# Predicting Disease Complications

## Using a Stepwise Hidden Variable Approach For Learning Dynamic Bayesian Networks


**Leila Yousefi**

**Allan Tucker**
**Mashael Al-Luhaybi**
Brunel University London

**Lucia Saachi,**
**Riccardo Bellazzi**
University of Pavia

**Luca Chiovato**
Instituti Maugeri


Type 2 Diabetes
what people see
high blood sugar
what people don't see
blindness
blurred vision
boils
cataracts
depression
erectile dysfunction
foot ulcers
frequent urination
glaucoma
intense fatigue
intense hunger
intense thirst
itchiness
kidney disease
numbness
pain
sexual dysfunction
skin infections

# Outline

- Motivation
- Data
- Problem
- Solution
- Dynamic Bayesian Networks (DBNs)
- Hidden variable discovery approach
  - Pair sampling and Stepwise approach procedure
- Results
- Conclusions and future works

# Type 2 Diabetes Mellitus (T2DM)


Mortality due to diabetes age 20-79 in 2017 (in millions)

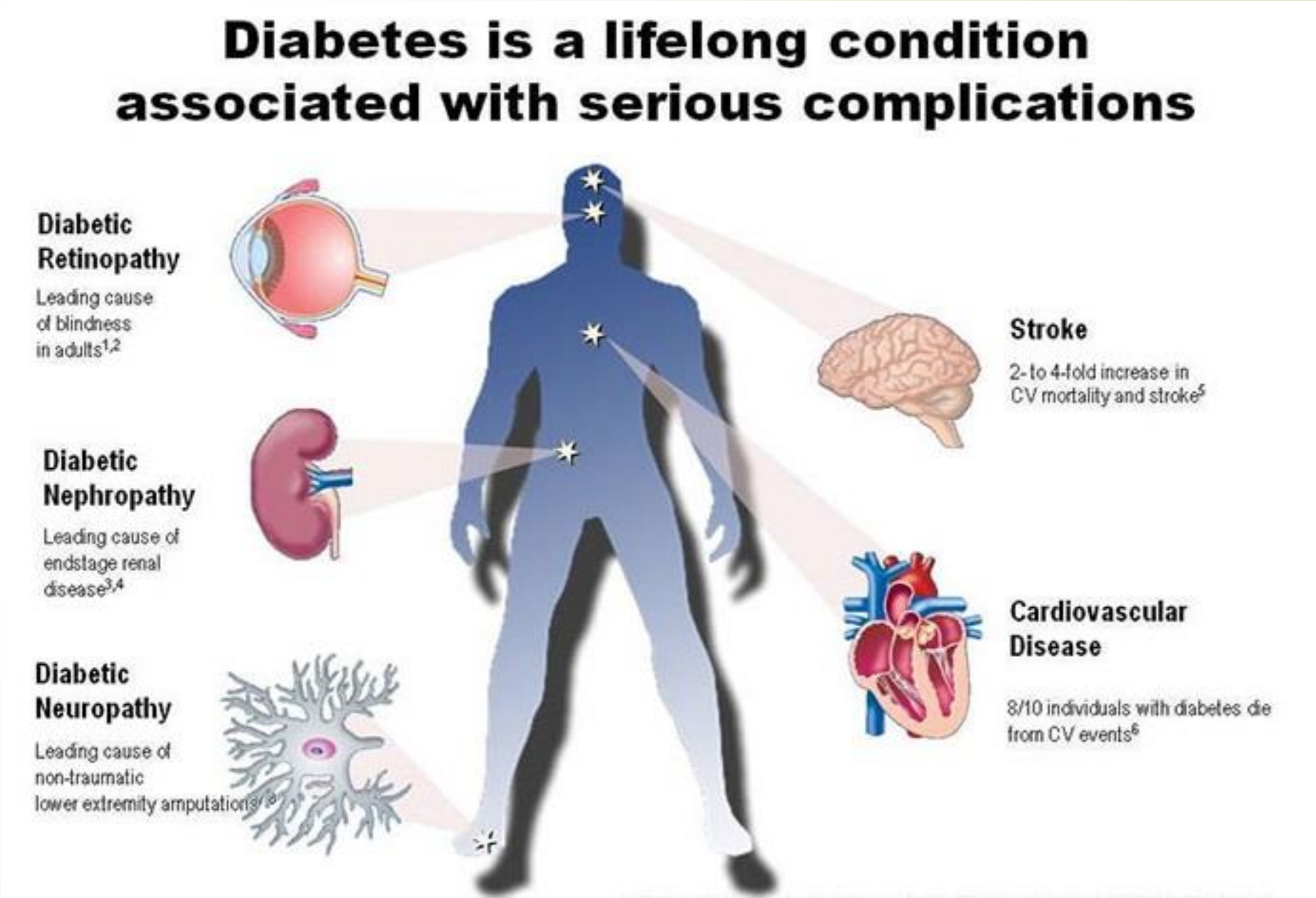
International Diabetes Federation

# The Data at Maugeri, Pavia:

- 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

- MATLAB and Bayes Net toolbox (murphy,2001)
- Visualization we used Graphviz.

