Diabetes Prediction Using Artificial Neural Network

Submitted By

Bonnhi Shikha Parna
ID:1903610201820

Md Hafizul Islam
ID:1903610201822

Submitted To

Mr. Faisal Ahmed
Assistant Professor
Department of CSE
Premier University Chattogram

Introduction

- Diabetes is one of the biggest health problems that affect millions of people across the world. Uncontrolled diabetes can increase the risk of heart attack, cancer, kidney damage, blindness, and other illnesses.
- Our goal is we will be predicting that whether the patient has diabetes or not on the basis of the features.

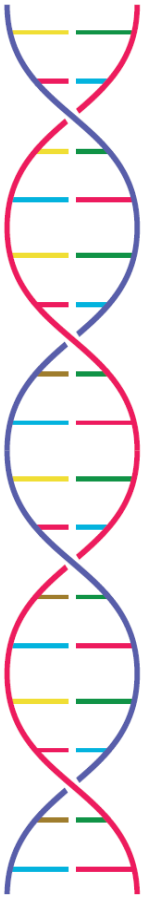
Objective of Project

The objective of diabetes prediction using Artificial Neural Network (ANN) is to develop a model that can accurately predict the presence or absence of diabetes in a patient based on their input features.

❑The Goal of our prediction

Positive result = **1**, the patient will be diagnosed with diabetes.

Negative result = **0**, the patient will not be diagnosed with diabetes.



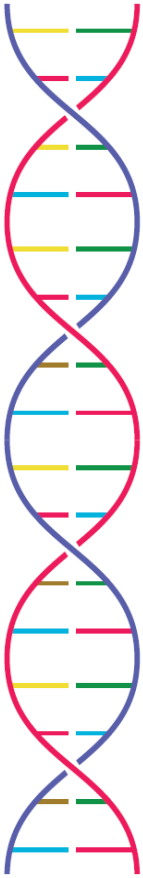
Dataset

	Pregnancies	Glucose	BloodPressure	SkinThickness	Insulin	BMI	DiabetesPedigreeFunction	Age	Outcome
0	6	148	72	35	0	33.6	0.627	50	1
1	1	85	66	29	0	26.6	0.351	31	0
2	8	183	64	0	0	23.3	0.672	32	1
3	1	89	66	23	94	28.1	0.167	21	0
4	0	137	40	35	168	43.1	2.288	33	1

- ❑Pregnancies: To express the Number of pregnancies.
- ❑Glucose: To express the Glucose level in blood.
- ❑Blood Pressure: To express the Blood pressure measurement.
- ❑Skin Thickness: To express the thickness of the skin.
- ❑Insulin: To express the Insulin level in blood.
- ❑ BMI: To express the Body mass index.
- ❑ Diabetes Pedigree Function: To express the Diabetes percentage.
- ❑ Age: To express the age.
- ❑ Outcome: To express the final result 1 is Yes and 0 is No.

Data Source:

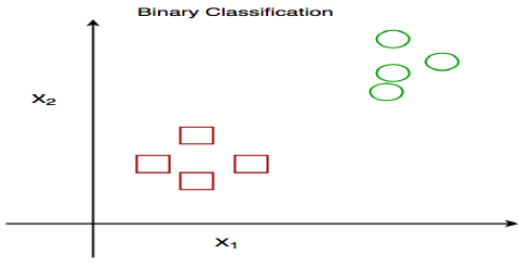
This dataset is originally from the National Institute of Diabetes and Digestive and Kidney Diseases.We collected it from Kaggle.Dataset Link[26.01.2023]



Exploratory data analysis basis of neural network

- There are seven features(Pregnancies,Glucose,BloodPressure,SkinThickness,Insulin. DiabetesPedigreeFunction,Age) and one label(Outcome) in our dataset. Therefore, we must use all of those features.

- This is binary classification problem.So,there are two classes.



Preprocessing

This dataset contains missing values. So we imputed missing values for few selected attributes like Glucose level, Blood Pressure, Skin Thickness, BMI and Age because these attributes cannot have values zero. Then we scale the dataset to normalize all values.

- **The Min-Max scaling is done using:**

$$X_i = \frac{X - Xmin}{Xmax - Xmin}$$

- If the scaled value is above 0.5, the result will be 1, that is, diabetes & if it is below 0.5, it predicts that diabetes will not occur.

