1. **Diabetes Detection with Machine Learning**

# I. Dataset

1. A publicly available dataset from Kaggle (<https://www.kaggle.com/datasets/ankitbatra1210/diabetes-dataset>) was used to train the models.
2. Out of 33 attributes, 9 were selected. These attributes have been described in brief :

## i. Insulin Levels

1. Measured in microunits per milliliter.

## ii. Age

## iii. BMI

## iv. Waist Circumference

1. Measured in inches.

## v. Cholesterol Levels

## vi. Blood Glucose Levels

## vii. Pancreatic Health

## viii. Neurological Assessments

## ix. Glucose Tolerance Test

1. Out of 13 types of diabetes, 4 major types were chosen, and the following mapping was applied to facilitate working with models that took only numberic values as input :
2. 'Prediabetic': 0,
3. 'Type 1 Diabetes': 1,
4. 'Type 2 Diabetes': 2,
5. 'Type 3c Diabetes (Pancreatogenic Diabetes)': 3

After this filtering, data from 21,539 patients were available.

# **II. Methodology**

For model training, 60% of the data was utilised, and the remaining 40% was used for testing.

6 different models were used, and each of the results have been shown in the next section.

## i. Logistic Regression

# References

1. 1.