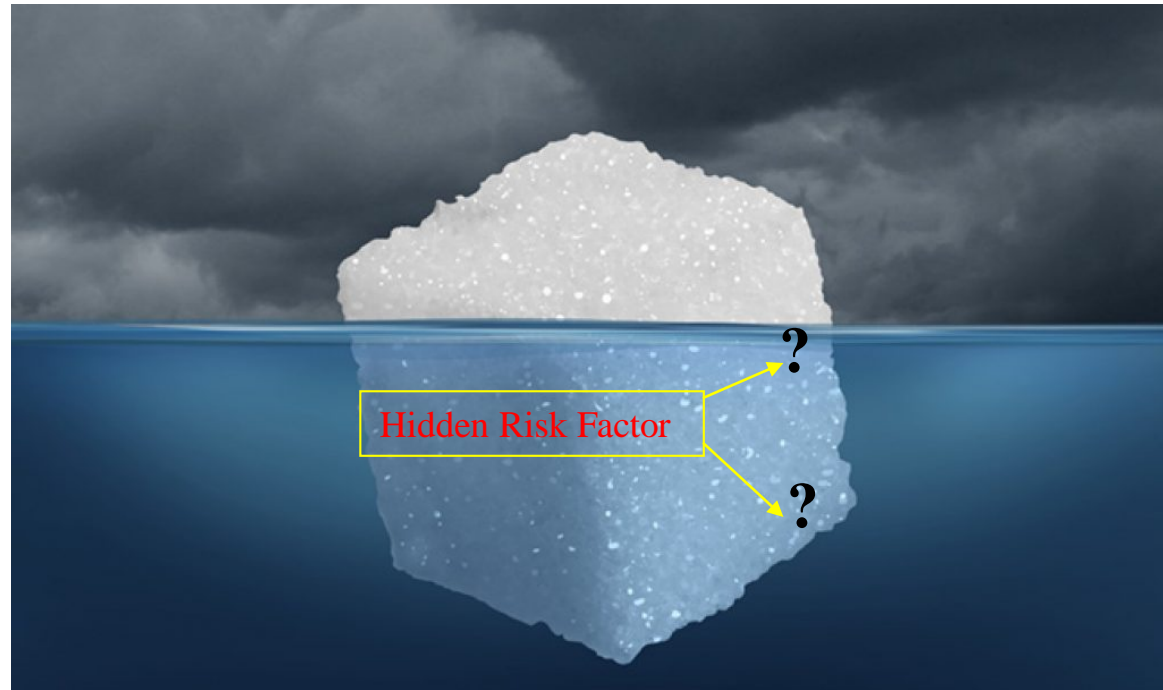


# Opening the Black Box: Discovering and Explaining Hidden Variables in Patient Modelling



Leila Yousefi, Stephen Swift, Mahir Arzoky, Allan Tucker

Brunel University London

Lucia Saachi , Luca Chiovato

University of Pavia, Istituti Maugeri, Italy

