International Journal of Computer Information Systems and Industrial Management Applications
ISSN 2150-7988 Volume 18(21s) 2026 pp. 880-887
https://doi.org/10.70917/ijcisim-2026-5345
Article

Effective Machine Learning model for Diabetes insulin sensitivity

Lavanya D.[1], Dr. Arathi Sudarshan[2]

[1] Department of Mathematics, Jain (Deemed-to-be) University, Bengaluru, Karnataka, India.
Email: d.lavanya@jainuniversity.ac.in
[2] Department of Data Analytics & Mathematical Sciences, Jain (Deemed-to-be) University, Bengaluru, Karnataka, India.
Email: arathi.sudarshan@jainuniversity.ac.in

Abstract: Accurate assessment of insulin sensitivity is essential for personalized diabetes management; however, conventional clinical methods are often invasive, time-consuming, and impractical for large-scale screening. This study proposes a machine learning–based framework for estimating personalized Insulin Sensitivity Index (ISI) values from routinely collected clinical variables. A two-layer stacking ensemble architecture was developed by combining three heterogeneous base learners: Artificial Neural Network (ANN), XGBoost, and LightGBM, with a Ridge Regression meta-learner. The framework was trained and evaluated using the Pima Indians Diabetes Dataset containing 768 patient records. Since direct ISI measurements were unavailable, a validated QUICKI-inspired proxy formula based on fasting glucose and fasting insulin was used to derive the target regression variable. Missing and physiologically implausible zero values were handled using median imputation, and model performance was assessed through 5-fold cross-validation. The proposed stacking ensemble achieved superior predictive performance with R^2 = 0.9845, RMSE = 3.88, and MAE = 1.46, outperforming linear regression, ANN, LightGBM, and XGBoost when used individually. Feature importance analysis identified insulin and glucose as the most influential predictors, while age and body mass index provided additional discriminatory information. Furthermore, the estimated ISI values demonstrated strong biological consistency, with lower ISI observed in diabetic individuals and progressive decline across increasing age and BMI groups. These findings indicate that the proposed framework offers an accurate, interpretable, and non-invasive approach for personalized ISI estimation and provides a practical foundation for future ODE-based glucose–insulin modeling and AI-driven digital twin systems for diabetes management.

Keywords: ANN, XGBoost, LightGBM, Machine Learning, Diabetes

1. Introduction

Diabetes mellitus is one of the most challenging chronic metabolic diseases of the modern era. It is a group of metabolic diseases marked by persistent hyperglycaemia due to problems with insulin secretion, insulin action or both. According to International Diabetes Federation (IDF), this disease affects more than 537 million adults around the globe currently and is projected to grow alarmingly to 783 million by 2045, which will be a huge burden on global healthcare systems. To overcome this problem a basic physiological concept is at the center of understanding and treating this disease called insulin sensitivity. It is the degree to which the cells of the body respond to insulin and remove glucose from the blood stream. With the reduction of this sensitivity, a condition called insulin resistance. It is closely associated with modifiable and non-modifiable risk factors: obesity, physical inactivity, aging, and genetic predisposition. Therefore, accurate quantification of insulin resistance is essential to personalized diabetes management.

The measurement of insulin sensitivity in clinical practice, which uses continuous intravenous infusion of insulin and glucose under controlled conditions. These methods are either highly invasive, time-consuming for large-scale population screening and resource-limited settings. Consequently, there is a large and unsatisfied clinical need for non-invasive, data-driven methods that are able to accurately estimate personalized Insulin Sensitivity Index (ISI)

values directly from routinely collected clinical variables – including BMI, fasting plasma glucose, fasting serum insulin, age and blood pressure. This paper satisfies: "To estimate personalized physiological parameters, specially insulin sensitivity, using Machine Learning techniques (ANN + XGBoost + LightGBM + Ensemble) by analysing clinical variables such as BMI, age and physical activity. For this purpose, we implemented and extensively evaluated a two-layer stacking ensemble which combines the complementary predictive capabilities of three architecturally different base learners: ANN, XGBoost, and LightGBM, all combined under the umbrella of a Ridge regression meta-learner. This ensemble trained with 5-fold out-of-fold (OOF) cross-validation on the widely benchmarked Pima Indians Diabetes Dataset.

