



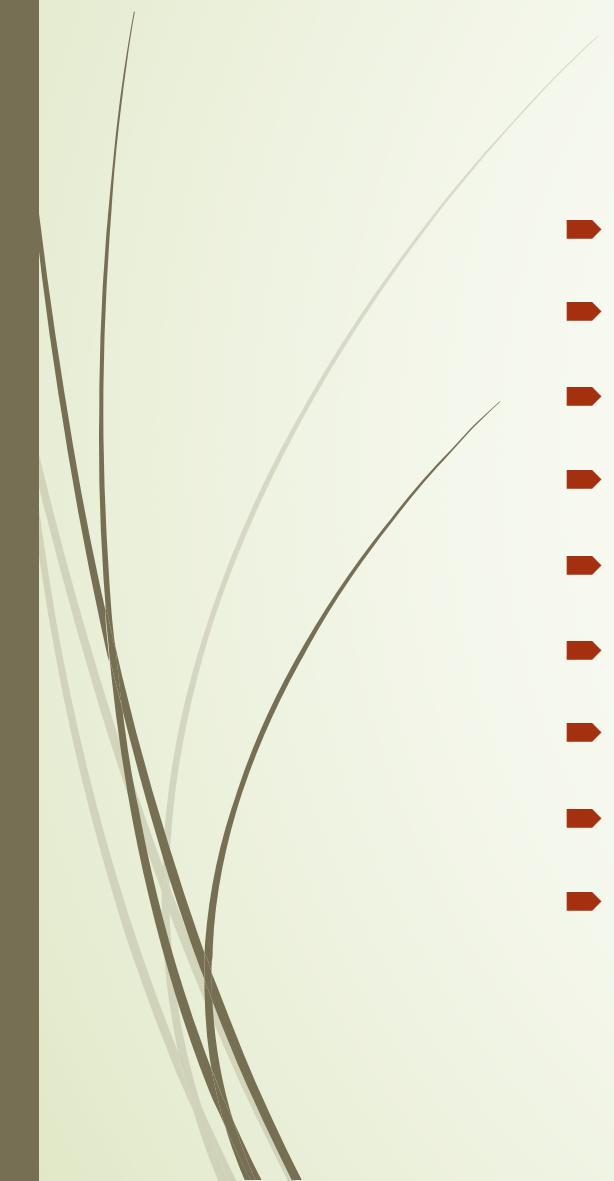
# Clustering of Diabetes

A short presentation on finding subgroups among adult-onset diabetes [\[1\]](#)