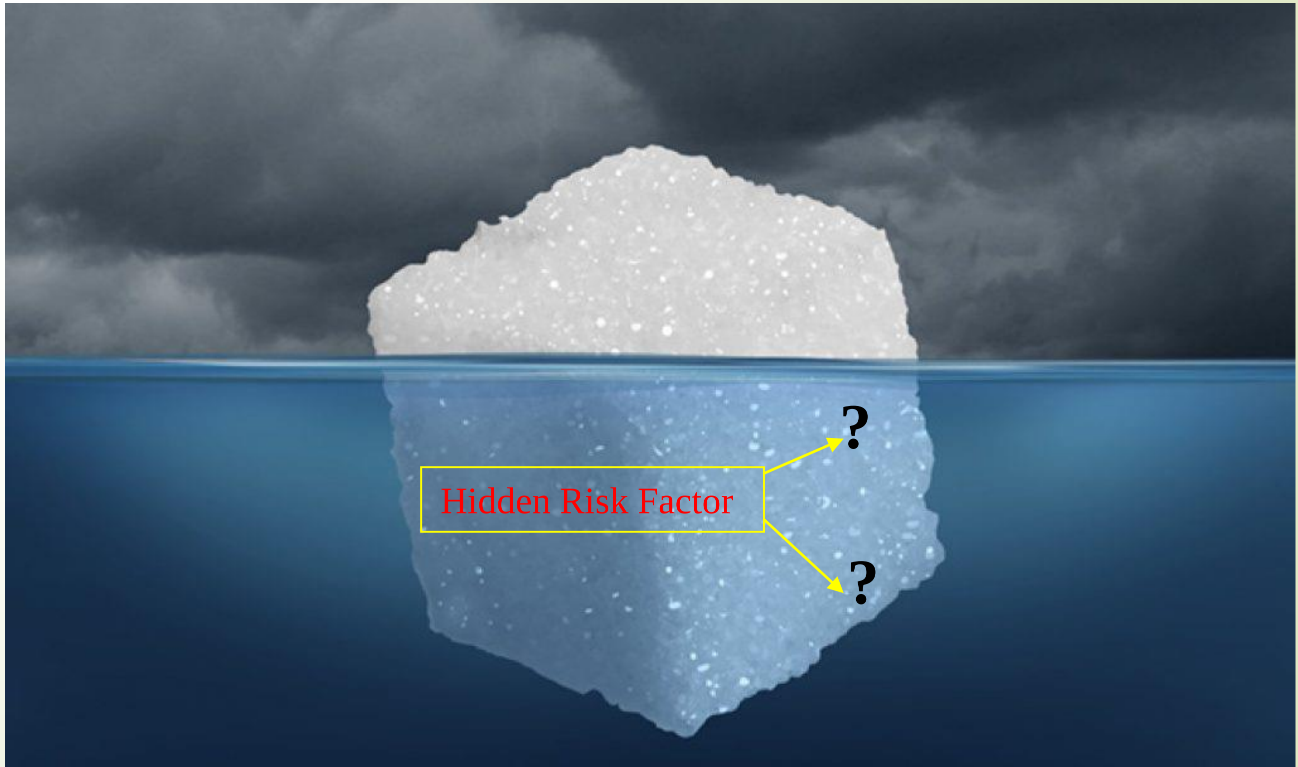


# T2DM Data

| Visit NO | Patient ID | HbA1c | Retinopathy | Neuropathy | Nephropathy | Liver disease | Hypertension | BMI | Creatinine | Cholestrol | HDL | DBP | SBP | SMK |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 1 | 885 | 0.769 | 0 | 0 | 0 | 0 | 1 | 0.286 | -0.391 | 2.082 | 0.020 | 1.705 | 0.286 | 1.335 |
| 2 | 885 | 0.769 | 0 | 0 | 0 | 1 | 1 | 0.286 | -0.391 | 2.082 | 0.020 | 1.705 | 0.286 | 1.335 |
| 3 | 885 | 0.769 | 1 | 0 | 0 | 1 | 1 | 0.286 | -0.391 | 2.082 | 0.020 | 1.705 | 0.286 | 1.335 |
| 4 | 885 | 0.769 | 1 | 0 | 1 | 1 | 1 | 0.286 | -0.391 | 2.082 | 0.020 | 1.705 | 0.286 | 1.335 |
| 1 | 894 | 0.151 | 0 | 0 | 1 | 1 | 1 | 2.782 | -0.511 | -0.149 | -0.053 | 0.297 | 0.286 | 1.335 |
| 2 | 894 | 0.151 | 0 | 0 | 1 | 1 | 1 | 2.782 | -0.511 | -0.149 | -0.053 | 0.297 | 0.286 | 1.335 |
| 3 | 894 | 0.151 | 0 | 0 | 1 | 1 | 1 | 2.782 | -0.511 | -0.149 | -0.053 | 0.297 | 0.286 | 1.335 |
| 4 | 894 | -0.056 | 0 | 0 | 1 | 1 | 1 | 2.937 | -0.511 | -0.017 | -0.343 | 0.297 | 0.794 | 1.335 |
| 5 | 894 | -0.056 | 0 | 0 | 1 | 1 | 1 | 2.937 | -0.511 | -0.017 | -0.343 | 0.297 | 0.794 | 1.335 |
| 6 | 894 | -0.056 | 0 | 0 | 1 | 1 | 1 | 2.937 | -0.511 | -0.017 | -0.343 | 0.297 | 0.794 | 1.335 |
| 7 | 894 | -0.262 | 0 | 0 | 1 | 1 | 1 | 2.782 | -0.511 | 0.534 | -0.488 | 0.297 | 0.540 | 1.335 |
| 8 | 894 | -0.262 | 0 | 0 | 1 | 1 | 1 | 2.782 | -0.511 | 0.534 | -0.488 | 0.297 | 0.540 | 1.335 |
| 9 | 894 | -0.262 | 0 | 0 | 1 | 1 | 1 | 2.782 | -0.511 | 0.534 | -0.488 | 0.297 | 0.540 | 1.335 |
| 10 | 894 | 0.151 | 0 | 0 | 1 | 1 | 1 | 2.906 | -0.511 | 0.744 | -0.488 | -0.642 | -0.223 | 1.335 |
| 11 | 894 | 0.151 | 0 | 0 | 1 | 1 | 1 | 2.906 | -0.511 | 0.744 | -0.488 | -0.642 | -0.223 | 1.335 |
| 12 | 894 | 0.151 | 0 | 0 | 1 | 1 | 1 | 2.906 | -0.511 | 0.744 | -0.488 | -0.642 | -0.223 | 1.335 |
| 13 | 894 | 0.151 | 0 | 0 | 1 | 1 | 1 | 3.557 | -0.391 | 0.376 | 0.455 | -0.642 | -0.223 | 1.335 |
| 14 | 894 | 0.151 | 0 | 0 | 1 | 1 | 1 | 3.557 | -0.391 | 0.376 | 0.455 | -0.642 | -0.223 | 1.335 |
| 15 | 894 | 0.151 | 0 | 0 | 1 | 1 | 1 | 3.557 | -0.391 | 0.376 | 0.455 | -0.642 | -0.223 | 1.335 |
| 16 | 894 | 0.013 | 0 | 0 | 1 | 1 | 1 | 3.324 | -0.235 | 0.744 | -0.125 | -0.642 | -0.223 | 1.335 |
| 1 | 1010 | 1.388 | 0 | 0 | 1 | 0 | 0 | 0.162 | -0.630 | 2.450 | -0.779 | 2.175 | 2.827 | 1.335 |
| 2 | 1010 | 1.388 | 0 | 0 | 1 | 0 | 1 | 0.162 | -0.630 | 2.450 | -0.779 | 2.175 | 2.827 | 1.335 |
| 3 | 1010 | 1.388 | 0 | 0 | 1 | 0 | 1 | 0.162 | -0.630 | 2.450 | -0.779 | 2.175 | 2.827 | 1.335 |
| 4 | 1010 | 1.388 | 0 | 0 | 1 | 0 | 1 | 0.162 | -0.630 | 2.450 | -0.779 | 2.175 | 2.827 | 1.335 |
| 5 | 1010 | 2.350 | 0 | 0 | 1 | 0 | 1 | 0.206 | -0.511 | 0.875 | 0.818 | 0.297 | -0.223 | 1.335 |
| 6 | 1010 | 2.350 | 0 | 0 | 1 | 0 | 1 | 0.206 | -0.511 | 0.875 | 0.818 | 0.297 | -0.223 | 1.335 |
| 7 | 1010 | 2.350 | 0 | 0 | 1 | 0 | 1 | 0.206 | -0.511 | 0.875 | 0.818 | 0.297 | -0.223 | 1.335 |
| 8 | 1010 | 2.350 | 0 | 0 | 1 | 0 | 1 | 0.206 | -0.511 | 0.875 | 0.818 | 0.297 | -0.223 | 1.335 |
| 9 | 1010 | 2.350 | 0 | 0 | 1 | 0 | 1 | 0.078 | -0.750 | 1.636 | -0.053 | -0.642 | 0.286 | 1.335 |

# The 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.




# The solution, Personalising Medicine

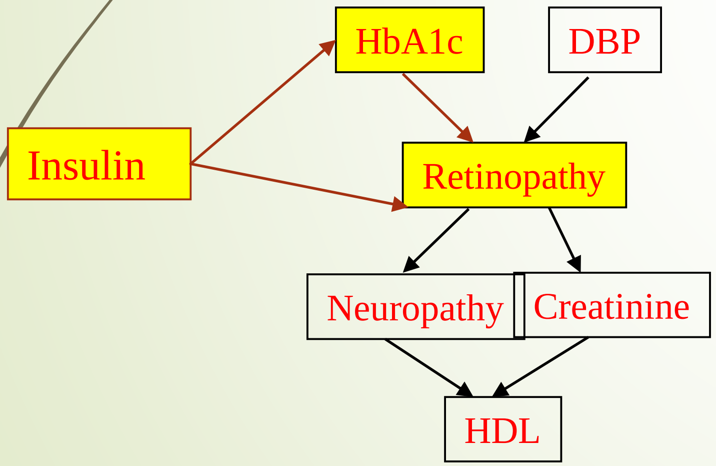
## Hidden Variable discovery approach

- Finding methods to assess the influences of these latent variables,
- Discover the dependencies between the latent variable and the observed variables.
- Discover Diabetic trigger and eliminate diabetes forever!
- Determining the precise position of the latent variable