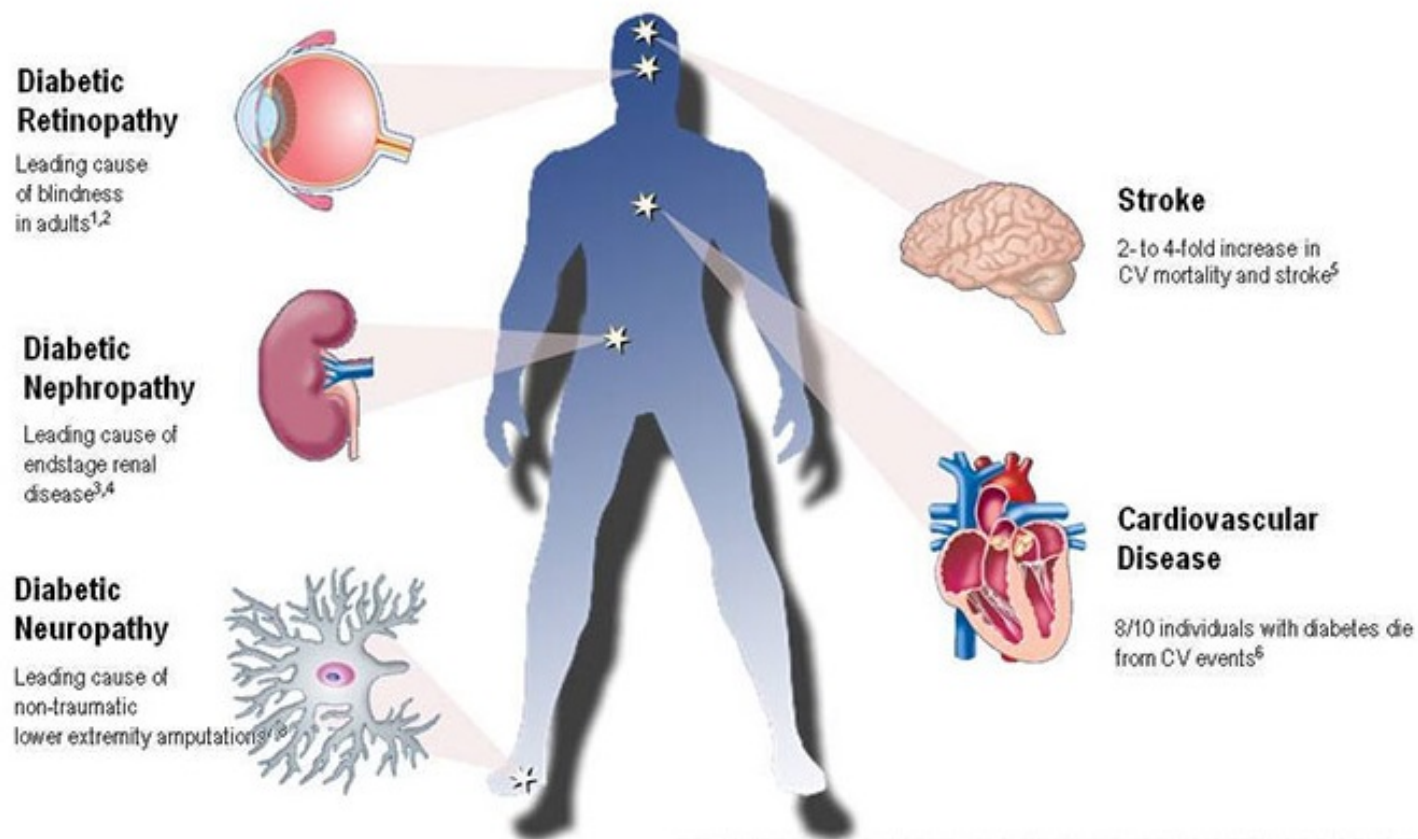
# The Data at Maugeri, Pa

- Type 2 Diabetes Mellitus (T2DM)
- Patients aged 25 to 65 years.
- 2009 and 2013.
- IRCCS Istituti Clinici Scientifici Maugeri of Pavia, Italy.
- MOSAIC project funded by the European Commission.
  - T2DM risk factors:
  - Physical examination
  - Laboratory data

