Innovations and Industrial Technologies 2025

Multifaceted Research on Diabetes Diagnosis Based on Medical and Non-Medical Data Using Machine Learning, Deep Learning, and Artificial Neural Network Techniques

dr Paweł Sobczak[1] | Antoni Ciesielski[2]

[1]Department of Computer Science, Polytechnic Faculty, The President Stanislaw Wojciechowski Calisia University, Kalisz,Poland;
[1]Department of Technical Sciences, Faculty of Economics and Technical Sciences, University of Applied Sciences in Konin, Konin, Poland.

[2]Technical Schools Complex named after Waldemar Gostomczyk in Ostrów Wielkopolski, Ostrów Wielkopolski, Poland

University of Kalisz

Kalisz, 23–24.10.2025

ABSTRACT

This study presents a comprehensive research compendium on the application of Artificial Intelligence (AI) and Machine Learning (ML) in diabetes diagnosis and prediction. It covers four main research areas: analysis of non-medical data, analysis of medical data, retinal image classification for diabetic retinopathy, and prediction of diabetic patient rehospitalization. The objective was to compare the performance of various ML models across these four dimensions of diabetic diagnostics and healthcare support. A wide range of algorithms was applied, including Extra Trees, XGBoost, Random Forest, Bagging, Gradient Boosting, CatBoost, and neural networks (MLP, Keras). The models were trained on real-world datasets (Diabetes Health Indicators Dataset, Healthcare Diabetes Dataset), achieving very high accuracy — up to 99% for the best-performing classifiers. To enhance model transparency and verify predictive reliability, a SHAP (SHapley Additive exPlanations) analysis was conducted, allowing precise interpretation of feature importance and individual prediction contributions. As part of the project, a web-based system — GlucoCheck AI — was developed, integrating all four modules into one cohesive platform built with Flask, HTML, CSS, and JavaScript. The system enables risk prediction, feature-importance visualization, medical data analysis, and retinal image classification using deep learning techniques. The findings confirm the potential of AI to improve diagnostic accuracy, interpretability, and accessibility in diabetes care. The study highlights the need for further clinical validation, data standardization, and integration of AI with healthcare infrastructures to advance digital and personalized medicine.

OBJECTIVES

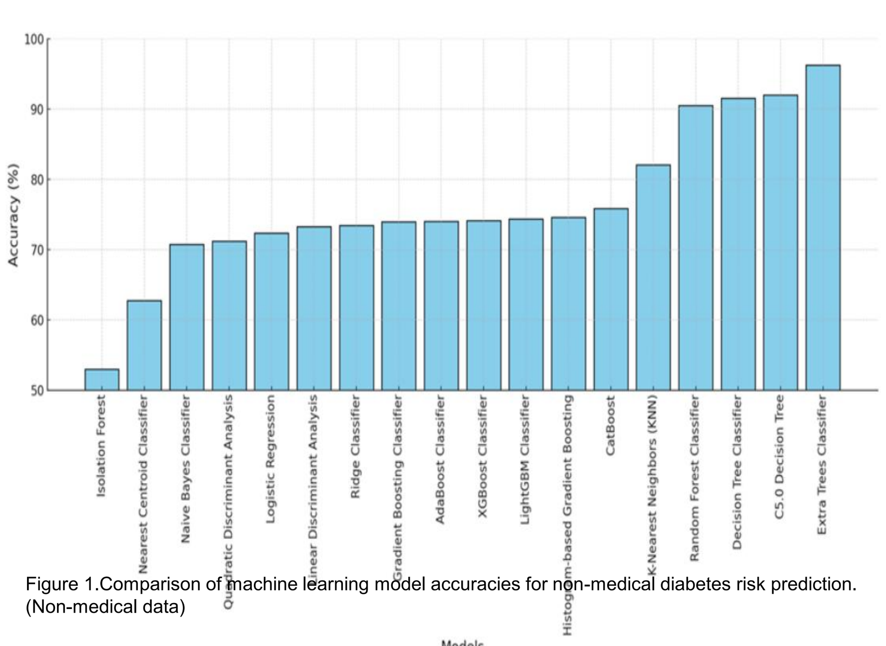


Figure 1.Comparison of machine learning model accuracies for non-medical diabetes risk prediction. (Non-medical data)