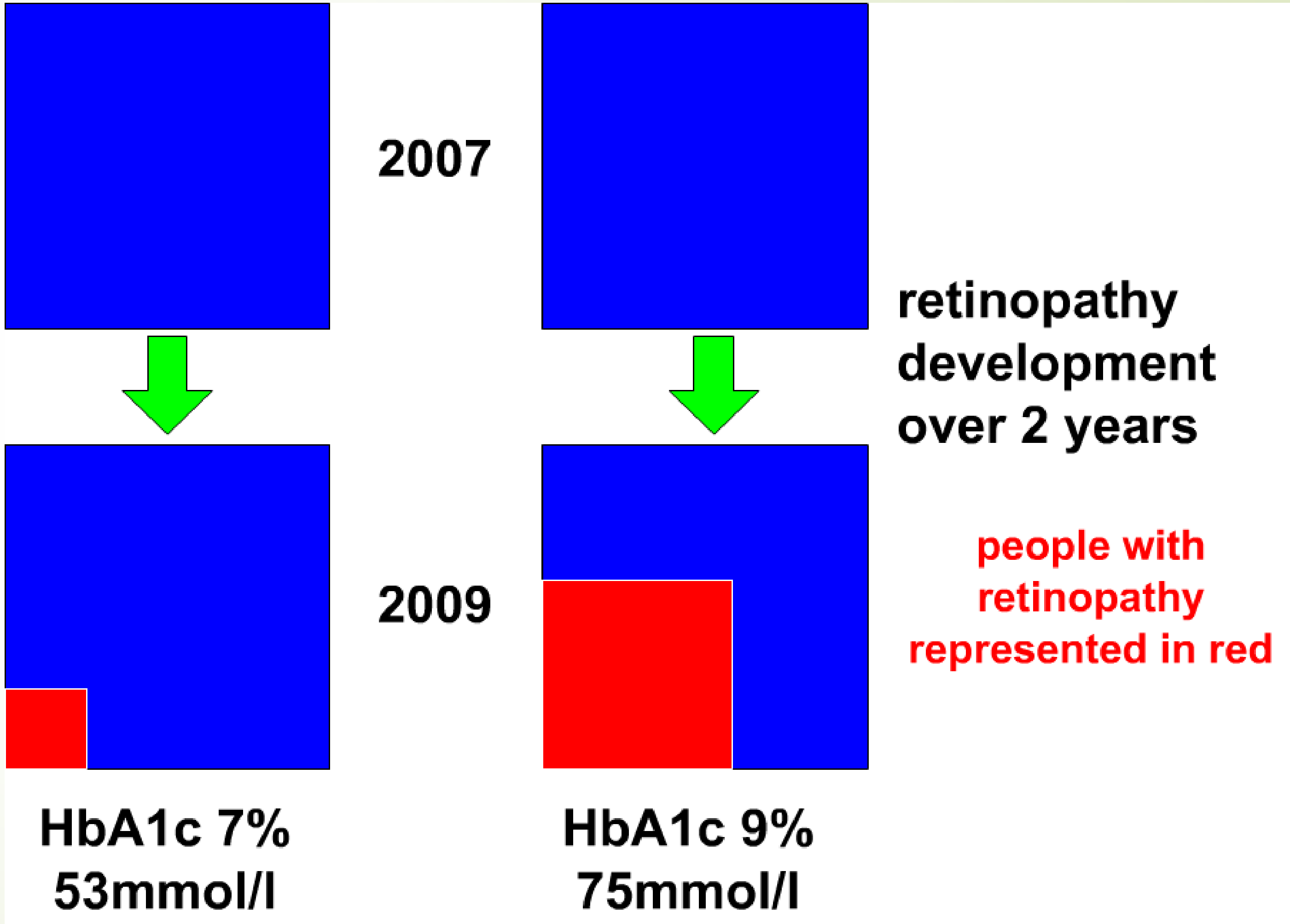
- Our key contribution is the combination of the IC* algorithm to identify latent variables within Dynamic Bayesian Networks.

# An example

**Our hypothesis is:**
"If Glycated Hemoglobin (HbA1c) is less than about (7%), then retinopathy may never develop, or develop very slowly."




Our observations….

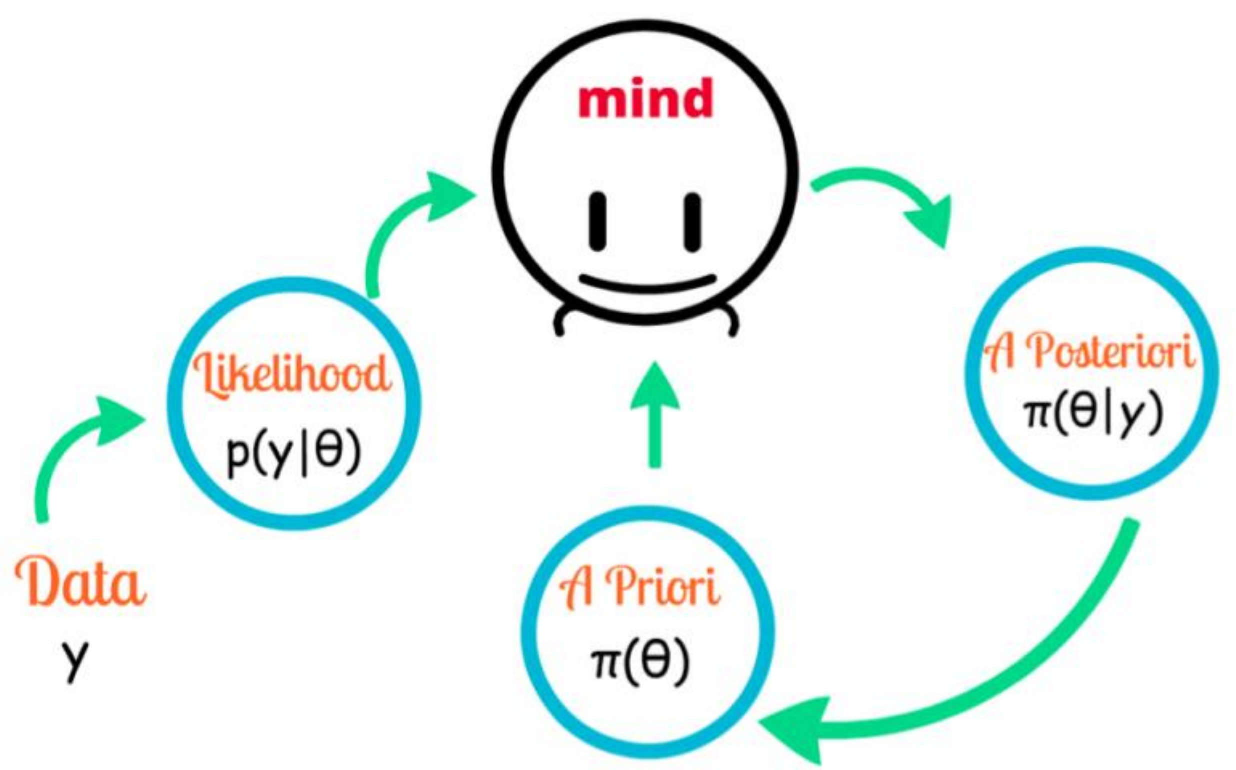



Leila.Yousefi@brunel.ac.uk

http://www.diabeticretinopathy.org.uk/prevention/hba1c_and_retinopathy.htm

# Dynamic Bayesian Networks

- Ideal for clinical data:
  - Flexibility in continuous and discrete variable;
  - Handling uncertainty through the modelling of probability distributions;
  - Enables prediction through inference;
  - No limit for minimum sample size;
  - Transparent (querying the model, graphical structure
  - It can naturally facilitate latent variables …

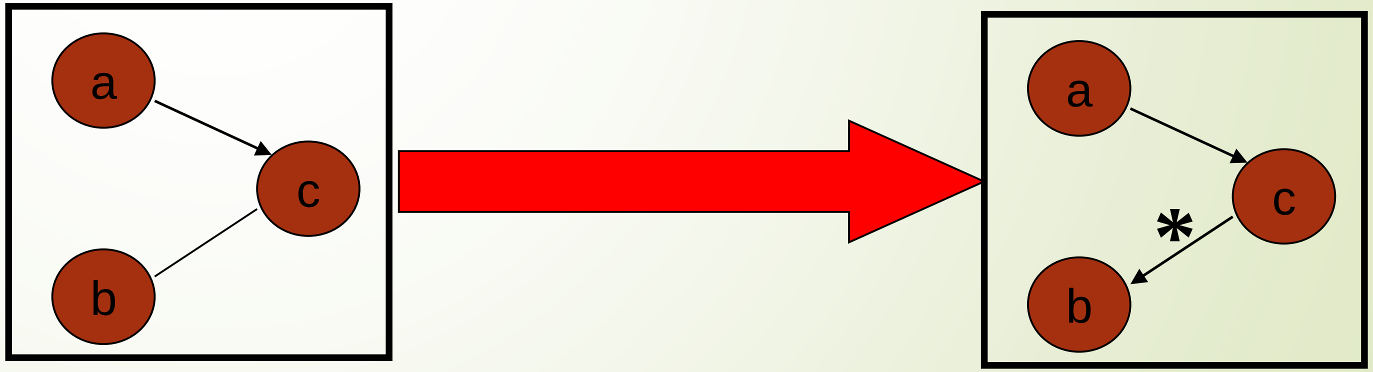



Leila.Yousefi@brunel.ac.uk

# IC (Inductive Causation) algorithm

- It Applies conditional independence analyses to infer causal structures;
- IC* algorithm (an extension of IC) learns a partially oriented Directed Acyclic graph (pattern) with latent variables.