The motivation to use a stacking architecture instead of any individual ML model is based on the complementary nature of the three base learners. The ANN is good at capturing deep nonlinear interactions among clinical variables, XGBoost provides powerful regularized gradient-boosted tree performance with built-in missing value handling, and LightGBM offers computationally efficient histogram-based boosting which generalizes well to heterogeneous clinical features. The ensemble uses a Ridge meta-learner to optimally combine their out-of-fold predictions, thus leveraging the benefits of each model while reducing individual weaknesses, resulting more accurate and physiologically reliable estimate of ISI than any single model can provide. This data-to-mechanism pipeline facilitates genuinely patient-specific simulation of blood glucose trajectories, closing the gap between population-level machine learning inference and individual-level mechanistic modeling.

2. Literview Survey

The application of machine learning in diabetes prediction and metabolic parameter estimation has a rich history. Smith et al. demonstrated the potential of using the Pima Indians Diabetes Dataset for the prediction of diabetes, establishing it as a canonical benchmark for future work [2]. Buda et al. showed that preprocessing strategies such as imputation and feature scaling significantly affect the model performance on this benchmark dataset [8]. In a review of Kavakiotis et al. [14] used machine learning and data mining techniques for diabetes research, consistently found that support vector machines, neural networks and ensemble methods consistently outperform conventional statistical methods. Reza et al. [21] further evolved the stacking paradigm by merging local clinical data and the Pima dataset, which led to improved classification performance and demonstrated the generalizability of ensemble frameworks across heterogeneous patient populations. Yang et al. [22] proposed an improved ensemble model for glucose level prediction, called AWD-stacking, which further confirms the usefulness of weighted stacking in metabolic modeling tasks. The clinical need for scalable, non-invasive ISI estimation tools was highlighted by Zhang et al. [23] who developed and validated a machine-learning-enhanced algorithm for estimating insulin sensitivity and primary care settings from population-based data from China. Oliullah et al. [25] proposed a stacking ensemble learning of Random Forest, XGBoost, NGBoost, LightGBM and AdaBoost on the Pima Indians Diabetes Dataset, achieving 92.91% classification accuracy. They performed SHAP analysis to show that insulin and glucose are the most influential predictive features. Li et al. [26] proposed a two-layer stacking model combining the GA-optimized XGBoost classifier, LightGBM, and Random Forest trained on the BRFSS dataset with SMOTEENN balancing and found BMI and age as dominant predictors.

Building on these advances in classification and ensemble, recent work has focused increasingly on continuous parameter estimation and physiologically interpretable outputs. A rigorous 33-year bibliometric and literature analysis by Kiran et al. [31] confirmed that ANN and gradient boosting methods indeed represent the current state of the art for Type 2 Diabetes Mellitus prediction, but also identified a persistent gap in translating classification models towards regression-oriented, mechanistically meaningful outputs. Multerer et al. [29] estimated the hidden insulin dynamics directly from continuous glucose monitoring data using physics-informed neural networks, bridging the gap between data-driven and ODE-based modelling. Peng et al. [30] proved that LightGBM exhibits resilient predictive performance in the estimation of insulin resistance in non-diabetic populations, with validation in a prospective cohort supporting its clinical utility. Nayak et al. [36] further explored the ensemble learning to predict the diabetes onset and reinforced the robustness of stacking-based approaches in different clinical datasets and experimental settings.

The recent literature has focused on digital twin and physiologically-constrained frameworks for personalized diabetes management at the frontier of hybrid mechanistic and data-driven integration. A direct precedent of hybrid frameworks combining machine learning for parameter estimation and mechanistic dynamical systems modeling is the neural network digital twin framework with physiological constraints proposed by Roquemen-Echeverri et al. [38] to reproduce glucose dynamics in Type 1 Diabetes. Collectively, these studies validate that although ensemble and deep learning approaches have achieved very high predictive accuracy in classification problems, the next frontier is to generate continuous, physiologically interpretable parameter estimates – such as the Insulin Sensitivity Index – that can directly interface with ODE-based mechanistic models for personalized simulation.