## Problem

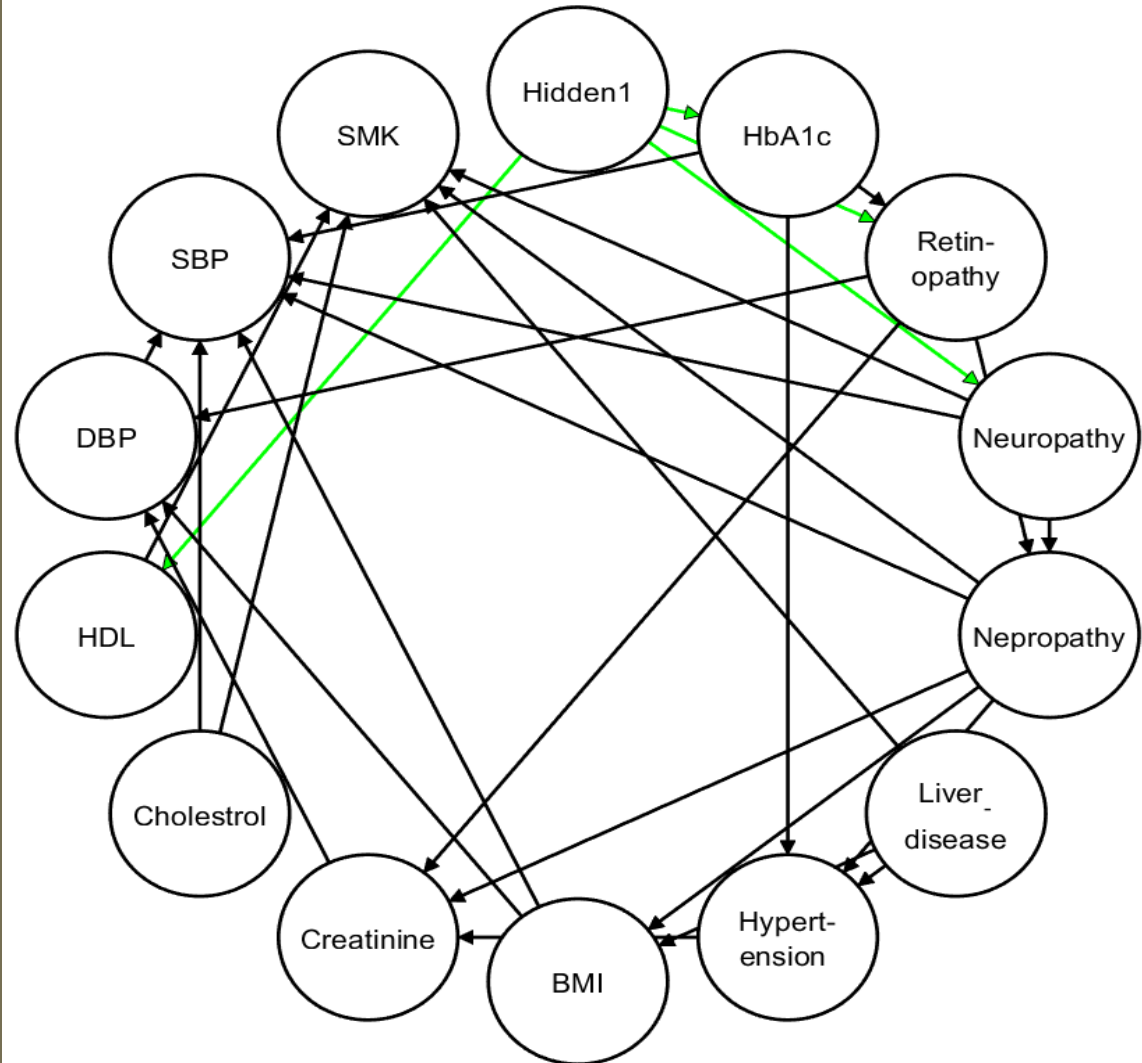
- Predicting comorbidities at the earliest from time-series data is challenging.
- Each patient has a dynamic and unique profile.
- Comorbidities interact.
- Many unmeasured effects.