**Whenever a then b but not vice versa**
**-------------**
**Possibly a => b**




Leila.Yousefi@brunel.ac.uk

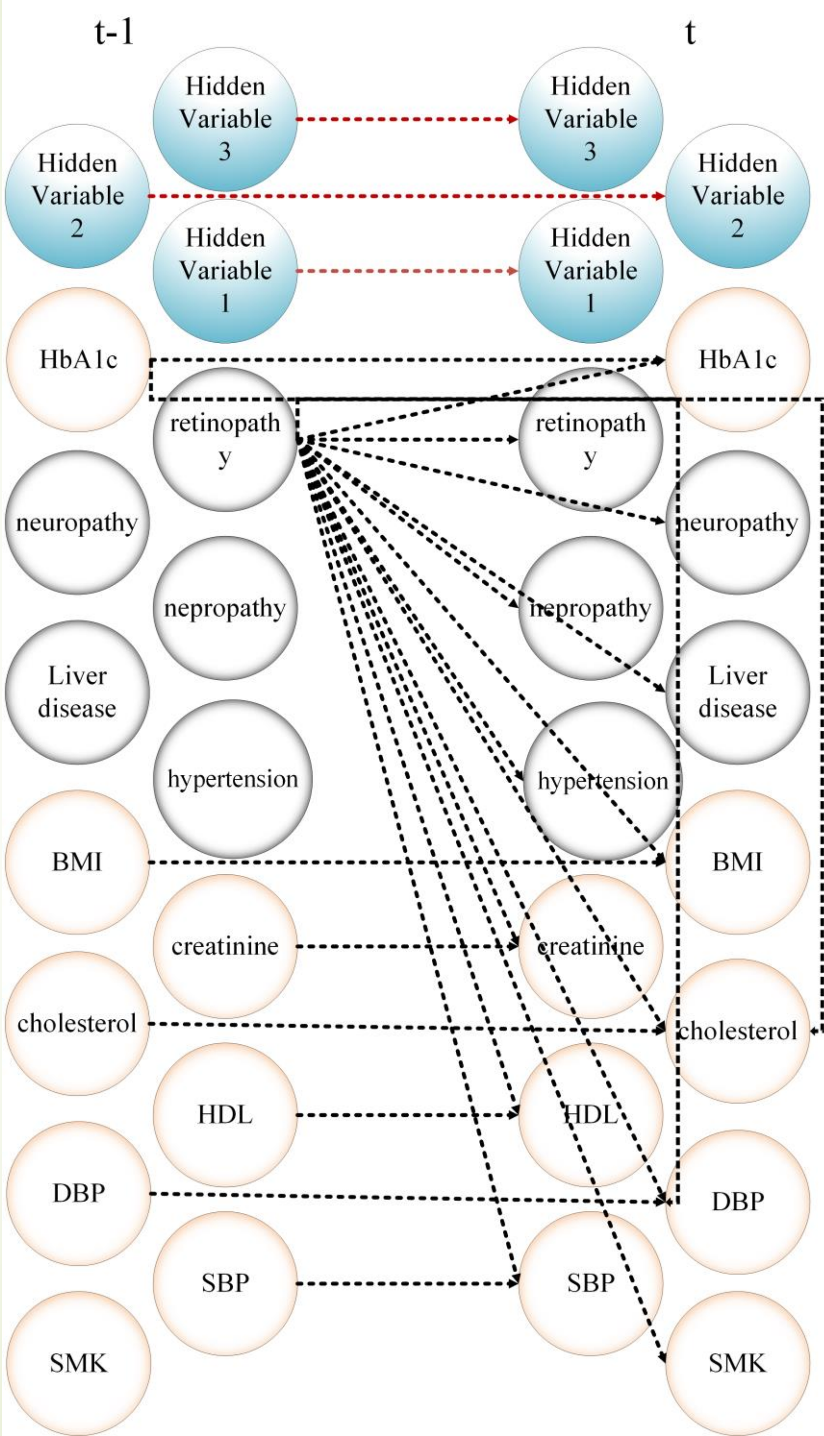
See Pearl, "Causality: Models, Reasoning, and Inference", 2000, p52 for more details.

# Dynamic links

- Learning the temporal links of our DBNs,

Using REVerse Engineering ALgorithm (REVEAL) (Liang1998)

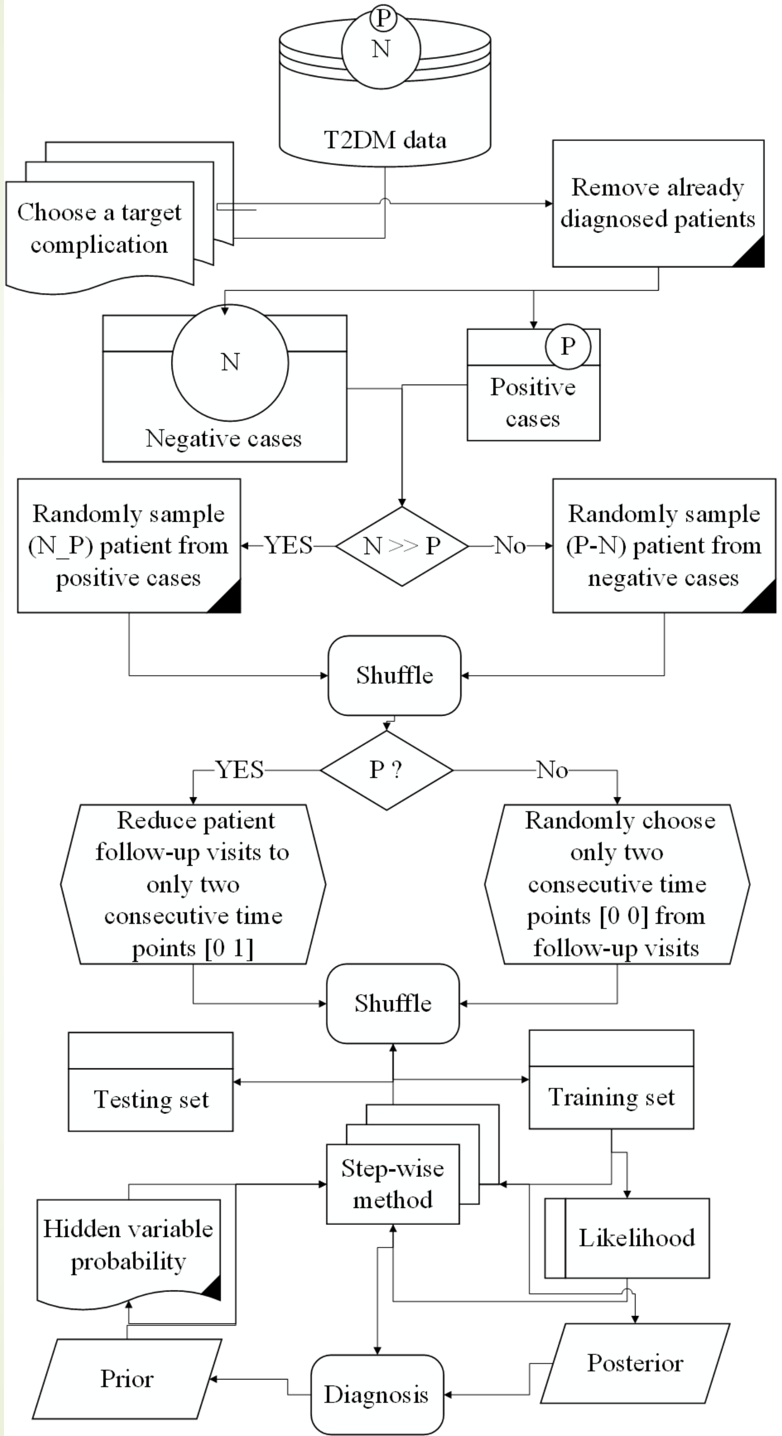
- We assumed hidden variable status at time *t* depends on the corresponding hidden variable at a previous time (*t-1*).