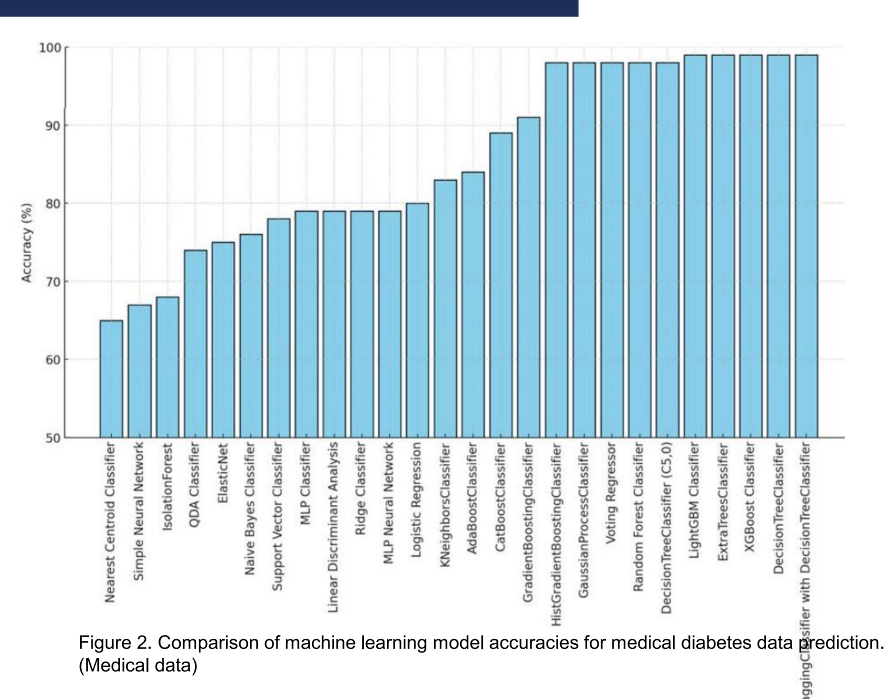


Figure 2. Comparison of machine learning model accuracies for medical diabetes data prediction. (Medical data)

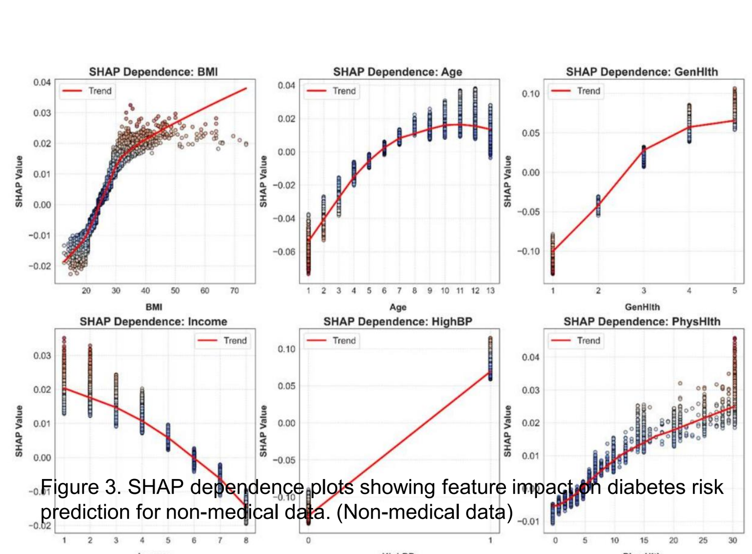


Figure 3. SHAP dependence plots showing feature impact on diabetes risk prediction for non-medical data. (Non-medical data)

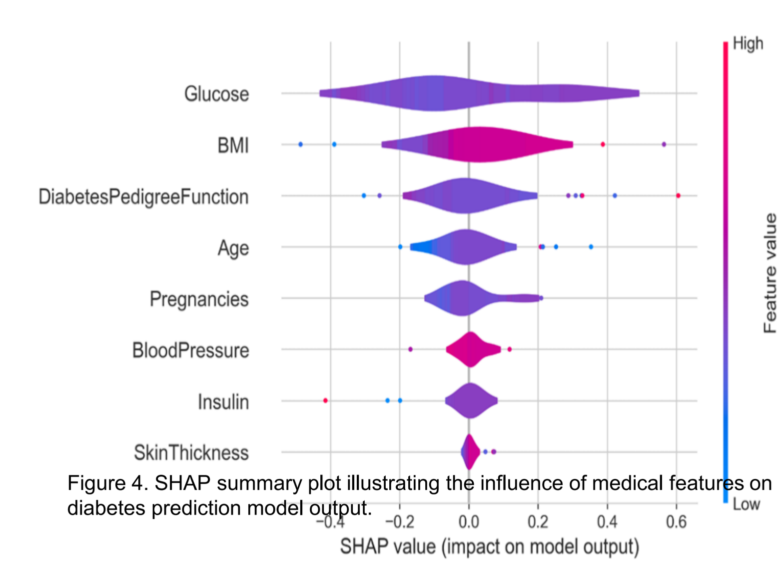


Figure 4. SHAP summary plot illustrating the influence of medical features on diabetes prediction model output.