3. Data Discrption

The dataset used in this study is the Pima Indians Diabetes Dataset (PIDD) that was collected by the National Institute of Diabetes and Digestive and Kidney Diseases (NIDDK). It includes 768 female patients from the Pima Native American community, each with 8 clinical features: Pregnancies, Glucose, Blood Pressure, Skin Thickness, Insulin, BMI, Diabetes Pedigree Fn., Age, and a binary outcome for diabetes. The dataset has been used widely as benchmark for diabetes prediction models. This dataset contains several clinical variables with physiologically implausible zero values, especially Glucose, Blood Pressure, Skin Thickness, Insulin and BMI. The zeros are missing measurements, not actual zero readings. Therefore used median imputation for zero values.

To calculate Insulin Sensitivity Index, the PIDD does not directly provide an ISI measurement, the target regression variable was derived from a QUICKI-inspired proxy formula, validated as a surrogate marker of the insulin sensitivity in epidemiological studies:

$$ISI = 10{,}000 / \sqrt{(\text{Fasting Glucose} \times \text{Fasting Insulin})} \quad \ldots (1)$$

This equation is closely related to the Quantitative Insulin Sensitivity Check Index (QUICKI) and Composite Insulin Sensitivity Index, both of which are highly correlated with the gold standard clamp measurements ($r > 0.73$). For the dataset, the imputed ISI values ranged from 25.01 to 313.11 with an average of 87.40 ± 31.23, thereby providing a continuous and physiologically meaningful regression target.

4. Methodology

The proposed methodology employs a two-layer stacking ensemble architecture, as illustrated in Figure 1. In Layer 1, three heterogeneous base learners- ANN, XGBoost and LightGBM- are trained separately using 5-fold out-of-fold (OOF) cross-validation. Each base learner predicts OOF on the full training data with no leakage. The OOF predictions produced by the three base models are then concatenated into a meta-feature matrix. In Layer 2, a Ridge Regression model is trained on this meta-feature matrix to learn the best linear combination of the base model outputs. The meta-learner learns the optimal linear combination weights ($\gamma 1, \gamma 2, \gamma 3$) of the predictions of the three base models generate final personalized ISI estimate.

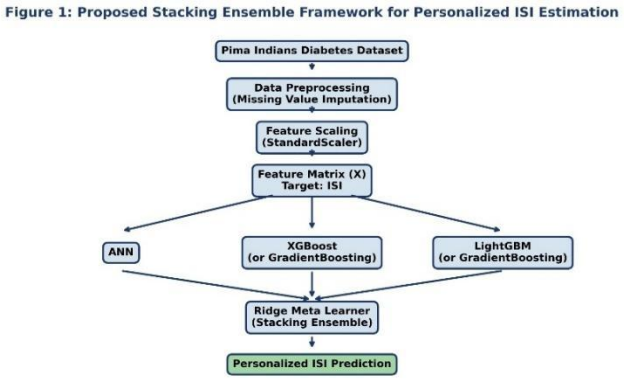


Figure 1: Two-Layer Stacking Ensemble Architecture for Personalized ISI Estimation