Leila.Yousefi@brunel.ac.uk

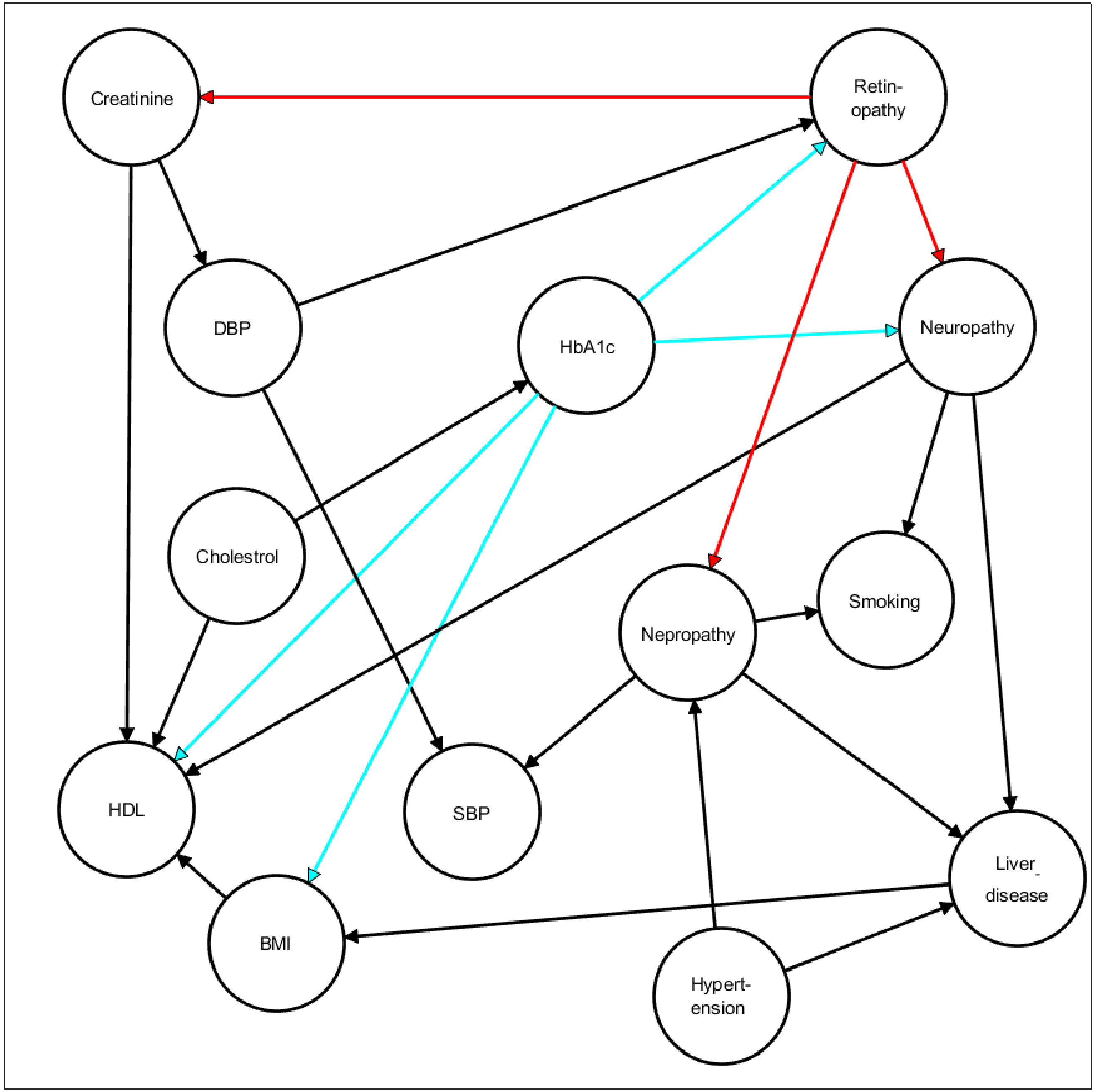
# Pair sampling and Stepwise approach procedure

- Apply IC* algorithm on the balanced data.
- Provide probabilities of states by applying inference rules on all discovered hidden variables.
- Treat the discovered hidden variables as an observed variables.
- Re-apply the IC* and repeat all Steps until no new hidden variables are discovered.
- Having discovered the hidden variables, we build a predictive DBN model.
- Parameter estimation using the expectation-maximization (EM) algorithm.



