

Diabetes Detection System using Machine Learning

Soma Das[†], Sagarika Ghosh[‡], Sourav Kumar[‡], Gaurav Ganguly[‡], G Uma Devi[‡],

[†]Institute of Engineering and Management, University of Engineering and Management, Kolkata 700091, India

[‡]University of Engineering and Management Jaipur, Jaipur 303807, India

* corresponding author E-mail: soma.das@iem.edu.in, kumar.sourav@uem.edu.in

Abstract—The increasing prevalence of diabetes in contemporary lifestyles underscores the need for swift and accurate diagnostic methods. Early detection is pivotal for effective management and prevention of complications, such as heart attacks, blindness, and kidney diseases, associated with diabetes mellitus. This research explores the application of machine learning in expediting the diagnosis of diabetes, offering a preliminary assessment that not only aids individuals but also supports governmental efforts in surveying diabetes prevalence in specific regions. Traditional diagnostic procedures often entail prolonged waiting periods and visits to diagnostic centers, impeding timely intervention. Leveraging machine learning and artificial intelligence, this study proposes an automated approach to analyze medical datasets, facilitating early detection and diagnosis of diabetes. The integration of advanced technologies promises to revolutionize diagnostic processes, offering a more efficient and timely response compared to conventional manual methods. This study demonstrates that the GridSearchCV-XGBClassifier performs better than other machine learning models, including K-Nearest Neighbours, Support Vector Machine, Random Forest, AdaBoost, Gradient Boosting, Voting Classifier-XGB-SVM, and XGBoost, with an accuracy of 91.41%.

Keywords: Machine Learning, Supervised Learning, K Nearest Neighbor, Healthcare, Diabetes, Hyperparameter tuning, XGBoost

I. INTRODUCTION

The escalating prevalence of diabetes in contemporary lifestyles constitutes a pressing global health challenge, necessitating the development of swift and accurate diagnostic methodologies. Early detection is paramount for the effective management and prevention of complications such as heart attacks, blindness, and kidney diseases that are intricately linked with diabetes mellitus. This research explores the revolutionary possibilities of machine learning in accelerating the diabetes diagnosis process in response to this necessity. Through the deployment of machine learning algorithms, this study aims to provide a preliminary assessment that benefits individuals seeking early intervention and contributes to governmental endeavours in systematically surveying diabetes prevalence in specific regions.

After heart disease and cancer, diabetes is the third most common cause of mortality. It is a chronic illness marked by high blood glucose levels. The pancreas secretes the hormone insulin, which aids in the uptake of glucose by cells for use as fuel. However, excess glucose builds up in the blood and can cause a number of health issues when the body is unable to produce enough insulin or use it efficiently. Although there is presently no cure for diabetes, there are effective management

techniques that can be used to lessen its effects and preserve general health.

Machine Learning (ML) is becoming a central component of Medical Science, an important field in the healthcare industry. ML applications are revolutionising medical procedures, which span from drug development to diagnosis and prognosis. In particular, ML algorithms have proven remarkably accurate in identifying cancer patients who are at risk of dying soon, highlighting the prospective uses of these developments in the medical field.

Our study aims to predict whether a patient has diabetes, a problem statement in the medical science domain. The Centers for Disease Control and Prevention (CDC) estimate that over 34 million Americans alone have diabetes, with 1 in 5 of them being ignorant of their illness. This concerning figure highlights the necessity for an infrastructure that uses predictive models to help people who might be unaware that they have diabetes. The ultimate goal is to improve patient outcomes by making a proactive contribution to healthcare initiatives that can support diabetes early detection and management.

Timely intervention is hindered by traditional diagnostic procedures, which are typified by extended periods of inactivity and visits to diagnostic centers. On the other hand, our research suggests an automated method to examine large medical datasets using machine learning and artificial intelligence, which may help with diabetes early detection and diagnosis. The main objectives of this study are outlined as follows:

- Explore the application of machine learning in the early detection and diagnosis of diabetes.
- Investigate the potential of using medical datasets for developing predictive models in diabetes detection.
- Develop and train machine learning models to predict the likelihood of diabetes in individuals.
- Evaluate the performance of different machine learning algorithms in diabetes detection.
- Utilize data mining techniques to extract valuable insights from medical datasets.

The application of machine learning models in diabetes detection yielded promising results. This task provides a comprehensive overview of the performance metrics for different models. Notably, the XGB Classifier demonstrated the highest accuracy (91.41%) by hyperparameter tuning among the tested models, closely followed by the XGBoost (89.91%) and Support Vector Machine (88.96%). Precision, recall, and F1 score metrics were consistent across the models, showcasing their robustness in predicting diabetes.