

Diabetes detection using machine learning

Submitted to: Bhupendra sir
Submitted by: Jainam Jain (20BEC043)




Introduction

- Diabetes is a chronic disease that occurs when the body cannot produce enough insulin or cannot use insulin effectively. According to the World Health Organization, the number of people with diabetes has risen from 108 million in 1980 to 422 million in 2014, and it is expected to continue to increase in the future.
- The detection of diabetes is a critical aspect in one's life. Early we get to know about disease, less damage it can give to the body.
- Traditionally, the diagnosis of diabetes has been based on blood tests, including fasting blood sugar, oral glucose tolerance tests, and HbA1c levels. However, these tests can be time-consuming, costly, and require a skilled healthcare professional to administer. This is where machine learning comes in.



Why Machine learning in disease detection

- Machine learning has the potential to revolutionize the way we detect and manage diabetes. By analyzing large datasets, machine learning algorithms can learn to identify patterns and relationships that may be too complex for humans to detect.
- These algorithms can then be used to develop predictive models that can accurately identify individuals at risk of developing diabetes or those who may already have the disease.
- In this presentation, I will discuss the use of machine learning algorithms for diabetes detection, including the types of data used, the different machine learning algorithms employed, and their performance



So basically i have used a dataset and 4 different machine learning models to predict, whether a person is diabetic or not.

Following are the traning models used by me to predict the diabetes

- Support vector machine
- Random forest
- Naive bayes classifier
- Logistic regression





Description of Dataset

- There are in total 768 rows and 8 columns in the csv data set
- The Diabetes dataset has features based on which model evaluates the output
- The features or inputs to the dataset are number of pregnancies, glucose concentration, diastolic blood pressure, insulin, body mass index, age.
- The output simply states whether the person is diabetic or not
- If the person is diabetic the output is 1 else the output is zero.

Source of dataset : Kaggle (website: [Pima Indians Diabetes Database | Kaggle](https://www.kaggle.com/uciml/pima-indians-diabetes))


- Different machine learning model gives different accuracy rate
- The methodology or rather the algorithm to train the model and test it is similar for all the machine learning models





Models and their accuracy

Machine learning model	Accuracy
Naive bayes classifier	73.2%
Random forest	73.6%
Logistic regression	74.8%
Support vector machine	75%



There are some advanced of machine learning learning algorithms which have lead to 90% of accuracy

GA, PSO, k-means, and HR-kmeans are all algorithms used in machine learning and data mining for clustering or optimization tasks. Here is a brief overview of each algorithm:

1. Genetic Algorithm (GA): GA is a metaheuristic algorithm that is inspired by the process of natural selection. It is used for optimization tasks, including clustering. GA works by generating a population of potential solutions (chromosomes), and then applying selection, crossover, and mutation operations to evolve the population towards better solutions.
2. Particle Swarm Optimization (PSO): PSO is another metaheuristic algorithm that is used for optimization tasks. It is inspired by the collective behavior of swarms of particles. In PSO, a population of particles searches the solution space to find the optimal solution by adjusting their positions and velocities based on their own best position and the best position of their neighbors.

Source: [Research paper \(Improving the Accuracy of Diabetes Diagnosis Applications through a Hybrid Feature Selection Algorithm\)](https://www.ncbi.nlm.nih.gov/pmc/articles/) <https://www.ncbi.nlm.nih.gov/pmc/articles/>



Conclusion:

the use of machine learning for diabetes detection has shown promising results. While achieving an accuracy of 75% using the support vector machine algorithm is a significant step forward, some more advanced machine learning algorithms have achieved up to 90% accuracy. Nevertheless, even with 75% accuracy, machine learning can play a significant role in diabetes detection, making it a valuable tool in healthcare.

The real-life importance of using machine learning for diabetes detection cannot be understated. Achieving higher accuracy is desirable, the use of machine learning for diabetes detection can significantly improve patient care and outcomes, making it an essential tool in modern healthcare.



Thank you for patiently observing the
presentation