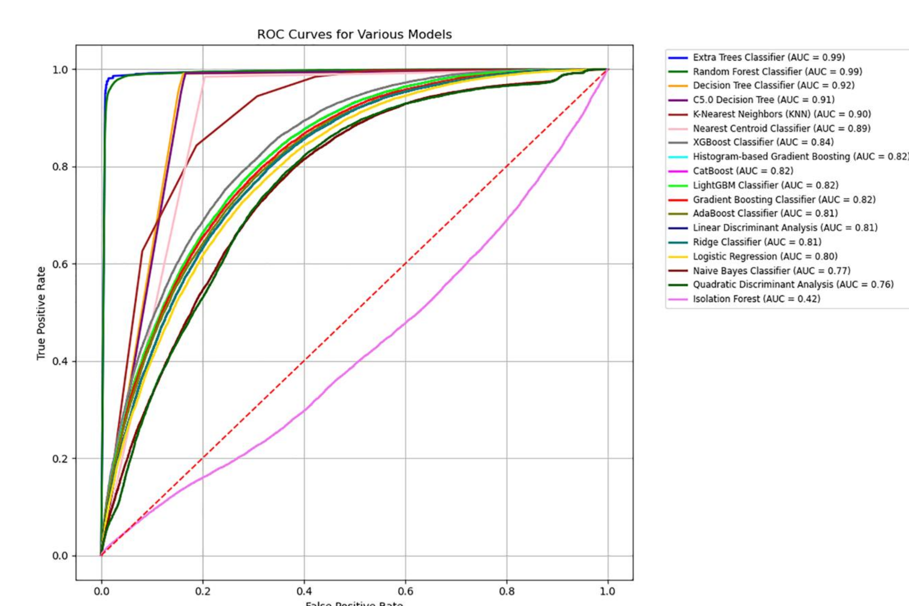


Figure 6. ROC curves comparing the performance (AUC) of various machine learning models for diabetes risk prediction. (Non-medical data)

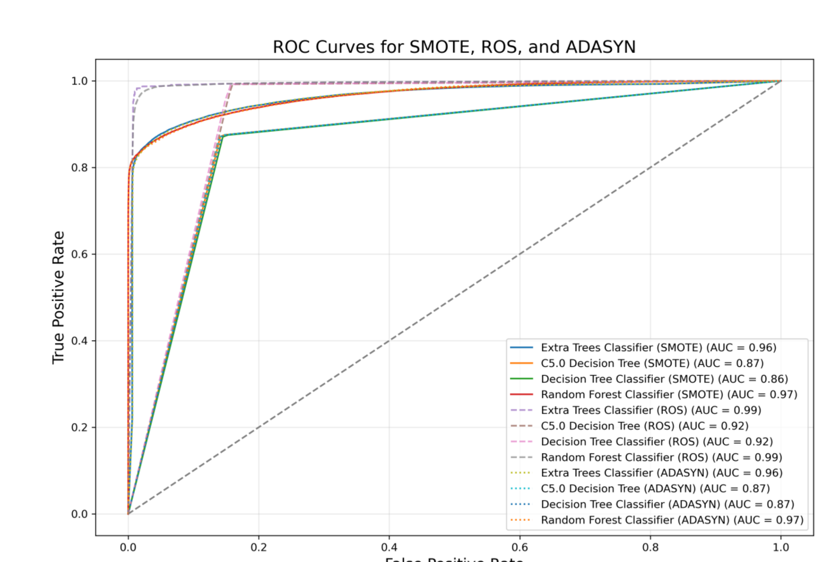
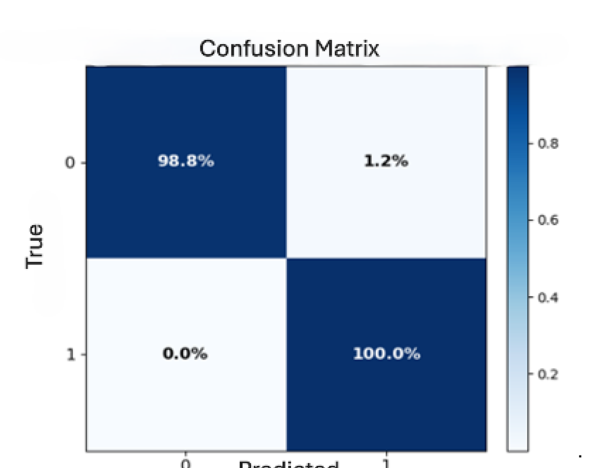
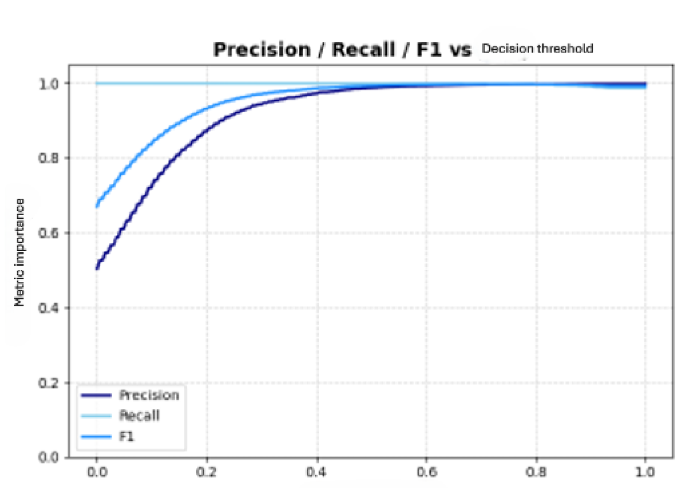