## Diabetes is a lifelong condition associated with serious complications

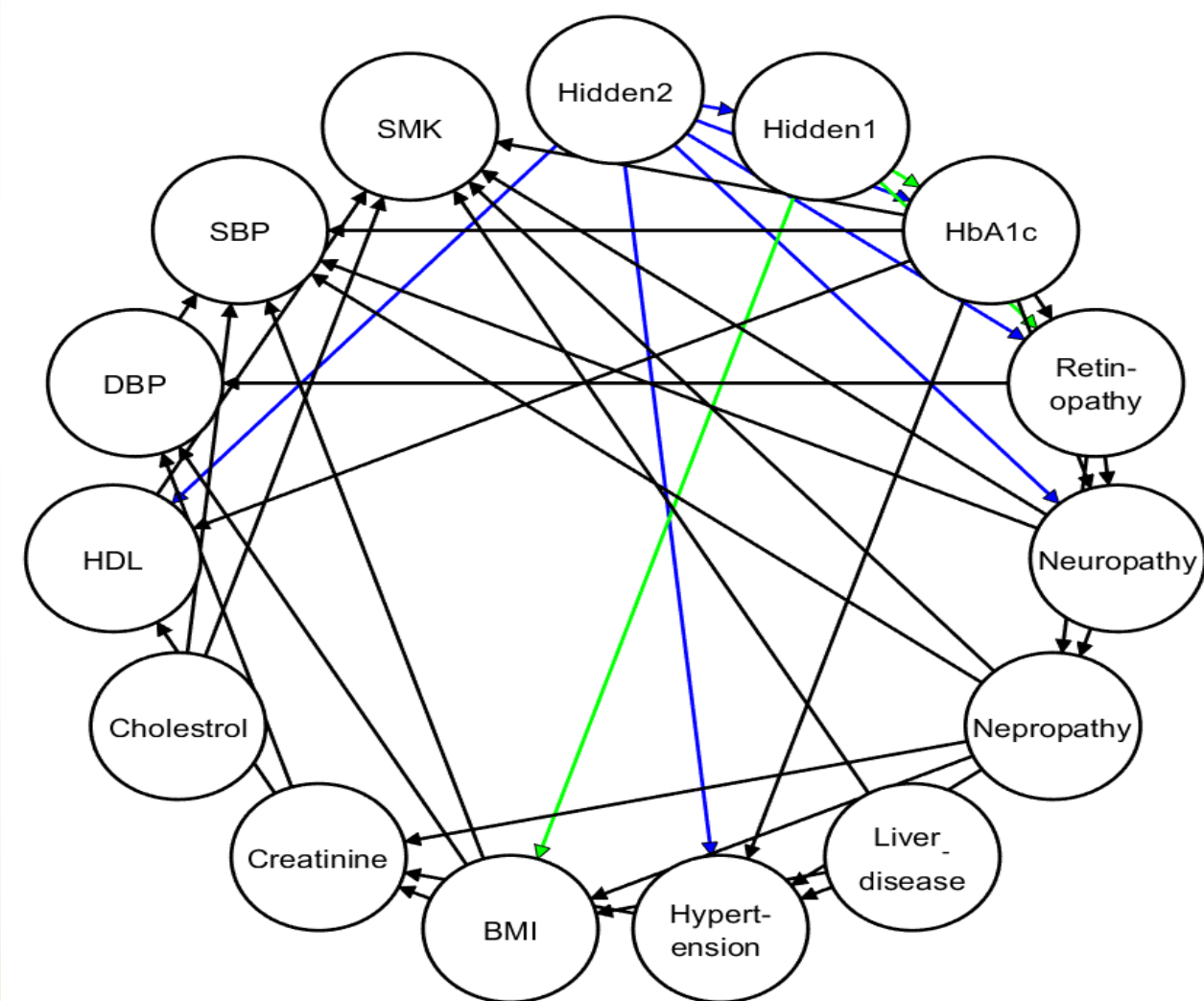


# Solution

Step 1:



Step 2:



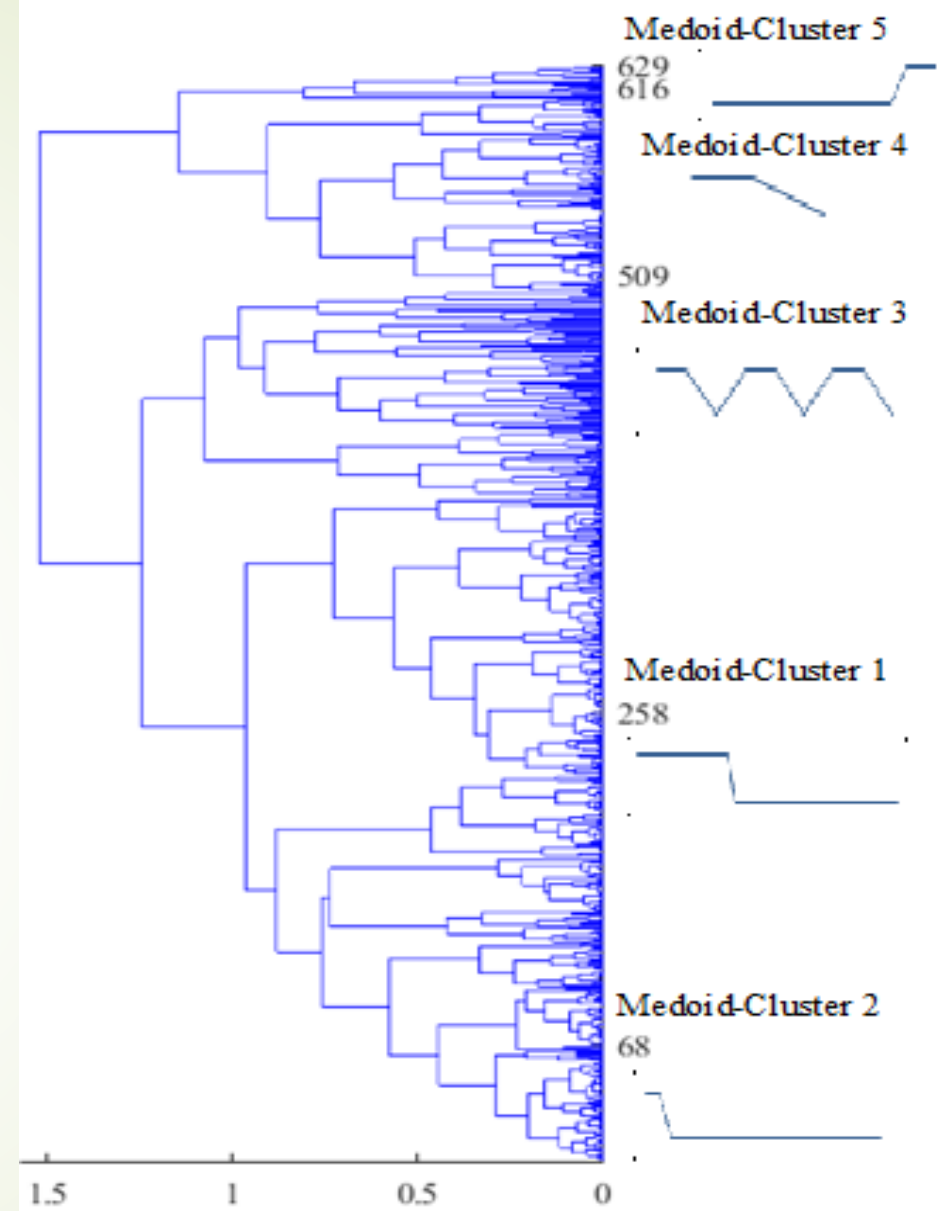
# Hierarchical Clustering and discovering Phenotypes

C1	Temporal Phenotype For Hidden variable pattern 1	C2	Temporal Phenotype For Hidden variable pattern 2
Cluster1		Cluster1	
Cluster2		Cluster2	
Cluster3		Cluster3	
Cluster4		Cluster4	
Cluster5		Cluster5	

C3	Temporal Phenotype For Hidden variable pattern 3	C4	Temporal Phenotype For Hidden variable pattern 4
Cluster1		Cluster1	
Cluster2		Cluster2	
Cluster3		Cluster3	
Cluster4		Cluster4	
Cluster5		Cluster5	

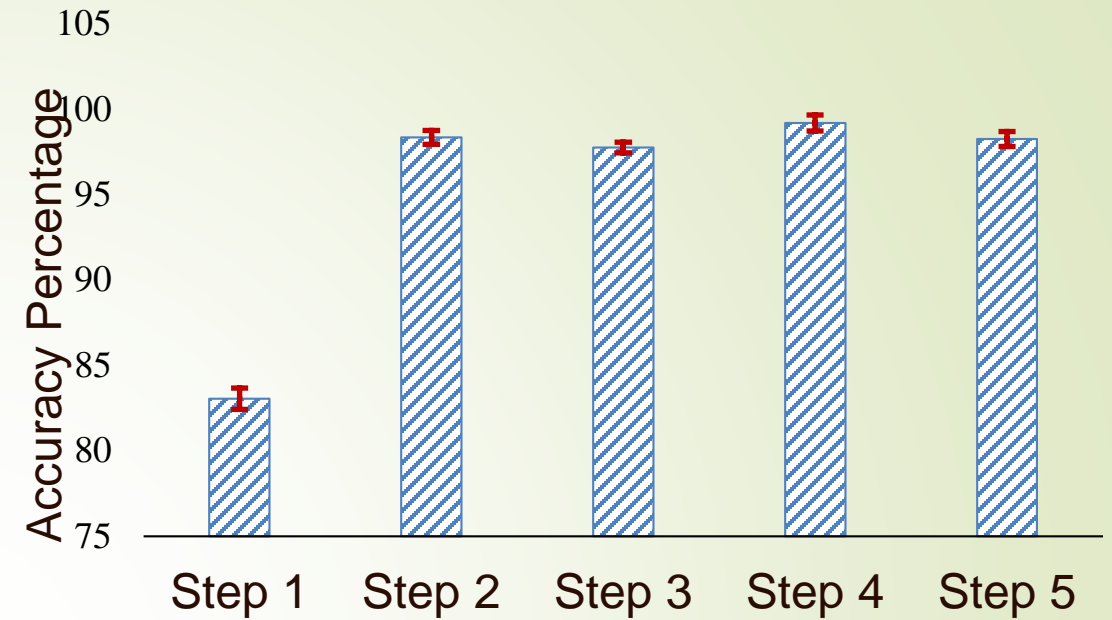
Using the Medoid hidden variable cluster profile for "deep temporal phenotype".