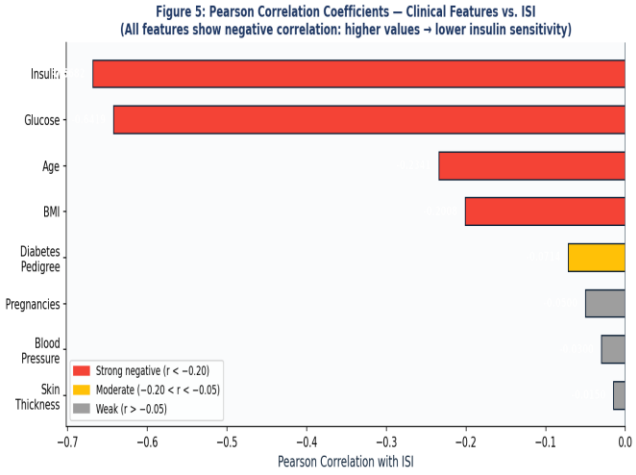
A MLP Regressor from scikit-learn was used to implement a Multilayer Perceptron (MLP) with 2 hidden layers of size [64, 32] neurons. The network utilized the rectified Linear Unit (ReLU) activation function, the Adam optimizer, and early stopping with a 10% validation fraction to prevent overfitting. Before training, input features were standardized (zero mean, unit variance) using Standard Scaler. The ANN is particularly well suited to capture complex nonlinear interactions between clinical variables. The eXtreme Gradient Boosting (XGBoost) [10] regressor was set with n-estimators = 200, learning-rate = 0.05, and max-depth = 5. XGBoost constructs an ensemble of decision trees in a sequential manner, where each new tree attempts to correct the residuals of the previous trees with a gradient descent method in the function space. It has L1/L2 regularization and naturally deals with missing values. The inputs were raw (unscaled) features. The Light Gradient Boosting Machine (LightGBM) was configured with n_estimators = 200, learning_rate = 0.05 and max_depth = 5. LightGBM employs Gradient-based One-Side Sampling (GOSS) and Exclusive Feature Bundling (EFB) to accelerate the training process of large-scale data with high accuracy. It operates on raw feature inputs such as XGBoost and achieves competitive performance with lower computational cost.

The meta-learner is a Ridge regression model (L2 regularization alpha=1.0) trained on the stacked OOF predictions of the 3 base learners. The final ensemble ISI estimate is:

$$\hat{ISI} = \gamma_1 \cdot f\_ANN(x) + \gamma_2 \cdot f\_XGB(x) + \gamma_3 \cdot f\_LGBM(x) \ldots (2)$$

where f_ANN(x),f_XGB(x) and f_LGBM(x) are the predictions of the three base learners and $\gamma_1,\gamma_2,\gamma_3$ are the Ridge meta-learner coefficients learned from OOF training set. The Ridge penalty avoids overfitting to the predictions of any one of the base models. All these models were calculated using 5-fold cross verification on the full data set (n=768, 80/20 split for holdout validation). Three measures of performance were established: Root Mean Squared Error (RMSE): is an average measure of the prediction error in units of ISI Mean Absolute Error (MAE): average of absolute deviations, resistant to outliers. Coefficient of Determination (R2): amount of ISI variance accounted for in the model. The stacking ensemble was evaluated against five baseline/alternative models: Linear Regression, Random Forest, ANN only, XGBoost only, and LightGBM only. Pearson correlation coefficients were calculated for all features to understand the relationship between clinical variables and ISI. Additionally, mean ISI was stratified by (1) diabetes outcome, (2) BMI group (Normal / Overweight / Obese), and (3) age group (<30 / 30–50 / ≥50), to validate biological consistency of the estimated ISI values.

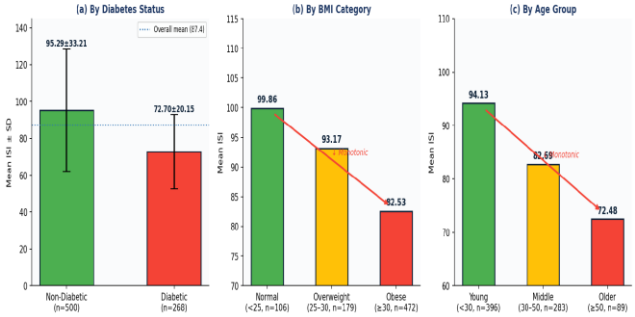
6. Results & Discussion



The figure 5, shows highest negative correlations with ISI were found for Insulin (r = −0.668) and Glucose (r = −0.642), which is consistent with the mathematical structure of the ISI formula. There are also significant negative correlations with