Figure 7. ROC curves illustrating model performance (AUC) using SMOTE, ROS, and ADASYN oversampling techniques for diabetes prediction. (Non-medical data)





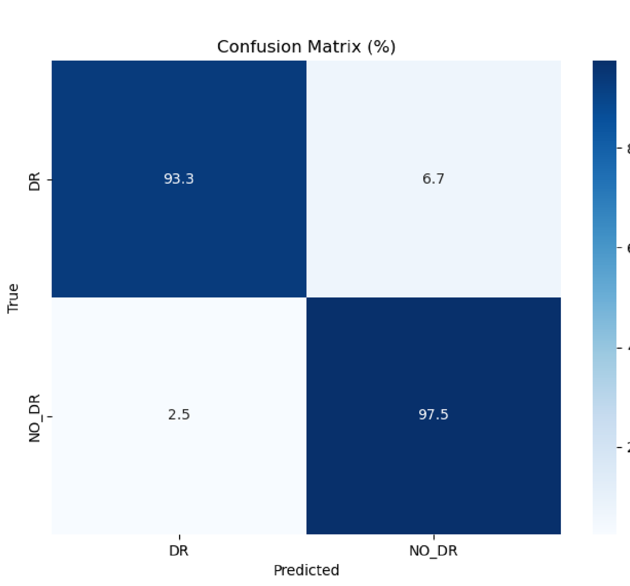


Figure 9. Confusion matrix illustrating classification performance for diabetic retinopathy detection. (Retinal (ophthalmic) data)

