Predicting Diabetes Using Machine Learning  
and Optical Character Recognition

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Joderick Sherwin J**  *Dept. of Computer Science and Engineering* | **Keerthana MG**  *Dept. of Computer Science and Engineering* | | **Hari Balaji JC**  *Dept. of Computer Science and Engineering* | | **Jayashrinidhi V**  *Dept. of Artificial Intelligence and Machine Learning* |
| **Jyothi Sakthi H**  *Dept. of Computer Science and Engineering* | | **Hariharaviswanathan**  *Dept. of Computer Science and Engineering* | | **Jayajothi Kumar**  *Dept. of Computer Science and Engineering* | |

**Abstract- Diabetes, also known as diabetes mellitus, is a group of common** [**endocrine diseases**](https://en.wikipedia.org/wiki/Endocrine_disease) **characterized by sustained high sugar levels in the blood or hyperglycaemia. There are two main causes of diabetes. The first cause can be due the pancreas not producing enough insulin, the second cause is the cells of the patient’s body will not respond properly to the insulin produced by the pancreas. Over time, this can cause serious health problems, such as** [**heart disease**](https://www.cdc.gov/diabetes/library/features/diabetes-and-heart.html)**,** [**vision loss**](https://www.cdc.gov/diabetes/managing/diabetes-vision-loss.html)**, and** [**kidney disease**](https://www.cdc.gov/diabetes/managing/diabetes-kidney-disease.html)**. The** [**considered**](https://www.thesaurus.com/browse/considered) **algorithm analysis the medical report of the patient and predicts if the person has diabetes. The Machine learning model compares and analysis the probabilities of the patient being diabetic with the past data and produces the output.**

1. Introduction

In the India, 840 million adults—more than 6 in 10—have [prediabetes](https://www.cdc.gov/diabetes/basics/prediabetes.html). Most of them don’t know they have it. With prediabetes, blood sugar levels are higher than normal, but not high enough for a type 2 diabetes diagnosis. Prediabetes raises your [risk](https://www.cdc.gov/diabetes/basics/risk-factors.html) for type 2 diabetes, heart disease, and stroke. It can lead to diabetes and other chronic diseases if not treated early.

To predict diabetes, various laboratory tests are performed on the patient by laboratories. This can include the blood glucose levels, the blood pressure, the skin thickness, the insulin levels as produced by the pancreas, the age and the body mass index of the patient. This data can then be fed into the machine learning model to get the output. As it is not easy to get an accurate prediction, in the proposed system we have used eight machine learning models to get multiple results and the best-case output is derived by further compiling and analysing the eight outputs to get the result. For our model we have provided 768 data sets with varying data parameters to increase the precision of the output of the machine learning models there by getting an accurate result. The probability for a person to be categorised as diabetic depends on several parameters given below. When the values of the different parameters are keyed into the algorithm the final output will indicate on the diabetic status of the patient. This will help in ruling out any anomaly in the diagnosis.

1. Related Work