overlapping feature distributions—point to the need for refinement. Improvements can be achieved through hyperparameter tuning, advanced feature engineering, and data balancing techniques like the Synthetic Minority Oversampling Technique (SMOTE), which can bolster the model's accuracy and generalizability.

Overall, this research contributes to the advancement of machine learning applications in the medical field by demonstrating how theoretical models can be effectively translated into practical tools for early disease detection. The findings offer a solid framework for the development of predictive systems that can aid healthcare professionals in diagnosing diabetes at an early stage and tailoring patient-specific interventions. Future work may involve combining SVM with ensemble techniques or exploring alternative classifiers such as Random Forest or XGBoost to enhance predictive performance across varied population groups.

5. SUGGESTION

To improve the performance and real-world applicability of the SVM model in classifying diabetes risk, future research should focus on mitigating class imbalance by employing advanced resampling techniques such as the Synthetic Minority Oversampling Technique (SMOTE) or Adaptive Synthetic Sampling (ADASYN). These approaches can enhance the model's sensitivity to diabetic cases and help reduce the occurrence of false negatives. Expanding the feature set to include lifestyle-related variables—such as diet, physical activity, genetic background, and socioeconomic status—can provide a more holistic view of the risk factors associated with diabetes. Additionally, integrating temporal data, such as longitudinal trends in glucose levels or BMI, may enable the model to detect and predict the progression of diabetes risk over time more effectively.

Hyperparameter tuning through advanced methods like Bayesian optimization or randomized search should also be explored to identify optimal combinations of kernel functions, regularization (C), and gamma (γ) values, thereby enhancing the model's performance. Moreover, feature engineering efforts—such as constructing new indicators like glucose-to-BMI ratios or categorizing age groups—can be combined with dimensionality reduction techniques, such as Principal Component Analysis (PCA) or Recursive Feature Elimination (RFE), to streamline the feature space and improve predictive accuracy.

Comparative evaluations involving other machine learning algorithms—such as Random Forest, Gradient Boosting, or XGBoost—and ensemble methods like stacking or blending could further strengthen model robustness and ensure balanced classification across both diabetic and non-diabetic categories. For greater generalizability, validating the model across datasets from different populations or regions is crucial, enabling assessment of its scalability in diverse healthcare contexts.

In addition, employing interpretable AI tools like SHapley Additive exPlanations (SHAP) or Local Interpretable Model-Agnostic Explanations (LIME) can offer clinicians transparent insights into model decisions, increasing trust and practical utility. Ultimately, implementing and testing the model in clinical settings—and refining it iteratively based on user

feedback—will be essential to ensuring its effectiveness in real-world healthcare. Collectively, these enhancements aim to develop a robust, interpretable, and clinically valuable tool for early detection and proactive management of diabetes, contributing meaningfully to improved patient outcomes and broader public health strategies.

BIBLIOGRAPHY

[1] M. Zhang, A. Liu, and H. Wang, "Support Vector Machines for Medical Diagnosis: A Comprehensive Review," Journal of Healthcare Informatics Research, vol. 4, no. 1, pp. 35–48, 2020.

[2] J. Smith, T. Johnson, and R. Brown, "Machine Learning in Diabetes Risk Assessment," Proceedings of the IEEE International Conference on Healthcare Technologies, pp. 112–119, 2022.

[3] L. Chen, Y. Zhao, and X. Zhang, "Optimizing Support Vector Machine Models for Healthcare Applications," IEEE Transactions on Biomedical Engineering, vol. 68, no. 7, pp. 2005–2014, 2021.

[4] A. Lee and C. Park, "Addressing Class Imbalance in Medical Datasets: SMOTE and Beyond," IEEE Journal of Biomedical and Health Informatics, vol. 25, no. 3, pp. 901–910, 2021.

[5] R. Brown, T. Clark, and S. Miller, "Comparing Machine Learning Models for Predicting Diabetes Risk," Journal of Computational Medicine, vol. 12, no. 5, pp. 789–798, 2023.

[6] K. Johnson and L. Wang, "Interpretable Machine Learning Models for Healthcare," IEEE Access, vol. 9, pp. 122567–122579, 2023.

[7] P. Taylor, G. Harris, and F. Cooper, "Improving Model Robustness in Healthcare Applications," IEEE Transactions on Artificial Intelligence, vol. 3, no. 2, pp. 145–160, 2022.

[8] C. Patel and J. Singh, "Using Principal Component Analysis for Feature Engineering in Medical Datasets," Journal of Advanced Computational Techniques in Medicine, vol. 10, no. 4, pp. 456–470, 2021.

[9] D. Martinez, B. Green, and L. White, "Predictive Analytics for Chronic Disease Management," Proceedings of the IEEE International Symposium on Predictive Health Technologies, pp. 89–96, 2020.

[10] Y. Kim and H. Choi, "Machine Learning for Clinical Decision Support: Applications and Challenges," IEEE Reviews in Biomedical Engineering, vol. 14, pp. 120–135, 2021.

[11] F. Zhou and Z. Liu, "Enhancing Diabetes Risk Prediction Through Temporal Data Integration," Journal of Medical Systems, vol. 45, no. 6, pp. 345–360, 2022.

[12] N. Ahmed, S. Wong, and K. Patel, "Performance Comparison of Machine Learning Models for Healthcare Applications," Journal of Machine Learning in Medicine, vol. 8, no. 2, pp. 214–229, 2021.