A Step Towards precision (personalised) medicine

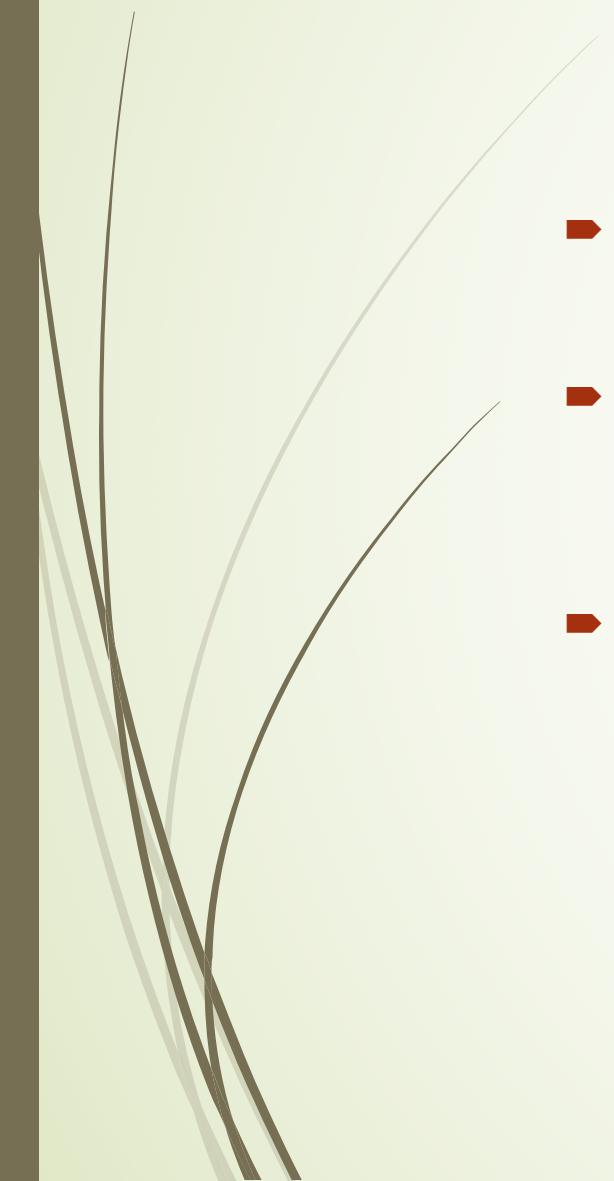
By

Prem Prakash



# Outline

- ▶ Objective
- ▶ Traditional Classification Method
- ▶ Drawback of Traditional Approach
- ▶ Proposed Methodology
- ▶ Data Cohort
- ▶ Results
- ▶ Limitation of the new Approach
- ▶ Conclusion
- ▶ Future Works



# Objective

- ▶ Stratification of diabetes types beyond **Type-1** and **Type-2**.
- ▶ **Identify** individual with **increased risk** for different types complications at diagnosis.
- ▶ A **refined** classification could pave the way for **personalised** medication thus provide optimal treatment.

# Traditional Classification Method

- ▶ **Diabetes**- excess of sugar in blood.
- ▶ Glycated Haemoglobin (**HbA<sub>1c</sub>**) test is used to get the average amount of glucose attached to haemoglobin over the period of past three months.
- ▶ HbA<sub>1c</sub> between **0% - 5%** is normal, a slight increase to **5% - 6.5%** is a cause of concern, and above **6.5%** is considered diabetic.
- ▶ Traditionally diabetic patients are classified into **two category** Type-1 and Type-2
- ▶ **Type-1** is when the pancreas can not produce **insulin**- a peptide hormone that helps carry sugar from the blood to cells for energy production.
- ▶ **Type-2** is when vital organs (liver, heart, kidney) becomes resistant to insulin or not enough insulin to carry out the task. 75–85%. It is **highly heterogeneous** therefore a reason for further stratification.
- ▶ A **third group** latent autoimmune diabetes (LADA) in adults affecting less than 10% of people with diabetes defined by the presence of glutamic acid decarboxylase antibodies (GADA)

# Drawback of Traditional Approach

- ▶ The complications developed are different in diabetic patients even in the same type (Type-2).
- ▶ Complication ranges over liver, kidney, cardiovascular organs.
- ▶ **Question:** Can we find a pattern at the onset of diabetes which can inform about the complication a patient is later going to develop?
- ▶ Broad classification impedes the precision in treatment.
- ▶ **Solution:** Find a further level of stratification among diseased that can help in providing more accurate treatment.

# Proposed Methodology

- ▶ The paper [1] uses **K-means clustering** to find subgroups in adult-onset diabetic patient.
- ▶ **Six features** extracted from blood-sample (easily obtainable, a minimum number of tests needed) and person features, are as follows:
  - ▶ Body Mass Index (BMI)
  - ▶ Age at onset of diabetes (diagnosis)
  - ▶ Glycated haemoglobin (HbA1c)
  - ▶ Homoeostasis model assessment (HOMA2) estimates of
    - ▶  $\beta$ -cell function (HOMA2-B)
    - ▶ insulin resistance (HOMA2-IR) based on C-peptide concentration
  - ▶ Glutamic Acid Decarboxylase Antibodies (**GADA**) - binary (0/1)- presence or absence

# Proposed Methodology...

- ▶ Data is **normalized** with mean zero and standard deviation one. Because K-means gives equal importance to all the features as it uses Euclidean distance.
- ▶ **GADA** positive and negative cohort were trained **separately** since K-means does not accommodate binary variables.
- ▶ **Men and women** are clustered **separately** to avoid sex-dependent differences.
- ▶ To get an optimum number of cluster (K) **silhouette** algorithm is used. There are other techniques as well like **elbow, gap-statistics** (GS). GS and silhouette are the most stable ones [\[2\]](#).
- ▶ **Five** optimum clusters were found for both men and woman, GADA positive and negative as well.