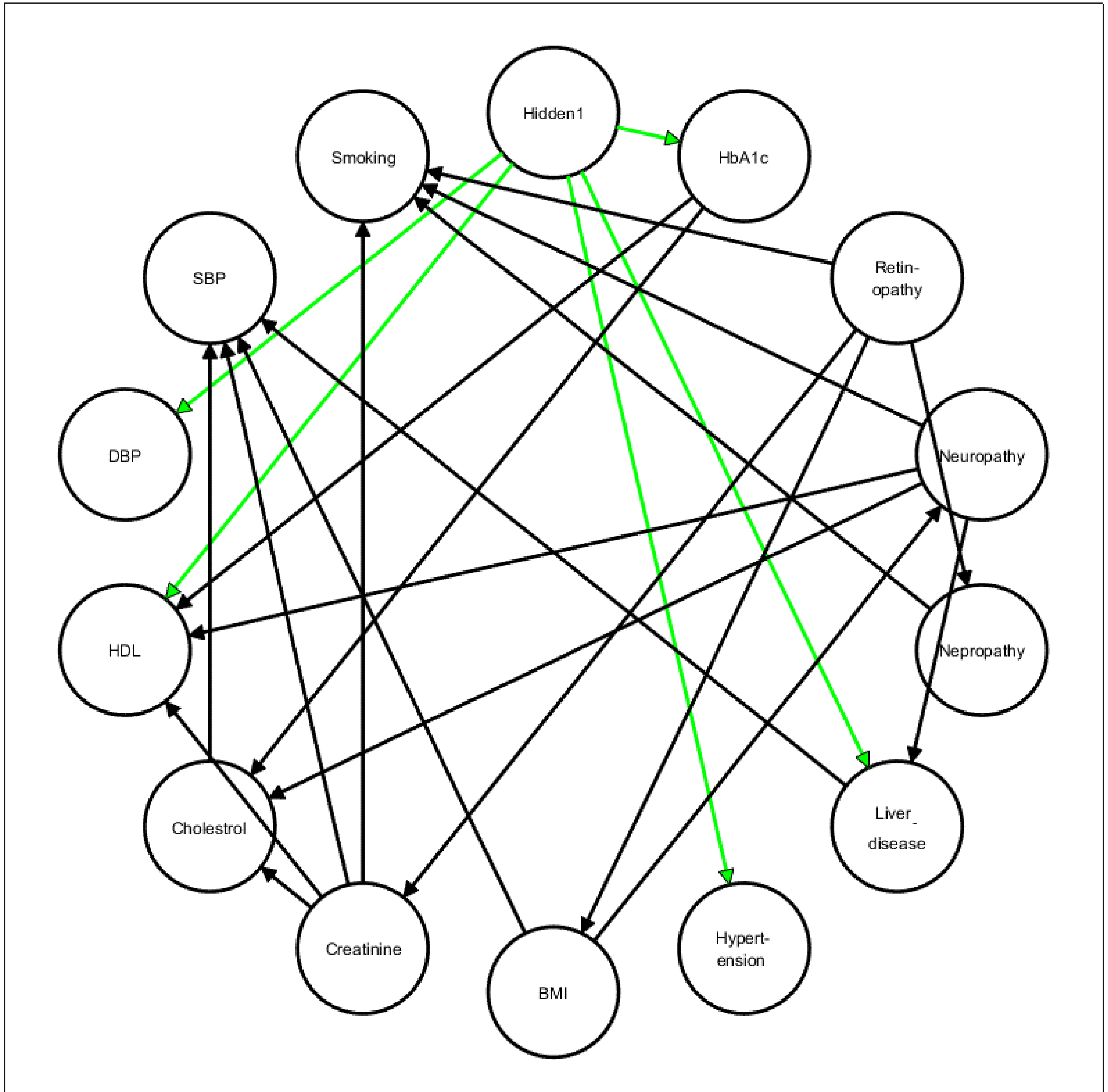

Leila.Yousefi@brunel.ac.uk

# Static structure with no hidden variable



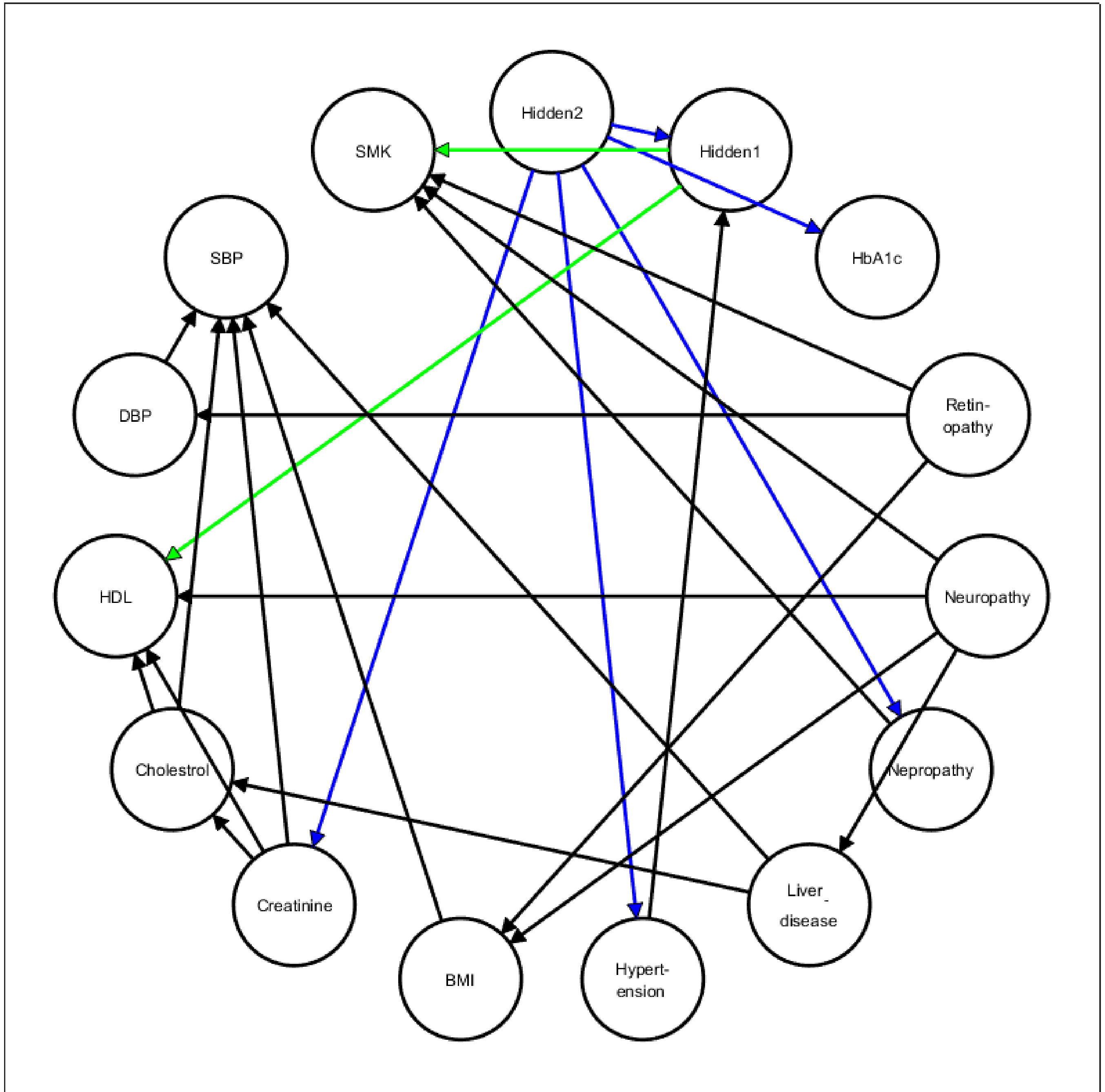
# Step 1

## Static structure with the first hidden variable



Step 2

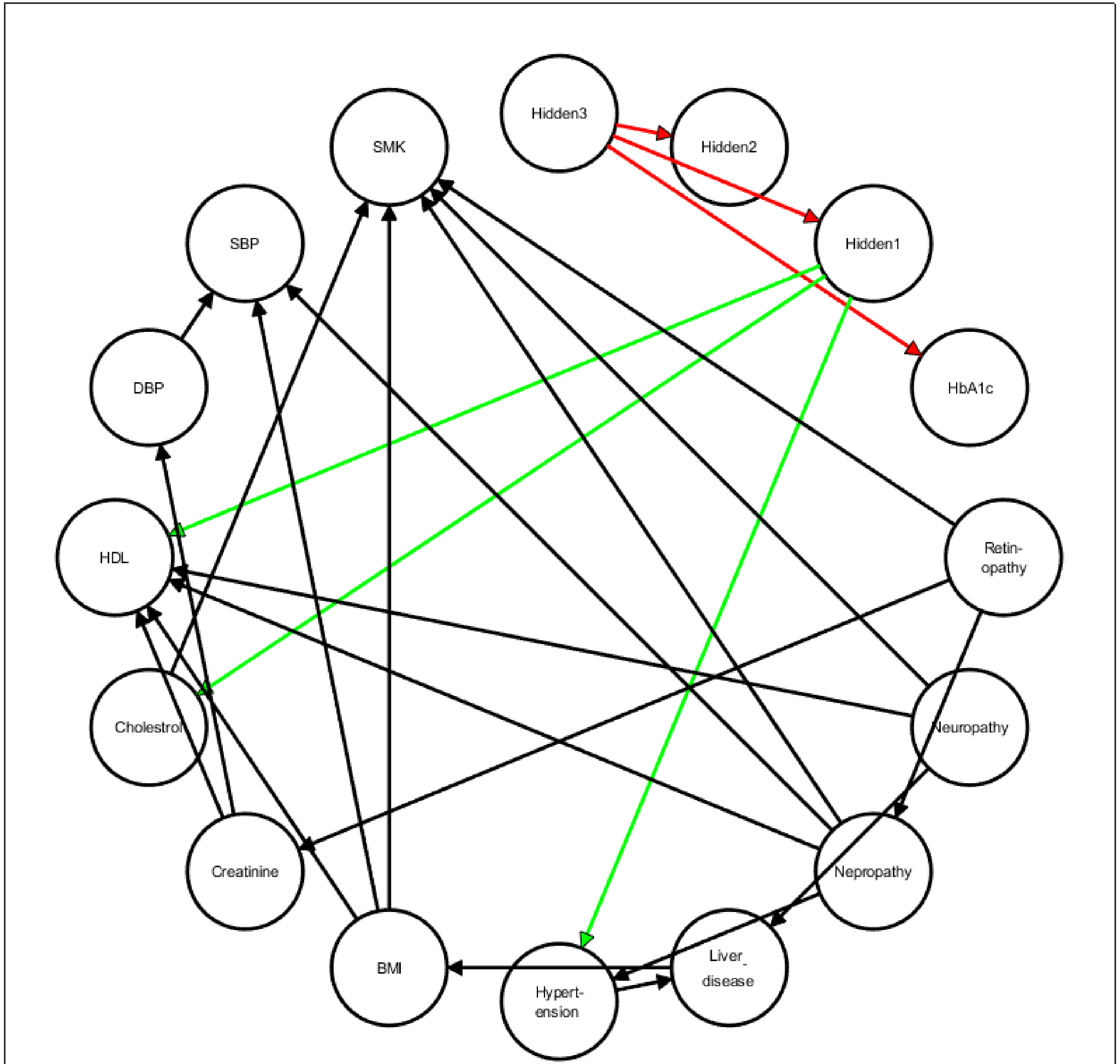
# Static structure with the second hidden variable




Leila.Yousefi@brunel.ac.uk

Step 3

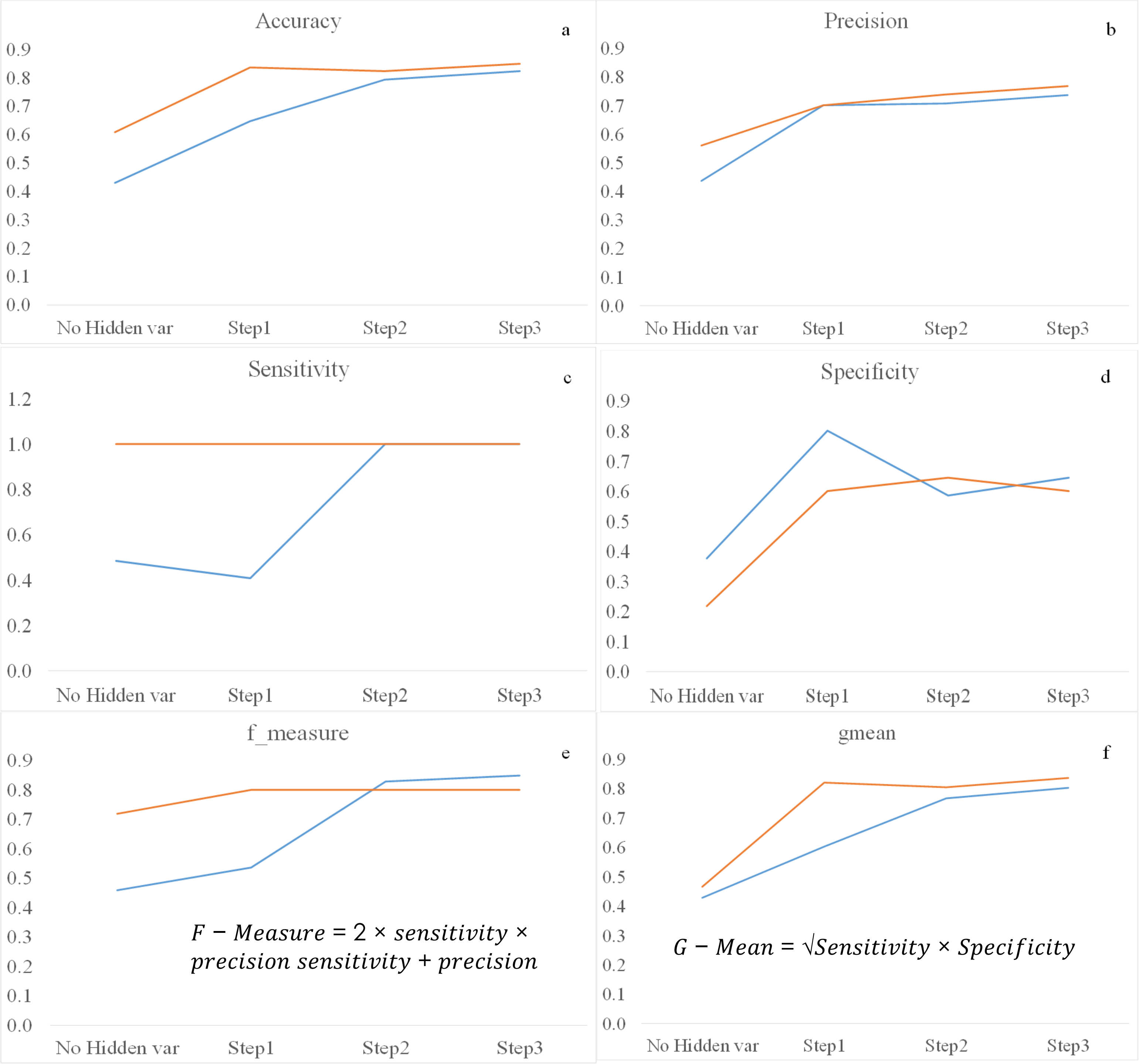
# Static structure with the third hidden variable




Leila.Yousefi@brunel.ac.uk

# Confusion Matrix Results

The step-wise approach with a generally improving classification accuracy of diagnosing targeted complications as a number of hidden variables are added (the blue line represent retinopathy and the red line represents liver disease).

