Chapter 5
A Smart Healthcare Diabetes Prediction System Using Ensemble of Classifiers

Ayush Yadav
Vellore Institute of Technology, Chennai, India

Bhuvaneswari Amma N. G.
https://orcid.org/0000-0003-3660-380X
Vellore Institute of Technology, Chennai, India

ABSTRACT

Throughout the world, diabetes is a life-threatening disease. This research study aims to develop a smart healthcare machine-learning model for diabetes prediction. The dataset is pre-processed to handle missing data and outliers, and feature selection techniques are used to identify the most relevant variables for the model. An ensemble of classifiers is built by combining logistic regression, XGBoost, random forest, and support vector machine. The performance of the proposed model is assessed using metrics such as accuracy, precision, recall, and F1-score. The results show that the random forest algorithm outperforms other models in terms of accuracy, precision, recall, and F1 score. The model achieves an accuracy of 85%, indicating that it can correctly predict diabetes in 85% of cases. In conclusion, this study demonstrates the feasibility of using machine learning models for diabetes prediction based on patient data. The model can be further improved by incorporating more extensive and diverse datasets and exploring more advanced machine-learning techniques.

INTRODUCTION

Recent research by the World Health Organization (WHO) shows that the number of people with diabetes is rising worldwide, as is their death rate. Based on these trends, the World Health Organization projected that diabetes will be the sixth major cause of death by 2030. One of the illnesses with the fastest global growth rates is diabetes (Ihnaini et al., 2021). Diabetes is characterized as a group of metabolic

DOI: 10.4018/979-8-3693-0639-0.ch005

disorders that raise blood glucose levels in people. The following are the two underlying causes of elevated glucose levels: (1) insufficient insulin production by the human body, and (2) inadequate insulin response by the body's cells. The hormone released by the pancreas, insulin, aids in controlling blood sugar levels. The blood sugar level should remain within the recommended range (3.6–6.9 mmol/l or 70–120 mg/dl). Hypoglycemia is defined as glucose levels below 50 mg/dl, which can cause increased thirst, perspiration, seizures, and diabetic coma. Clinically relevant tasks in diabetes treatment include hypoglycemic prediction. As hypoglycemia can cause dangerous side effects including seizures and coma, it is important to detect it early on and take precautions. There is evidence that higher glucose levels (>200 mg/dl) represent hyperglycemia, which can cause long-term vascular problems such as diabetic retinopathy, neuropathy, and nephropathy (El-Sappagh et al., 2019).

To improve the quality of life, appropriate glucose regulation requires supervision. Type-1, Type-2, and gestational diabetes are the three subtypes of diabetes (Kibria et al., 2022). When the beta cells that produce insulin in the pancreas are destroyed by the immune system, type 1 diabetes results. Approximately 10% of people have type 1 diabetes. While it is challenging to prevent, a therapy that involves giving the body insulin externally is an option. type 2 diabetes, on the other hand, results from improper utilization of the insulin produced by the pancreas. Diabetes type 2 is the most prevalent and affects people older than 45 years in about 90% of instances. Patients with type 2 diabetes have a 2 to 4 times higher risk of developing heart disease. Gestational diabetes is a type of diabetes that affects pregnant women. Blood glucometers are employed to measure diabetes throughout certain periods. Continuous glucose monitoring devices, which offer a less intrusive way to record the patient's current glycemic level, are used to test diabetes continually. The majority of the body's organs, such as the kidneys, eyes, heart, and nerves, among others, are impacted by delayed diabetes identification. Consequently, getting a precise and timely diagnosis of diabetes is essential. Appropriate data analysis becomes essential for the diagnosis and interpretation of diabetes when handling data as a classification problem for machine learning (ML). Therefore, it is very valuable to employ artificial intelligence to accurately anticipate diabetes.

Consequently, getting a precise and timely diagnosis of diabetes is essential. Appropriate data analysis becomes essential for the diagnosis and interpretation of diabetes when handling data as a classification problem for ML. Therefore, it is very valuable to employ artificial intelligence to accurately anticipate diabetes. To estimate a patient's risk of acquiring diabetes, machine learning algorithms can examine their genetic predispositions, lifestyle choices, and medical history. Patients can better manage their diabetes by using machine learning algorithms to monitor glucose levels in real-time and send alarms when levels are outside of the usual range. ML models can be used to identify the most effective treatment strategies for individual patients based on their unique medical history and other factors. ML and AI can help healthcare professionals make more informed decisions about diabetes prevention, management, and treatment, which can improve patient outcomes and reduce healthcare costs. With the use of machine learning, patients may easily confirm their health in the early stages, and it will also help practitioners with future research. It can be applied to problems with both regression and classification (Ganie & Malik, 2022). Due to the classification challenge nature of diabetes prediction, we group individuals according to their diabetes status. Numerous ML techniques are useful for evaluating and synthesizing the data into pertinent knowledge from various perspectives. As a part of ML datasets must be prepared features must be selected and extracted, training and testing, and additional assessment. Clinical data, text data, and sensor data are just a few examples of the several types of data that are gathered utilizing various wearable technology, most of which is in raw form. Preprocessing, which includes managing