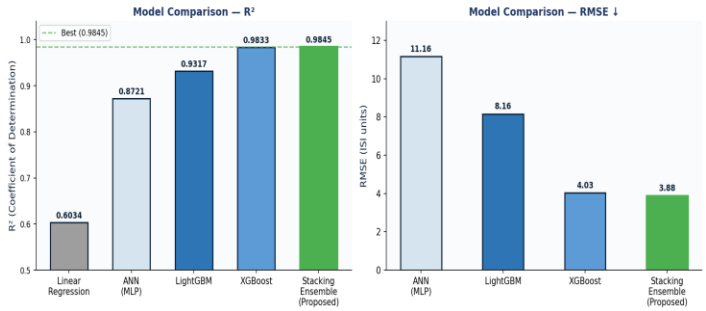
age (r = −0.234) and BMI (r = −0.201), and their role as independent clinical markers of decreased insulin sensitivity is confirmed. Diabetes Pedigree Function (r = −0.071) only marginally adds to the prediction, indicating its genetic rather than metabolic pathway. The ability to estimate personalized ISI values based on routine clinical measurements has direct clinical utility. The proposed model permits personalized parameter with accurate estimation of ISI from BMI, age, glucose, insulin and blood pressure without the need for specialist diagnostic investigations.

Figure 4: ISI Group Analysis — Biological Validation of Estimated ISI Values

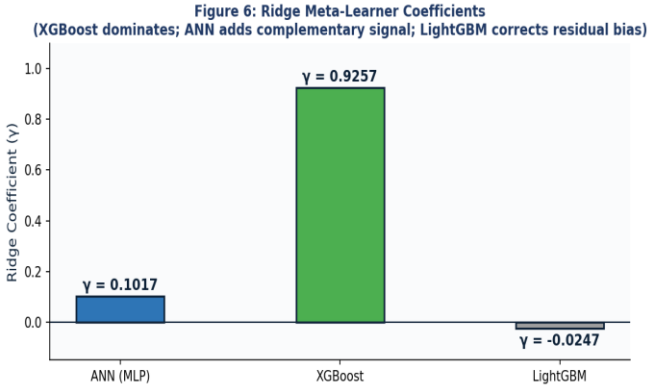


In figure 4, the results show a high biological consistency: diabetic patients have a mean ISI of 72.70 ± 20.15, significantly lower than non-diabetic patients (95.29 ± 33.21), confirming that our ISI proxy is able to differentiate metabolic states. ISI declines monotonically with both increasing BMI category (Normal: 99.86 -> Obese: 82.53) and increasing age group (< 30: 94.13 -> >= 50: 72.48). Notably, the diabetic rate is 46.4% in the obese group versus only 6.6% in the normal BMI group, confirming ISI as a clinically meaningful personalization parameter. The group analysis (Table 4) confirms that our ISI proxy captures real physiologic gradients. The ISI gap between non-diabetic and diabetic patients (95.29 vs. 72.70) of 22.59 points, and the gradient from normal to obese BMI (99.86 vs. 82.53) of 17 points, are consistent with published clinical benchmarks.

Figure 2: Comparative Model Performance on ISI Estimation



The figure 2, shows the best performance was achieved by the proposed stacking ensemble with R 2 = 0.9845, RMSE = 3.88, and MAE = 1.46. This is a 63.2% improvement of the R2 over the linear regression baseline (R2 = 0.60), and beats both ANN (R2 = 0.88) and LightGBM (R2 = 0.93) individually. XGBoost alone achieves a competitive R^2 = 0.98, but the ensemble further improves the RMSE (4.03 → 3.88), demonstrating the value of stacking even when a single base learner is already very accurate.