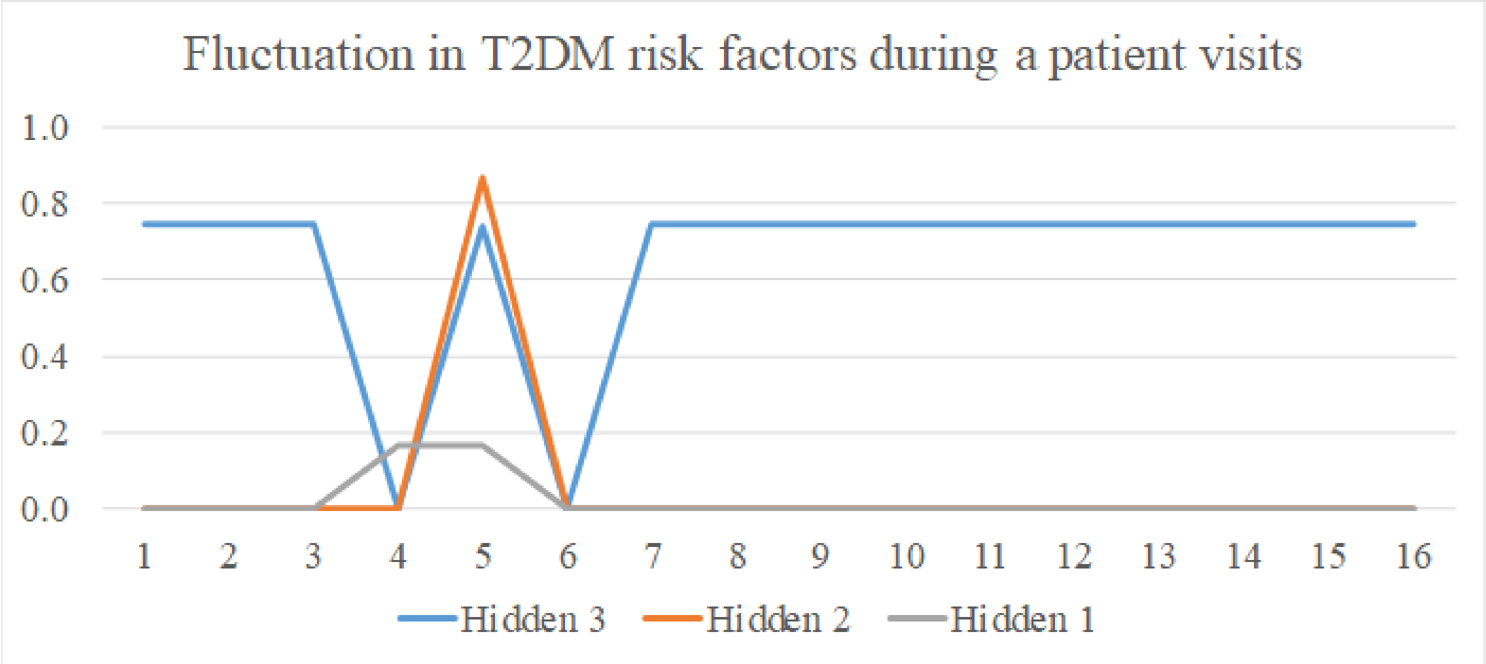


# Hidden variable fluctuation

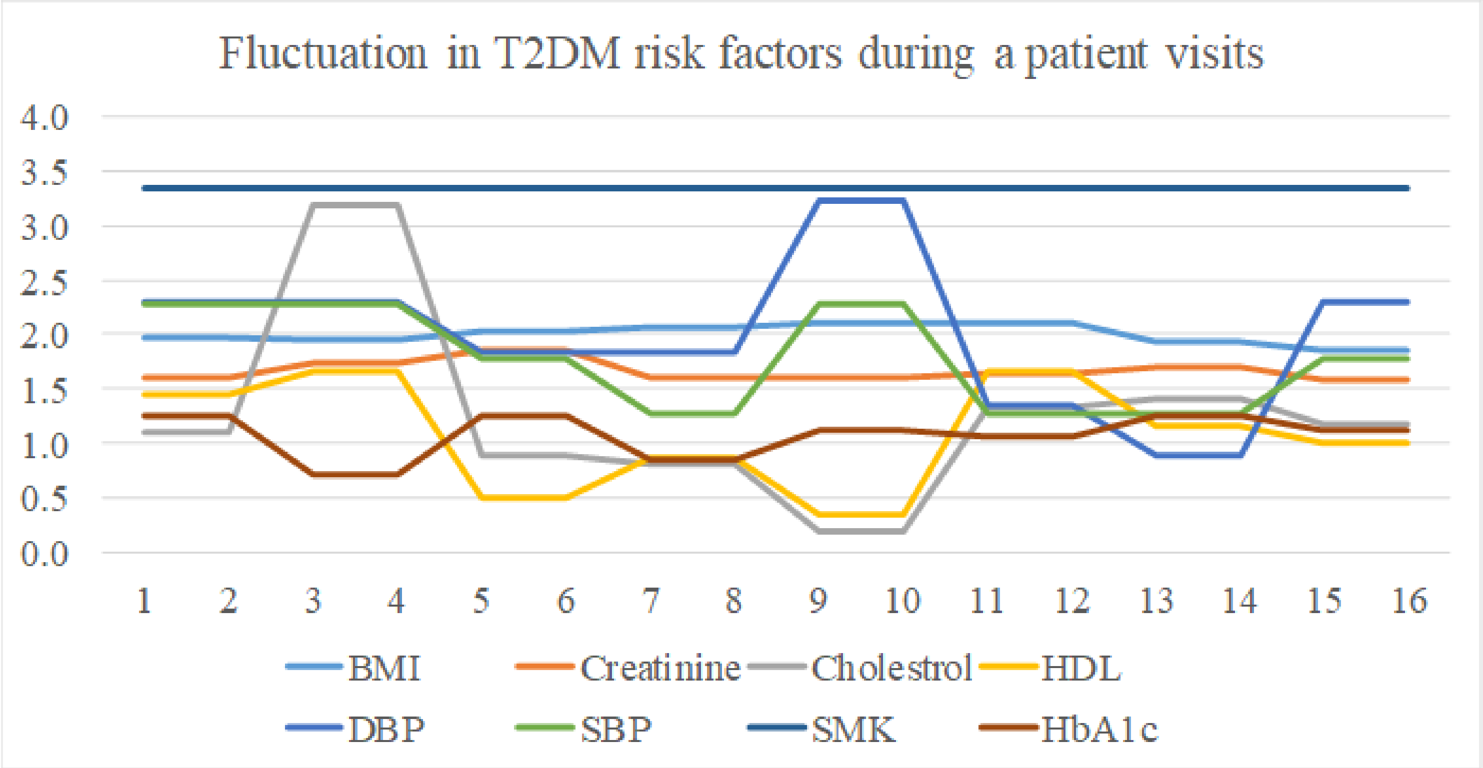
Predicted Latent Variable Pattern

VS

T2DM Complication and Features




a




b

Leila.Yousefi@brunel.ac.uk

# Inference: Query Rules

- QR =

P (complication=‘being at risk of retinopathy’ | Evidence)