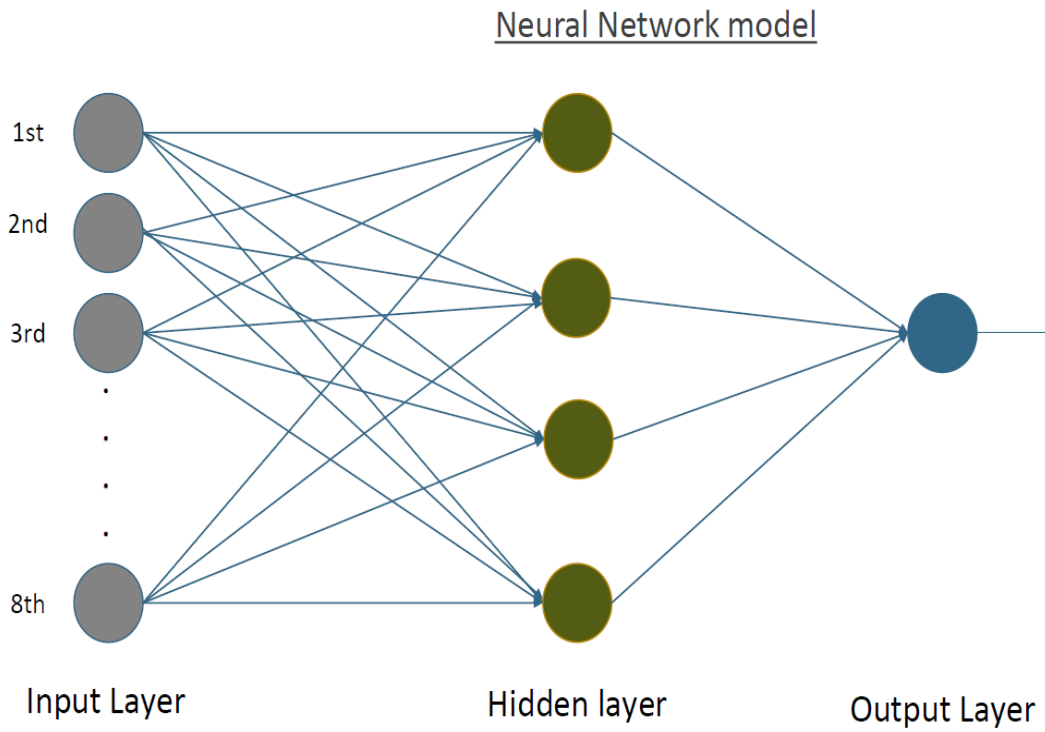


Fig 1: Neural Network Model

Activation Function

- Sigmoid / Logistic Activation Function-

  Activation function, $A = 1/(1 + e^{-X})$

  Value Range : 0 to 1
- Loss Function :

  Binary Cross-Entropy / Log Loss

$$\text{Log loss} = \frac{1}{N} \sum_{i=1}^{N} -(y_i * \log(p_i) + (1-y_i) * \log(1-p_i))$$

Graphical Representation:

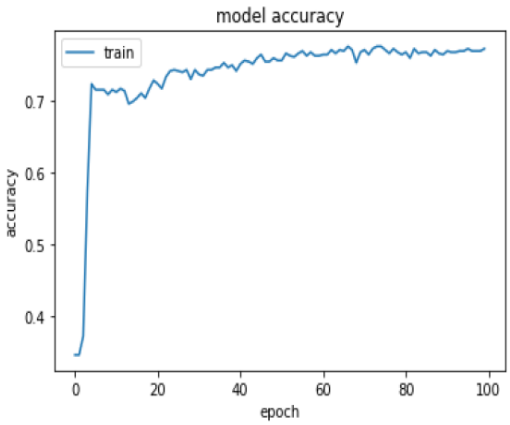


Fig 3: Accuracy of training data

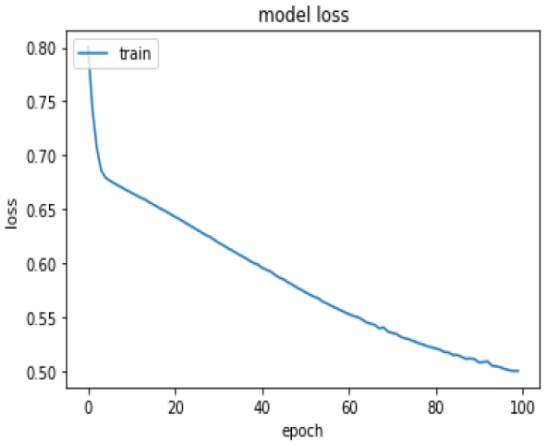


Fig 2: Loss of training data

```
Evaluation:
 loss: 0.5423
 accuracy: 0.7597
```

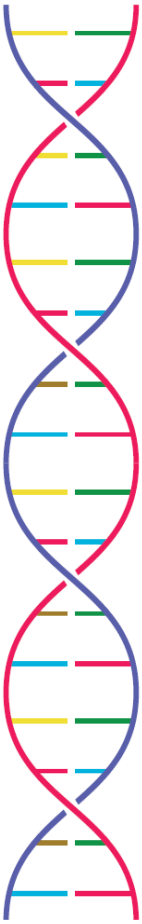
Results

Classification Report :

Classifier	precision	recall	F1-score	support
0	0.81	0.96	0.94	53
1	0.67	0.90	0.92	39
accuracy			0.76	154
macro avg	0.74	0.74	0.74	154
weighted avg	0.76	0.75	0.76	154

Table 1. Classification report.

Conclusion



One of the important real-world medical problems is the detection of diabetes at its early stage.
Experimental results determine the adequacy of the designed system with an achieved accuracy of **75%** using Decision Tree algorithm.We have used KNN Algorithm and Decision Tree to predict the diabetes where result is classified into **Yes** or **No** . We compared the testing data and actual data to get the accuracy of our project.

Thank you for your attention!

Any Questions or Comments?