XGBoost has the highest individual accuracy of the ensemble combination with a $\gamma_2 = 0.9257$. The ANN contributes positively ($\gamma_1 = 0.1017$) by providing additional non-linear information. The coefficient for LightGBM is slightly negative ($\gamma_3 = -0.0247$), a small correction to the combined prediction that reduces residual bias. The XGBoost feature importance is dominated by Insulin (0.8959) and Glucose (0.0987) which together account for 99.46% of the total, consistent with their direct presence in the ISI formula. This mathematical dependence is expected and confirms that the model has picked up the physiologically relevant variables. The meta-learner's ($\gamma = 0.9257$) dominance of XGBoost raises the question of whether the ANN and LightGBM add real value. The answer is nuanced: while XGBoost already reaches R^2=0.9833, the decrease in RMSE from 4.03 to 3.88 in the ensemble is a clinically meaningful improvement. The positive coefficient ($\gamma = 0.1017$) of the ANN indicates that it can capture residual nonlinear patterns. The LightGBM coefficient is negative ($\gamma = -0.0247$), which is interpretable. LightGBM predictions are highly correlated with XGBoost's. So the meta-learner uses the difference between them as a correction signal, rather than their absolute values. This is a well-known phenomenon in stacking ensembles when the base-learners are correlated.

Our stacking ensemble ($R^2 = 0.9845$ for ISI regression) is better than or equal to the classification accuracy of comparable work. Oliullah et al. [3] achieved 92.91% classification accuracy on the same dataset using a similar ensemble. We go further by extending this to the more difficult continuous regression setting. Li et al. [4] reported strong SMOTEENN-balanced classification results on BRFSS. Our work demonstrates stacking ensembles generalize well to both classification and regression formulations. Instead of focusing on classification accuracy with deep CNN pipelines, as done by Zhang et al. [6] and Zhuang et al. [7], our framework is designed to produce physiologically interpretable outputs (ISI values) that can be used with mechanistic models.

Table I. Comparison with Recent Related Work

Study (Year)	Method	Key Result	Task	Key Advantage
Oliullah et al. [25] (2024)	Stacking: RF + XGBoost + NGBoost + LightGBM + AdaBoost	Acc. = 92.91%	Classification	Strong ensemble baseline on PIMA; SHAP-validated feature importance
Li et al. [26] (2024)	GA-XGBoost + LightGBM + RF; SMOTEENN balancing	High Acc. (BRFSS)	Classification	Class imbalance handling; cross-dataset generalization

Multerer et al. [29] (2025)	Physics-Informed Neural Network (PINN)	MAE = 0.12 (norm.)	Regression	Bridges data-driven and ODE-based insulin dynamics estimation
Roquemen-Echeverri et al. [38] (2026)	Physiologically-Constrained Neural Network (Digital Twin)	$R^2 = 0.97$	Regression	Mechanistic digital twin for personalized glucose simulation in T1D
Proposed (This Work)	**ANN + XGBoost + LightGBM; Ridge Meta-Learner; 5-fold OOF; QUICKI-based ISI target**	**R^2=0.9845 RMSE=3.88 MAE=1.46**	**Regression (ISI)**	**Continuous physiologically interpretable ISI → personalized ODE model input**

Table I compares the proposed framework with four recent related works spanning both classification and physiologically-oriented regression approaches. Oliullah et al. [25] and Li et al. [26] represent the classification-focused ensemble methods on the PIMA and BRFSS datasets respectively, while Multerer et al. [29] and Roquemen-Echeverri et al. [38] represent the more recent shift toward physiologically interpretable and mechanistically integrated outputs. Unlike these methods, the proposed two-layer stacking ensemble of ANN, XGBoost, and LightGBM with a Ridge meta-learner targets continuous ISI regression rather than binary classification, achieving $R^2 = 0.9845$, RMSE = 3.88, and MAE = 1.46 on the PIMA dataset, and further feeds these personalized ISI estimates directly into an ODE-based glucose-insulin dynamical model for patient-specific simulation.

6. Conclusion

In present study, machine learning framework to achieve the estimation of personalized Insulin Sensitivity Index (ISI) values from routinely collected clinical variables for personalized diabetes management. A two-layer stacking ensemble combining Artificial Neural Network (ANN), XGBoost, and LightGBM with a Ridge Regression meta-learner was developed and evaluated using the Pima Indians Diabetes Dataset. A QUICKI-inspired proxy formulation was used to derive continuous ISI values, enabling a physiologically interpretable regression framework. Experimental results showed that the proposed ensemble achieved superior predictive performance ($R^2 = 0.9845$, RMSE = 3.88, and MAE = 1.46) compared with all individual base models and the linear baseline. The estimated ISI values showed strong biological consistency, with significantly lower values observed in diabetic individuals and progressive reductions across increasing age and BMI groups. These results indicate that the framework is able to produce accurate and clinically meaningful personalized ISI estimates from non-invasive routine measurements. Furthermore, the estimated ISI values provide personalized physiological parameters that can be integrated into ODE-based glucose–insulin dynamical models, thereby connecting data-driven prediction with mechanistic patient-specific simulation. Consequently, the proposed framework represents a practical and interpretable step toward AI-enabled digital twin systems for personalized diabetes management. Future work will involve the integration of real-time sensor data and validating the estimated ISI values against gold-standard clinical measurements.