# Data Cohort

- ▶ Five cohorts from **Sweden** and **Finland**.
  - ▶ All New Diabetics in Scania (ANDIS),
  - ▶ The Scania Diabetes Registry (SDR),
  - ▶ All New Diabetics in Uppsala (ANDIU),
  - ▶ Diabetes Registry Vaasa (DIREVA),
  - ▶ Malmö Diet and Cancer CardioVascular Arm (MDC-CVA).

# Results from Analysis of the new Subgroups

- ▶ **Five** optimum clusters were identified.
- ▶ On all of the subsets of data the clusters were similar.
- ▶ Groups with its characteristics:
  - ▶ Severe Autoimmune Diabetes (**SAID**): characteristics- early-onset disease, low BMI, poor metabolic control, insulin **deficiency**, and **presence** of **GADA**
  - ▶ Severe Insulin-deficient Diabetes (**SIDD**): similar to **SIDD** but GADA **negative**, low insulin secretion i.e. low HOMA2-B index, low metabolic control.
  - ▶ Severe Insulin-resistant Diabetes (**SIRD**): insulin resistance (high HOMA2-IR index) and high BMI.
  - ▶ Mild Obesity-related Diabetes (**MOD**): Characterised by obesity rather by insulin resistance.
  - ▶ Mild Age-related Diabetes (**MARD**): older than patients in other clusters and similar to MOD cluster, and showed modest metabolic resistance.
- ▶ **Note:** New set of incoming patients are clustered independently among themselves where K=5 is used.

# Inferences

- ▶ On further analysis of each group on mean HbA<sub>1c</sub>, antidiabetic therapy, progression of disease, genetic sequencing, etc., provided following inference:
  - ▶ **SAID** overlapped with **Type-1** diabetes and LADA (GADA)
  - ▶ The **SIRD** had a higher risk of **kidney** complications compared to other insulin resistance group MOD and MARD
  - ▶ **Genetic** association in the clusters **differed** from those seen in traditional Type-2 diabetes.
  - ▶ **Time** for sustained use of insulin varied.

# Limitations of The New Approach

- ▶ The new groups do not say anything about **aetiology** (cause/origin) of disease.
- ▶ Soft and hard rule for **moving patients between clusters**. A study on a large group is needed to understand better.
- ▶ **Limited diversity of the dataset**- patients from northern Europe and limited non-Scandinavian representation.
- ▶ Its applicability to **other ethnicities** needs to be assessed.



# Conclusion

- ▶ Information extracted from **variable central to the development** of diabetes is far richer than merely assessing with only one metabolite (HbA1c), glucose.
- ▶ The **sub-grouping** is a step towards an early prediction of chronic diabetic complications.
- ▶ In medical treatment **personalised** care is better than generalized care, and this an effective step toward precision medicine.



# Future Works

- ▶ Only **two types** of autoantibodies were used, effects of other antibodies are unknown.
- ▶ **Availability of data** on some **risk** factor such as blood pressure, blood lipids etc., can further improve the clustering.
- ▶ Improve the stratification through inclusion of **additional features** such as **biomarkers**, genotypes, genetic risk scores etc.
- ▶ Long way to go before developing precision medicine.

# References

- ▶ Ahlqvist, Emma, et al. "Novel subgroups of adult-onset diabetes and their association with outcomes: a data-driven cluster analysis of six variables." *The Lancet Diabetes & Endocrinology* 6.5 (2018): 361-369 [\[1\]](#).
- ▶ [Determining The Optimal Number Of Clusters: 3 Must Know Methods](#) [\[2\]](#)

THANK YOU