International Journal of Reliable and Quality E-Healthcare
Volume 12 • Issue 2

DDPIS: Diabetes Disease Prediction by Improvising SVM

Shivani Sharma, ABES Institute of Technology, India*
https://orcid.org/0000-0003-3381-269X

Bipin Kumar Rai, ABES Institute of Technology, India
https://orcid.org/0000-0002-9834-8093

Mahak Gupta, ABES Institute of Technology, India

Muskan Dinkar, ABES Institute of Technology, India

ABSTRACT

An illness that lasts longer and has continual repercussions is known as a chronic illness. Adults all across the world die as a result of chronic sickness. Diabetes disease prediction by improvising support vector machine is a platform that predicts diabetes based on the data entered into the system and offers reliable results based on that data. Earlier, the dataset consisted of a smaller number of features comprising the patients' medical details that were useful in determining the patient's health condition and was mainly focused on gestational diabetes, which only deals with pregnant women. In this work, the authors build a system that is more efficient than the previous system because of these reasons. It provides more accurate results by improvising the support vector machine, which includes more datasets and can predict the possibility of diabetes disease in both males and females.

KEYWORDS

Accuracy, Diabetes Disease Prediction, Machine Learning, Support Vector Machine

INTRODUCTION

Diabetes is one of the most widespread and fatal chronic diseases that harm the entire body system. The body of a diabetic patient has a high level of blood sugar (Lyngdoh et al., 2021). A person with a chronic illness has a condition that lasts longer and has ongoing consequences. One of the most significant disadvantages of chronic disorders is that they have a detrimental impact on people's standard of living. It is one of the most dangerous infections that may be discovered worldwide. This chronic illness costs the lives of adults all over the world.(Ahmed et al., 2021; Lai et al., 2019). Chronic diseases have a monetary burden attached to them and cost a lot of money for governments and people. As we all know, the operation cost is high and not every family can afford it. Two factors

DOI: 10.4018/IJRQEH.318090 *Corresponding Author

can cause diabetes: (1) the pancreas produces insufficient insulin, or (2) the body produces insufficient insulin. Only 5–10% of people with diabetes have this type of disease (Type-1) or (2). The produced insulin does not affect the cells (Type-2). Insulin is the hormone that controls the uptake of glucose from the bloodstream into most cells (muscles and fat cells). If there isn't enough insulin, glucose won't have the same effect as it usually does, and glucose won't be absorbed by the body cells that need it (Deberneh & Kim, 2021).

Diabetes mellitus is one of the leading causes of death in the United States. It requires detection and diagnosis at an early stage. Diagnosis of diabetes and interpretation of diabetes data is a significant categorization issue (Deberneh & Kim, 2021; Saeedi et al., 2019). Diabetes also afflicted approximately 463 million people aged 20 to 79 in 2019. (International Diabetes Federation-IDF) (Gulshan et al., 2016). Seventy-nine percent of the adult population live in low- and middle-income countries. According to estimates (IDF), approximately 700 million people will have diabetes by 2045 (Soni & Varma, n.d.). Every year, the number of instances grows, and the number of active cases continues to rise. Diabetes has become one of the most severe and rapid diseases to claim many people's lives worldwide, so it is essential to be concerned (Nayak & Pandi, 2021; Perveen et al., 2016). According to research, 70% of people in India suffer from this widespread disease, and 25% die due to early ignorance. The primary motivation for developing this project is so that a user can sit at their convenience and check their health (Vizhi & Dash, 2020; Zhou et al., 2020).

We developed the platform diabetes disease prediction by improvising a support vector machine to overcome diabetes disease in earlier stages. As we all know, in the competitive economic development environment, people are so busy making money and improving their lifestyle and future that they are not concerned about their health. The leading causes of ignorance are that they do not have time. They are so busy with their work that they neglect their health and do not go for regular body check-ups, which are essential for monitoring an individual's health to be free from any disease harmful to their body that may cost their life. People have become so preoccupied with their daily lives that they have no time to schedule appointments and consult a doctor, resulting in fatal conditions. Our diabetes prediction system helps individuals to predict the possibility of diabetes without taking more of their time. Whenever they are free from work, they can immediately check the likelihood of diabetes. They can consult the doctor for further treatment or assistance if the results are positive.

Machine learning is a kind of artificial intelligence (AI) that lets software applications become more accurate and efficient in predicting outcomes. ML algorithm uses historical data to anticipate improved output values (Kaur, 2019; Kumar et al., 2022).

Support Vector Machine, i.e., SVM, is a machine learning algorithm based on supervised learning. SVM can be used for classification and regression complications but mainly for classification problems. The main aim of the support vector machine is to find a hyperplane in n-dimensional space (where n is the total number of attributes). The dimension in the hyperplane depends on the number of attributes used (Pranto et al., 2020; Rani, 2020)

Let's consider an example where we have two independent variables x1, x2 and one of them is dependent on either the blue or red. From the first figure, we now have to choose the best line to segregate our data points.(Shafi & Ansari, 2021)

We choose the hyperplane whose distance from it to the nearest data point on each side is maximized. If such a hyperplane exists, it is known as the maximum-margin hyperplane/hard margin. So, from the above figure, we choose L2.

LITERATURE REVIEW

Arwatki Chen Lyngdoh et al. (2021) compared machine learning algorithms such as KNN, SVM, DT, RF, and Naive Bayes. They compared all the classifiers and obtained the highest accuracy of