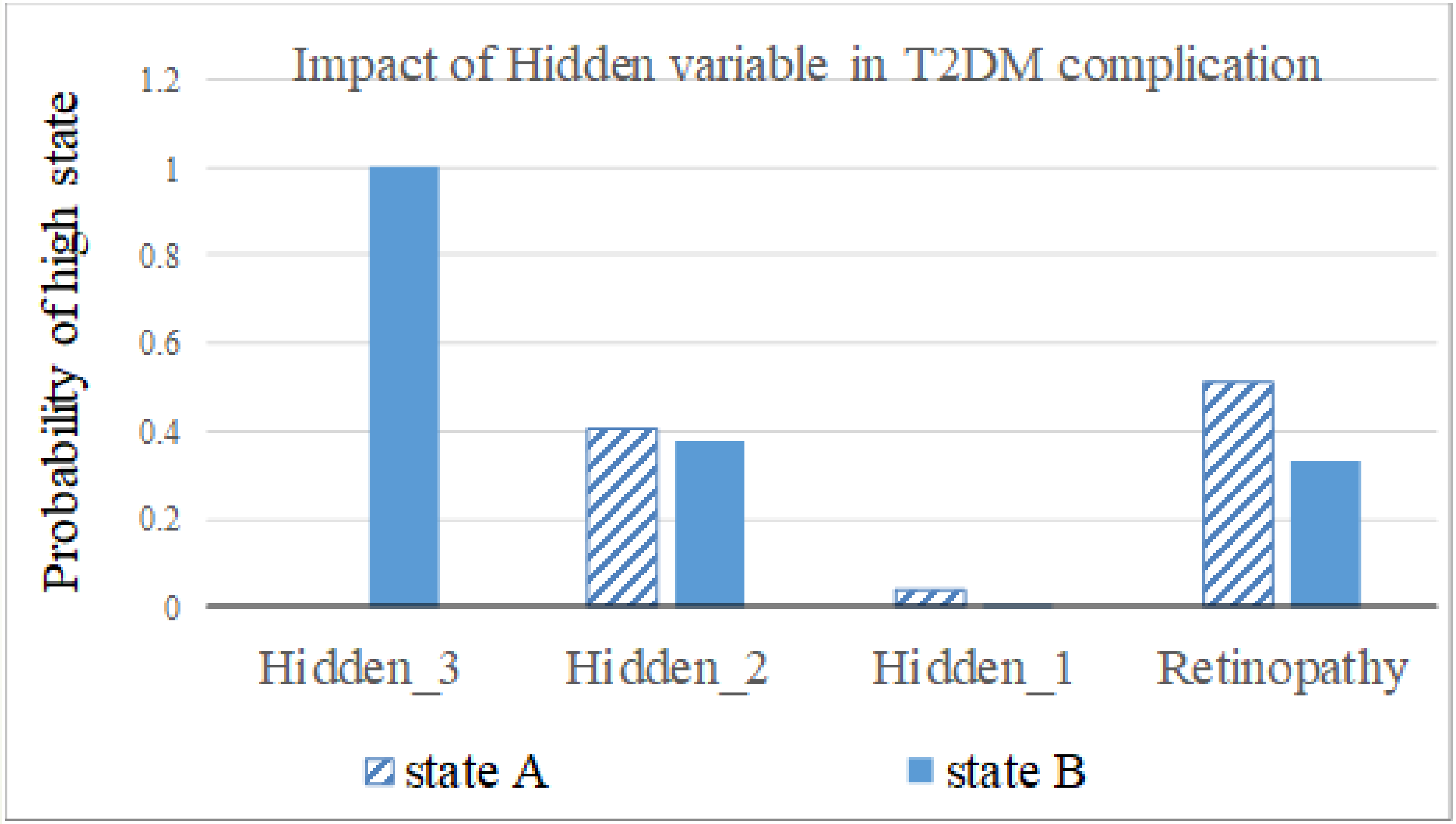
- Evidence =

{risk of having retinopathy is reduced =

’high level of hidden variable 3,

lower level of hidden variable 2,

and very low level of hidden variable 1}




The variations in the latent variable are affected by various comorbidities.

# Conclusion and Future Works

- Effectively integrates Bayesian methods with latent variables by adapting the prior probability of the event occurrence for future time points;
- The proposed method is more accurate than using one of hidden variable step or no hidden variables at all;
- Avoiding overfitting in the structure learning, using a stronger stopping rule in the step-wise approach;
- Exploiting mutual information metrics (Ebert,2007) to filter some of the hidden variable relationships;
- Discovering interesting dependencies between the latent variable and the observed variables;
- Interpreting the impact of hidden (latent) variables in finding temporal phenotypes in the presence of unmeasured diabetic disorders;
- Concentrating on the continuous investigation of features;
- Exploring Deep Learning methods;

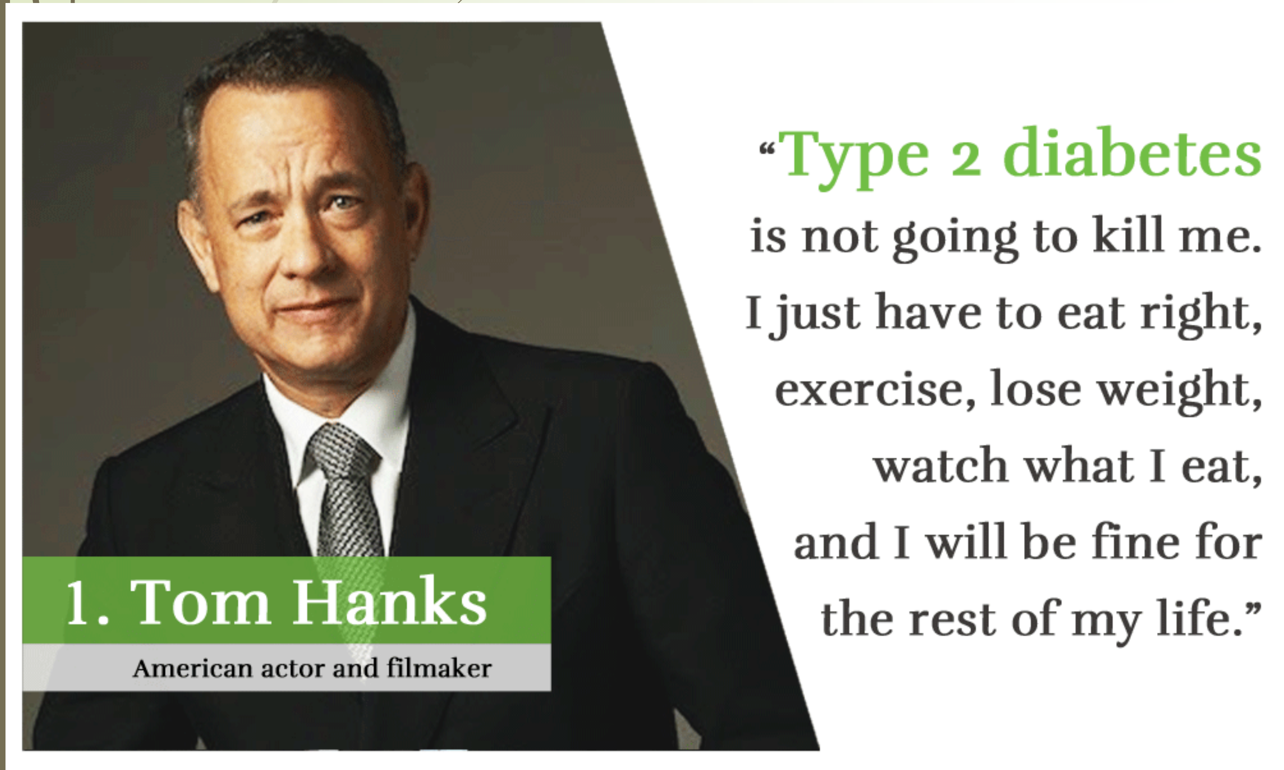
Leila.Yousefi@brunel.ac.uk

# References


- K. Murphy et al., “The bayes net toolbox for matlab,” Computing science and statistics, vol. 33, no. 2, pp. 1024– 1034, 2001.
- P. Spirtes, C. N. Glymour, and R. Scheines, Causation, prediction, and search. MIT press, 2000.
- S. Liang, S. Fuhrman, and R. Somogyi, “Reveal, a general reverse engineering algorithm for inference of genetic network architectures,” 1998.
- I. Ebert-Uphoff, “Measuring connection strengths and link strengths in discrete bayesian networks,” Georgia Institute of Technology, Tech. Rep., 2007.
- J. Pearl, “Probabilistic reasoning in intelligent systems. 1988,”San Mateo, CA: Kaufmann, vol. 23, pp. 33–34.
- L. Yousefi, L. Saachi, R. Bellazzi, L. Chiovato, and A. Tucker, “Predicting comorbidities using resampling and dynamic bayesian networks with latent variables,” in 2017 IEEE 30th International Symposium on Computer-Based Medical Systems (CBMS), June 2017, pp. 205–206.

# Thank you for listening!

# Any Question?




Leila.Yousefi@brunel.ac.uk