

The methodology included every phase, from data gathering and preprocessing through advanced model development and performance assessment. The experiment demonstrated how combining several machine learning and deep learning techniques might completely transform diabetes detection. While praising accomplishments, the methodology also highlighted flaws in the data collection process. The goal of the roadmap for future improvements was to use technology to better detect and treat diabetes, which would ultimately help people of all ages and backgrounds.

4. Results

The study emphasizes the potential impact on healthcare management as well as the effectiveness of the suggested methodology in detecting diabetes.

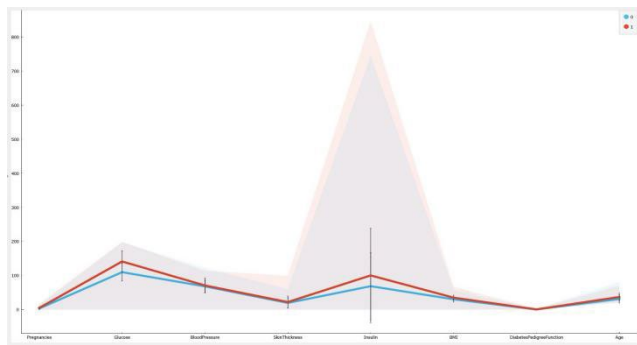


Figure 4. The line plot showcases a comprehensive comparison of attributes between Non-Diabetic (blue) and Diabetic (red) outcomes, effectively visualizing ranges, means, and error bars for each attribute.

Evaluation Metrics: After putting the various machine learning and deep learning algorithms into practice, a thorough review was carried out utilizing a number of metrics. These metrics, which offered numerical evaluations of the models’ performance, comprised, it was crucial to evaluate the models’ classification abilities using the metrics Area Under the Curve (AUC), Classification Accuracy (CA), F1-score, Precision, and Recall.

Performance Rankings: The integrated SGN model, which combines Stochastic Gradient Descent and Neural Networks, came up as the most promising performance among the many methodologies investigated. AUC, CA, F1, Precision, and Recall assessment scores were notable,

reaching values of 0.978, 0.967, 0.963, 0.966, and 0.967 for each parameter, respectively. AUC, CA, F1, Precision, and Recall scores for neural networks were notably excellent, coming in at 0.982, 0.944, 0.943, 0.943, and 0.944 respectively. Although it performed admirably, the Naive Bayes technique had somewhat lower results, with AUC, CA, F1, Precision, and Recall values of 0.966, 0.938, 0.937, 0.938, and 0.938, respectively. Similar results were obtained using the Random Forest technique,

which received scores of 0.951, 0.913, 0.910, 0.911, and 0.913 for AUC, CA, F1, and Precision, respectively.

Hyperparameter Optimization: To optimize the performance of each procedure, the project’s methodology included hyperparameter tuning. Different configurations for various models were produced through the refinement process. With a learning rate of 0.299 and a tree count of 100, Gradient Boosting adopted the “xgboost” method. Utilizing Rectified Linear Unit (ReLU) activation, the Adam solver algorithm, and a regularization parameter (alpha) of 0.0001, 300 neurons were added to neural networks in hidden layers. For classification and regression problems, stochastic gradient descent (SGD) was optimized with various loss functions. Lasso (L1) was used as the regularization model for SGD, with alpha set at 0.00005 and a constant learning rate of 0.01.

Future Implications and Conclusion: The project’s remarkable results demonstrate the legitimacy of the methodology chosen while also highlighting its potential to completely transform the diagnosis and treatment of diabetes. The SGN model’s crucial function, which is backed by Neural Networks and other methods, shows how many strategies were successfully combined to handle a significant healthcare concern. The study does, however, acknowledge some shortcomings, mostly attributable to irregularities in the database collection procedure. These flaws suggest chances for the accuracy and utility of the model to be improved and expanded. The initiative establishes a distinct course for the future when looking ahead. It seeks to improve the application’s accuracy, increase its accessibility by catering to a range of age groups, and enlarge the dataset to boost its capacity for prediction.

The findings of this project’s findings highlight the potential of combining different machine learning and deep learning approaches to develop an integrated system for diabetes detection. The SGN model’s performance, together with those of other strategies, highlights the potential of this ground-breaking application to transform diabetes diagnosis and care.

Table 1. The table demonstrating the values obtained after executing various models

MODEL	AUC	CA	F1	PRECISION	RECALL
SGN (Stochastic gradient descent + Neural networks)	0.978	0.967	0.963	0.966	0.967
Neural Network	0.982	0.944	0.943	0.943	0.944
Random Forest	0.951	0.913	0.910	0.911	0.913
Gradient Boosting	0.71	0.86	0.58	0.863	0.867
Naïve bayes	0.966	0.938	0.937	0.938	0.938

- diabetes mellitus disease using lifestyle data.” Predictive Modeling in Biomedical Data Mining and Analysis. Academic Press, 2022. 295-311.
- [14] Ge, Xiaochun, et al. ”Application of machine learning tools: Potential and useful approach for the prediction of type 2 diabetes mellitus based on the gut microbiome profile.” *Experimental and Therapeutic Medicine* 23.4 (2022): 1-10.
- [15] Ye, Jiancheng, et al. ”Predicting mortality in critically ill patients with diabetes using machine learning and clinical notes.” *BMC medical informatics and decision making* 20.11 (2020): 1-7
- [16] Alghamdi, Manal, et al. ”Predicting diabetes mellitus using SMOTE and ensemble machine learning approach: The Henry Ford Exercise Testing (FIT) project.” *PloS one* 12.7 (2017): e0179805.10.1109/AISP57993.2023.10134971.