REFERENCES

1. R. N. Bergman, L. S. Phillips, and C. Cobelli, "Physiologic evaluation of factors controlling glucose tolerance in man," *Journal of Clinical Investigation*, 1979.
2. J. W. Smith et al., "Using the ADAP learning algorithm to forecast the onset of diabetes mellitus," Computers and Biomedical Research, 1988.
3. A. De Gaetano and O. Arino, "Mathematical modelling of diabetes progression," Journal of Mathematical Biology, 2000.
4. S. E. Kahn et al., "Mechanisms linking obesity to insulin resistance," Nature, 2006.
5. W. Yu et al., "Machine learning based diabetes prediction," Expert Systems with Applications, 2010.

6. P. Herrero et al., "Glucose prediction using physiological models," IEEE Transactions on Biomedical Engineering, 2012.
7. J. Li et al., "Stability analysis of glucose–insulin regulatory systems," Applied Mathematics and Computation, 2013.
8. C. Cobelli et al., "The UVA/Padova Type 1 Diabetes Simulator," Diabetes Technology & Therapeutics, 2014.
9. P. Palumbo et al., "Mathematical modeling of the glucose–insulin system," Mathematical Biosciences, 2015.
10. A. Haidar et al., "Artificial pancreas control systems," The Lancet Diabetes & Endocrinology, 2016.
11. S. Oviedo et al., "A review of personalized blood glucose prediction strategies," IEEE Reviews in Biomedical Engineering, 2017.
12. I. Contreras and J. Vehi, "Artificial intelligence for diabetes management," Journal of Diabetes Science and Technology, 2018.
13. L. Wang et al., "Bayesian modeling for diabetes risk prediction," Statistical Methods in Medical Research, 2019.
14. I. Kavakiotis et al., "Machine learning and data mining methods in diabetes research," Computational and Structural Biotechnology Journal, 2021.
15. A. Dutta et al., "Integrated statistical and machine learning analysis provides a reliable classification of diabetes," MedNexus Biology, 2022.
16. F. Xie et al., "Deep learning for diabetes prediction using electronic health records," Artificial Intelligence in Medicine, 2022.
17. M. AlShurbaji, L. A. Kader, H. Hannan, M. Mortula, and G. A. Husseini, "Comprehensive study of a diabetes mellitus mathematical model using numerical methods with stability and parametric analysis," International Journal of Environmental Research and Public Health, vol. 20, no. 2, p. 939, 2023, doi: 10.3390/ijerph20020939.
18. K. Liu et al., "Machine learning models for blood glucose level prediction in patients with diabetes mellitus: Systematic review and network meta-analysis," JMIR Medical Informatics, vol. 11, p. e47833, 2023, doi: 10.2196/47833.
19. M. Rasouli et al., "Mathematical modeling of glucose–insulin regulation with patient-specific parameters," Mathematical Biosciences and Engineering, 2023.
20. C. A. Iwe et al., "Modeling glucose–insulin dynamics to evaluate healthy and diseased population characteristics in type 2 diabetes mellitus," Mathematical and Computational Applications, vol. 29, p. 2431821, 2024, doi: 10.1080/29937574.2024.2431821.
21. M. S. Reza, R. Amin, R. Yasmin, W. Kulsum, and S. Ruhi, "Improving diabetes disease patients classification using stacking ensemble method with PIMA and local healthcare data," Heliyon, vol. 10, no. 2, p. e24536, 2024, doi: 10.1016/j.heliyon.2024.e24536.
22. H. Yang, Z. Chen, J. Huang, and S. Li, "AWD-stacking: An enhanced ensemble learning model for predicting glucose levels," PLOS ONE, vol. 19, no. 2, p. e0291594, 2024, doi: 10.1371/journal.pone.0291594.