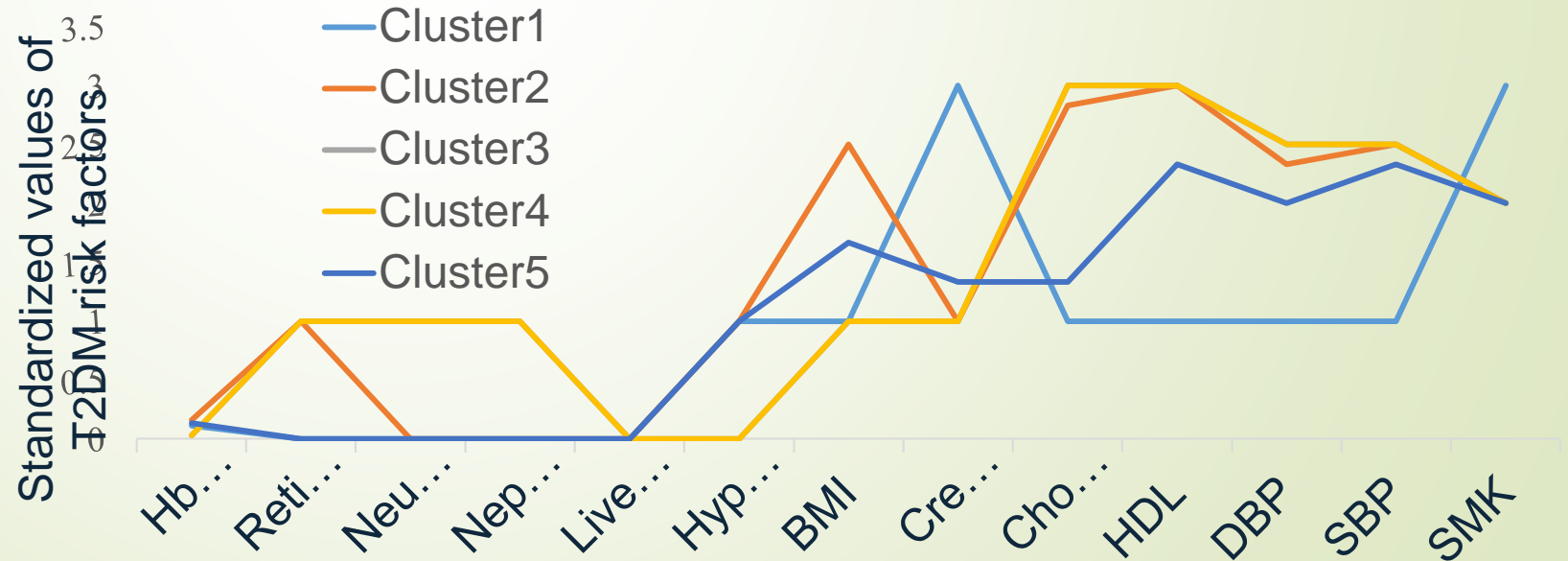
Dendrogram of Hierarchical clustering (complete) and Temporal phenotypes (The First Hidden Clusters "Profiles")

# Results

Accuracy and Errors Bar for Five Steps of Adding Hidden variables in predicting retinopathy.



Mean values of T2DM risk factors and complications clusters based on the Fourth Hidden variable (C4).





Thank you!

Happy to discuss more,  
Please come and see  
me at the poster.