

# WEBAPP FOR DISEASE DETECTION PROJECT

PROJECT SYNOPSIS  
OF MAJOR PROJECT

BACHELOR OF TECHNOLOGY  
COMPUTER SCIENCE

SUBMITTED BY  
ZUNAIRA KHAN  
&  
NOOR YESHFEEN  
DECEMBER 2021



SUPERVISOR  
ASFIA AZIZ  
&  
SHAH IMRAN ALAM

**JAMIA HAMDARD**  
**SCHOOL OF ENGINEERING SCIENCES AND**  
**TECHNOLOGY HAMDARD NAGAR, NEW DELHI, DELHI**  
**110062**

- **ABSTRACT: -**

Our project object is to detect whether patients have diabetes, breast cancer and coronavirus or not by given a number of features from patients. The motivation of our project is to save human resources in medical canters and improve accuracy of diagnosis. In our project we use different methods to detect diseases such as Logistic Regression, SVM

- **INTRODUCTION: -**

Our problem is that we want to predict whether patients have diabetes, breast cancer and coronavirus by given some features of users. This is important to medical fields. If such a prediction is accurate enough, we can not only avoid wrong diagnosis but also save human resources. When a patient without these diseases is diagnosed with these, he will fall into unnecessary panic and when a patient with these diseases is not diagnosed with these diseases, he will miss the best chance to cure his disease.

Such wrong diagnosis is painful to both patients and hospitals. With accurate predictions, we can solve the unnecessary trouble. Besides, if we can apply our machine learning tool into medical prediction, we will save human resource because we do not need complicated diagnosis process in hospitals. (Though it is a very long way to go.) The input to our algorithm has features with number values.

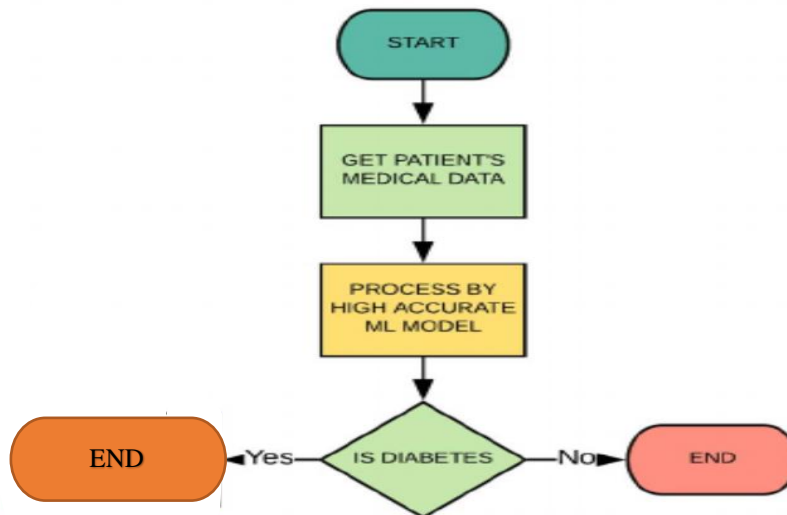
The Machine learning algorithms usually find out the hidden pattern from the large dataset and find out the desire approximate final result. Machine learning is a subfield of AI, And Machine learning algorithms can classify into three categories such as Supervised Learning, Unsupervised Learning and Reinforcement Learning. In our system, we use supervised learning algorithms for testing out accuracy among some sort of popular Machine Learning (ML) algorithms. Supervised Learning algorithms learn the pattern from pre-existing data and try to predict new result based on the previous learning.

ML algorithms are to identify existing data like probability-based, function-based, rule-based, tree-based, instance-based, etc. Different Machine Learning algorithms are introduced using various data mining algorithms for assisting medical experts. The effectiveness of the decision support system is recognized by its accuracy. So the main aim of building a decision support system is to predict and diagnose a particular disease with a maximum degree of accuracy . In this system, we use the pre-existing data sets

- **PROPOSED ACHITECHTURE: -**

Supervised Learning algorithms learns the pattern from pre-existing data and try to predict new result based on the previous learning. ML algorithms are used to identify existing data like probability-based, function-based, rule-based, tree-based, instance-based, etc. Following Fig.1. indicates the ample architecture of our proposed Model. According to our model patients are

required to provide their medical data for successful diagnosis of their diabetes test.



## • **METHODOLOGY: -**

### **1. Support Vector Machine (SVM)**

Support vector machine is a supervised learning algorithm mainly used for classification problem. Overfitting derives an ML model to misclassify the data from a given dataset. In that case SVM can prevent overfitting nature from samples data and produce better accuracy. SVM possess a linear hyperplane with a margin which divides the dataset into positive and negative samples

SVM select the hyper planes with maximum possible distance. The SVM decision boundary in mathematical expression as follows:

$$\min_{\theta} \frac{1}{2} \sum_{j=1}^n \theta_j^2$$

Where,

$$\begin{aligned} \theta^T x^{(i)} &\geq 1 \quad \text{if } y^{(i)} = 1 \\ \theta^T x^{(i)} &\leq -1 \quad \text{if } y^{(i)} = 0 \end{aligned}$$

### **2. Logistic Regression**

Logistic Regression is a supervised learning that computes the probabilities for classification problems with two outcomes. It can also be extended to predict several classes. In Logistic Regression model, we apply the sigmoid function, which is

$$\sigma(z) = \frac{1}{1 + e^{-z}}.$$

This function successfully maps any number into the value between 0 and 1 and we can regard this value as the probability of predicting classes. For example, we have two classes and they are presence of heart disease and absence of disease. If we set the threshold as 0.5, applying the sigmoid function gives us a value of 0.7, which means the man has the 70% probability of having heart disease so we will predict that he has heart disease.

- **FACILITIES REQUIRED FOR PROPOSED WORK: -**

**A) Hardware Components:**

Intel® Core™ i5-2450M processor, 4GB3 DDR3 SDRAM at 1600MHz, 500GB  
5400RPM SATA Hard Drive

**B) Software Requirement:**

Form Design (HTML 5.0, CSS, JS)  
Coding (Python)  
Framework (Flask)  
Reporting Tool (Data Report)  
Jupyter Notebook

**C) Server Requirement:**

OS (LINUX)  
4GB ram  
500 GB HDD