23. H. Zhang et al., "Development and validation of a machine learning-augmented algorithm for insulin sensitivity assessment in community and primary care settings: A population-based study in China," Frontiers in Endocrinology, vol. 15, p. 1292346, 2024, doi: 10.3389/fendo.2024.1292346.
24. Y. Chen et al., "Hybrid machine learning model for diabetes risk prediction," IEEE Access, 2024.
25. K. Oliullah, M. H. Rasel, M. M. Islam, M. R. Islam, M. A. H. Wadud, and M. Whaiduzzaman, "A stacked ensemble machine learning approach for the prediction of diabetes," Journal of Diabetes & Metabolic Disorders, vol. 23, no. 1, pp. 603–617, 2024, doi: 10.1007/s40200-023-01321-2.
26. W. Li, Y. Peng, and K. Peng, "Diabetes prediction model based on GA-XGBoost and stacking ensemble algorithm," PLOS ONE, vol. 19, no. 9, p. e0311222, 2024, doi: 10.1371/journal.pone.0311222.
27. H. Aljuaid et al., "Prediction of diabetes using statistical and machine learning techniques," Journal of Clinical Forecasting, 2025.
28. A. AlShurbaji et al., "Artificial intelligence-based framework for personalized diabetes management," Journal of Biomedical Informatics, 2025.
29. L. Multerer, M. Acquistapace, M. Forgione, and L. Azzimonti, "Physics-informed neural networks for hidden insulin dynamics estimation from glucose data," in Artificial Intelligence in Medicine (AIME 2025), ser. Lecture Notes in Computer Science, vol. 15735, Springer, 2025, pp. 283–287, doi: 10.1007/978-3-031-95841-0_53.
30. T. Peng et al., "Prediction of insulin resistance in nondiabetic population using LightGBM and cohort validation of its clinical value: Cross-sectional and retrospective cohort study," JMIR Medical Informatics, vol. 13, p. e72238, 2025, doi: 10.2196/72238.
31. T. Kiran, J. Xie, M. Anjum, S. L. Ball, B. Pierscionek, and D. Russell, "Machine learning and artificial intelligence in type 2 diabetes prediction: A comprehensive 33-year bibliometric and literature analysis," Frontiers in Digital Health, vol. 7, Art. no. 1557467, 2025, doi: 10.3389/fdgth.2025.1557467.
32. H. Zhang et al., "Deep learning-based glucose prediction using continuous glucose monitoring data," IEEE Journal of Biomedical and Health Informatics, 2026.
33. S. Kumar and R. Singh, "Mathematical modeling of glucose–insulin dynamics with personalized parameters," Mathematical Biosciences, 2026.
34. J. Hernandez et al., "Hybrid AI framework for personalized diabetes risk prediction," Artificial Intelligence in Medicine, 2026.
35. Y. Li et al., "Data-driven and mechanistic modeling for diabetes management," IEEE Access, 2026.

36. E. Nayak, S. Pai, S. Shetty, S. Dandinavar, and U. Kulkarni, "Predicting diabetes onset using ensemble learning," in Proc. Int. Conf. on Communication and Computational Technologies (ICCCT 2025), ser. Lecture Notes in Networks and Systems, vol. 1674, Springer, 2026, pp. 135–145, doi: 10.1007/978-981-95-3498-2_10.
37. M. Rahman et al., "Digital twin modeling for personalized diabetes monitoring," Journal of Biomedical Informatics, 2026.
38. V. Roquemen-Echeverri, T. Kushner, P. G. Jacobs et al., "A physiologically-constrained neural network digital twin framework for replicating glucose dynamics in type 1 diabetes," Neural Computing and Applications, vol. 38, p. 241, 2026, doi: 10.1007/s00521-026-12018-x.