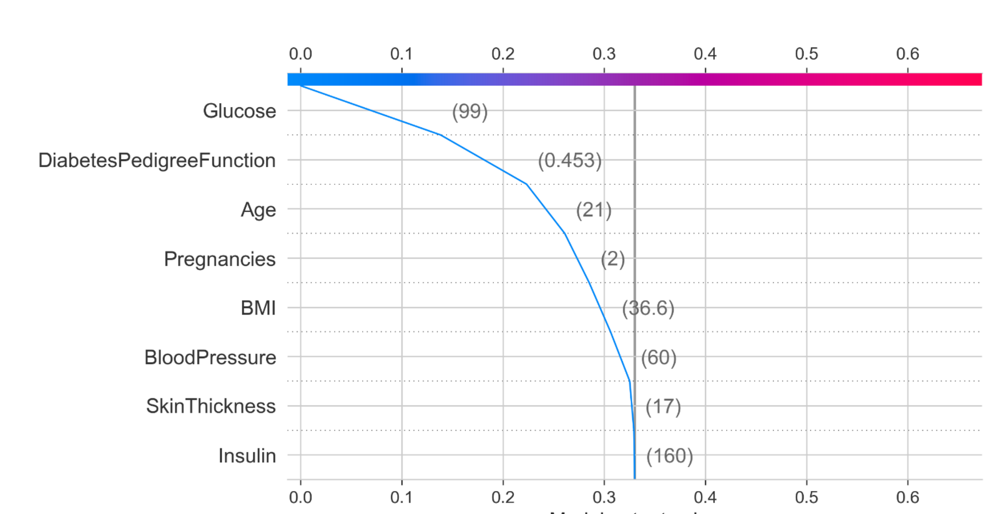


Figure 10. SHAP force plot illustrating individual feature contributions for medical diabetes data prediction. (Medical data)

Data Sources

The study utilized four publicly available datasets representing non-medical, medical, retinal, and hospital prediction domains.
Behavioral Risk Factor Surveillance System (BRFSS) 2015
Centers for Disease Control and Prevention (CDC), U.S. Department of Health and Human Services, Atlanta, USA.
https://www.cdc.gov/brfss/index.html (Non-medical data)
Pima Indians Diabetes Dataset (OpenML ID 43483)
https://www.openml.org/d/43483 (Medical data)
Diabetes 130-US Hospitals (1999–2008)
https://archive.ics.uci.edu/dataset/296/diabetes-130-us-hospitals-for-years-1999-2008 (Hospital prediction data)
Indian Diabetic Retinopathy Image Dataset (IDRiD)
https://ieee-dataport.org/open-access/indian-diabetic-retinopathy-image-dataset-idrid (Retinal imaging data)

CONCLUSIONS

The conducted research confirms the high efficiency and practical applicability of Artificial Intelligence (AI) and Machine Learning (ML) in diabetes diagnosis and prediction. The analysis across four main domains — non-medical data, medical data, retinal imaging, and rehospitalization prediction — demonstrated that advanced AI algorithms, including Extra Trees, XGBoost, and neural networks, can substantially improve classification accuracy compared to traditional statistical methods. The implementation of SHAP (SHapley Additive exPlanations) analysis allowed not only the verification of model reliability but also a detailed interpretation of feature contributions to individual predictions. This approach significantly enhanced model transparency and interpretability, strengthening trust among medical professionals. The strong alignment between AI predictions and real clinical data indicates the potential use of such models in Clinical Decision Support Systems (CDSS).The integration of four research modules within the GlucoCheck AI platform proved the feasibility of developing a modern, unified diagnostic and analytical tool that combines medical, behavioral, and image-based data. Such systems can support not only diagnosis but also prevention, treatment monitoring, and patient education, contributing to comprehensive diabetes management. Final conclusions emphasize the importance of further clinical validation, dataset expansion with diverse demographic groups, and the establishment of ethical and regulatory standards for AI applications in healthcare. The presented solution provides a robust foundation for the continued development and implementation of intelligent technologies toward personalized, preventive, and digital medicine.
Related publication (medical data module): Majcherek, D., Ciesielski, A., & Sobczak, P. (2025). AI-driven analysis of diabetes risk determinants in U.S. adults: Exploring disease prevalence and health factors. PLOS ONE. Available at: https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0328655

REFERENCES

1. Recent trends in diabetes mellitus diagnosis: an in-depth review of AI techniques. Frontiers in Artificial Intelligence, 2024. https://pubmed.ncbi.nlm.nih.gov/40328407
2. Artificial Intelligence Applications in Diabetic Retinopathy. Journal of Clinical Medicine, 2023. https://pmc.ncbi.nlm.nih.gov/articles/PMC11220221
3. Comparison of Machine Learning Models for Predicting 30-Day Readmission in Diabetes. Journal of Medical Artificial Intelligence, 2024. https://jmai.amegroups.org/article/view/9179
4. Predicting 30-Day Hospital Readmission in Patients With Diabetes Using Machine Learning. ResearchGate, 2025. https://www.researchgate.net/publication/390885160
5. DiabDeep: Pervasive Diabetes Diagnosis Based on Wearable Sensors and Neural Networks. arXiv preprint, 2024. https://arxiv.org/abs/1910.04925

CONTACT

p.sobczak@uniwersytetkaliski.edu.pl
+48660404803
antekciesielski2007@gmail.com
+48662670294

University of Kalisz

The project is co-financed from the state budget,
allocated by the Minister of Science under „Doskonała Nauka II" programme.

Ministerstwo Nauki i Szkolnictwa Wyższego