

# Diabetes Prediction Using Machine Learning Techniques

**Abhishek Khari**

Department of computer science & Engg, G L Bajaj Institute of Technology & Management, Gr Noida

**Anant Vajpeyi**

Department of computer science & Engg, G L Bajaj Institute of Technology & Management, Gr Noida

**Ankit Agrawal**

Department of computer science & Engg, G L Bajaj Institute of Technology & Management, Gr Noida

**Sargam Tiwari**

Department of computer science & Engg,  
G L Bajaj Institute of Technology & Management, Gr Noida

## ABSTRACT

Diabetes mellitus or just diabetes is an ailment caused because of the expansion level of blood glucose. Diabetes is an interminable malady with the possibility to cause an overall human services emergency. In any case, early forecast of diabetes is very testing task for clinical specialists because of complex relationship on different factors as diabetes influences human organs, for example, kidney, eye, heart, nerves, foot and so on. AI is a developing logical field in information science managing the manners by which machines gain as a matter of fact. One such assignment is to help make expectations on clinical information. The point of this paper is to think about precision of various calculations and build up a framework which can perform early expectation of diabetes for a patient with a higher exactness.

## Keywords

AI, Data Pre-processing, Classification, Machine Learning.

## 1. INTRODUCTION

Diabetes is the quickly developing malady among the individuals even among the youths. Diabetes is brought about by the expansion level of the sugar (glucose) in the blood. The diabetes can be arranged into two classifications, for example, type 1 diabetes and type 2 diabetes. Information expository is a procedure of looking at and distinguishing the concealed examples from huge measure of information for making inferences. In social insurance, this scientific procedure is completed utilizing AI calculations for breaking down the clinical information to assemble AI models to do the clinical judgments. The essential point of this paper is to utilize understanding clinical history and different AI calculations to characterize people with or without a conclusion of type II diabetes. And then look at the exactness of those calculations. A precise characterization calculation could fill in as a clinically valuable instrument to propel the avoidance and control of the ailment by enabling specialists to recommend precaution measures to in danger patients and make prior move to arrange indicative tests. The contributions to our calculations will be numeric highlights that we will created dependent on the clinical record information. We will utilize bolster vector machines, Random timberland and calculated relapse based calculations to foresee the twofold result of a positive or negative finding of type II diabetes.

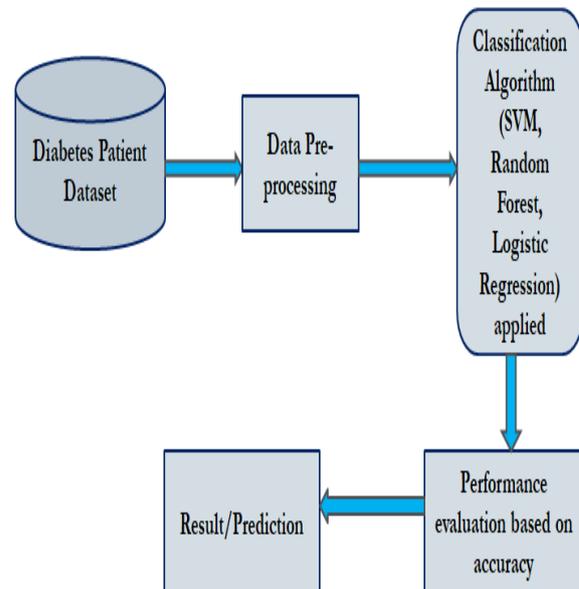
## 2. OBJECTIVE AND SCOPE

The purpose of this broadside is to compare accuracy of dissimilar algorithms and mature a system which can achieve initial forecast of diabetes for a patient with a advanced precision. And depending on that accuracy choosing the algorithm which performs best for predicting whether the patient has or will have diabetes or not.

Support vector machines, Random forest and logistic regression based algorithms are used to predict the outcome and application is built with the help of Tensorflow and Flask.

## 3. METHODOLOGY

Planned technique is concise in character under in the form of prototypical illustration. The character displays the flow of the research showed in assembling the typical.



**Figure 1: Proposed Method of Diabetes Prediction Using Machine Learning Techniques**

## 4. RESULTS & DISCUSSION

In this paper, three learning techniques were explored to predict Diabetes. We made the following observations after significant analysis

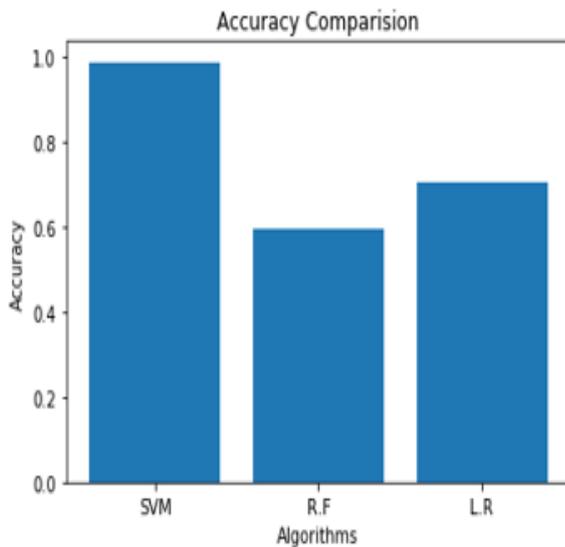
## Diabetes Prediction Using Machine Learning Techniques

- We utilized three diverse arrangement calculations to anticipate the nearness or nonattendance of a kind II diabetes finding dependent on highlights made from quiet clinical records. Among the three learning calculations utilized, the SVM model played out the best, trailed by strategic relapse and previous random forest.
- The data set is highly diverse and contained significant amounts of invalid entries. Pre-processing is the key to improve the accuracy.

[10] Tensorflow Estimator,  
<https://www.tensorflow.org/guide/estimators>

### 5. CONCLUSION

- Between the three learning systems recycled, the SVM model achieved the best, trailed by logistic regression and previous random forest.



**Figure 2: Accuracy Comparison of Diabetes Prediction Using Machine Learning Techniques**

- Android application can be made for users so that they can predict whether he/she has diabetes or not at any time.
- Web application can be deployed on the cloud so that anyone can access the url and make predictions.

### REFERENCES

- [1] M. W. Andy Liaw, "Classification and regression by randomForest," R news 2.3, pp. 18-22, 2002.
- [2] P. J. Lin, "Clinical Multiple Chronic Conditions in Type 2 Diabetes Mellitus: Prevalence and Consequences," vol. 21, no. 1, 2015.
- [3] Oliver Bear Don't Walk IV, David Joosten, Tim Moon , "Predicting the Diagnosis of Type 2 Diabetes Using Electronic Medical Records" , 2014
- [4] Duyun Chen, Yaxuan Yang , and Junrui Zhang . " Methods for Predicting Type 2 Diabetes", 2015
- [5] Scikit-learn: <http://scikit-learn.org/stable/>
- [6] A Library for Large LinearClassification: <http://www.csie.ntu.edu.tw/~cjlin/lib/linear/>
- [7] A Library for Support Vector Machines <https://www.csie.ntu.edu.tw/~cjlin/libsvm/>
- [8] L. Breiman. Random forests. Machine learning, 45(1):5–32, 2001.
- [9] Flask: <http://flask.pocoo.org>