

## Advancements in Predicting Diabetes Biomarkers: A Machine Learning Epigenetic Approach

**Authors :** James Ladzekpo

**Abstract :** Background: The urgent need to identify new pharmacological targets for diabetes treatment and prevention has been amplified by the disease's extensive impact on individuals and healthcare systems. A deeper insight into the biological underpinnings of diabetes is crucial for the creation of therapeutic strategies aimed at these biological processes. Current predictive models based on genetic variations fall short of accurately forecasting diabetes. Objectives: Our study aims to pinpoint key epigenetic factors that predispose individuals to diabetes. These factors will inform the development of an advanced predictive model that estimates diabetes risk from genetic profiles, utilizing state-of-the-art statistical and data mining methods. Methodology: We have implemented a recursive feature elimination with cross-validation using the support vector machine (SVM) approach for refined feature selection. Building on this, we developed six machine learning models, including logistic regression, k-Nearest Neighbors (k-NN), Naive Bayes, Random Forest, Gradient Boosting, and Multilayer Perceptron Neural Network, to evaluate their performance. Findings: The Gradient Boosting Classifier excelled, achieving a median recall of 92.17% and outstanding metrics such as area under the receiver operating characteristics curve (AUC) with a median of 68%, alongside median accuracy and precision scores of 76%. Through our machine learning analysis, we identified 31 genes significantly associated with diabetes traits, highlighting their potential as biomarkers and targets for diabetes management strategies. Conclusion: Particularly noteworthy were the Gradient Boosting Classifier and Multilayer Perceptron Neural Network, which demonstrated potential in diabetes outcome prediction. We recommend future investigations to incorporate larger cohorts and a wider array of predictive variables to enhance the models' predictive capabilities.

**Keywords :** diabetes, machine learning, prediction, biomarkers

**Conference Title :** ICDR 2024 : International Conference on Diabetes Research

**Conference Location :** Vancouver, Canada

**Conference Dates :** May 20-21, 2024