A Brief Introduction to Deep Learning and Bayesian Approach in Disease Prediction

IDA Research Group

Weibo Liu and Leila Yousefi

Dept. of Computer Science, Brunel University London,

London, UK

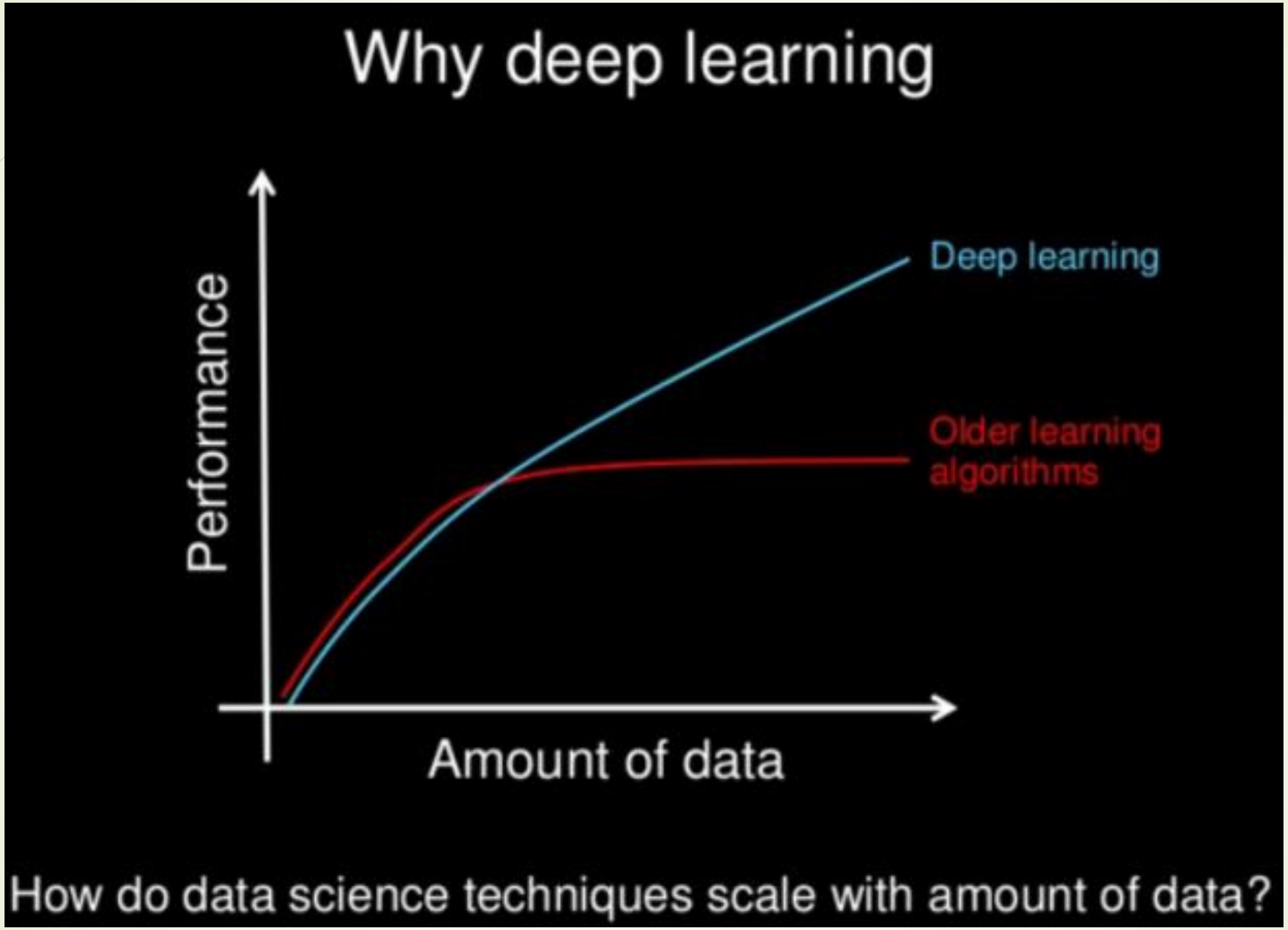
Outline

- Part 1:
  - Motivation
  - Three deep learning architectures
- Part 2:
  - Disease Prediction using Bayesian Network and Latent variable
  - Inductive Causation* algorithm
  - History of Bayesian Deep Learning
  - Bayesian deep Learning

Motivation

Deep Learning is a machine learning technique based on big data and aims to learning representations.

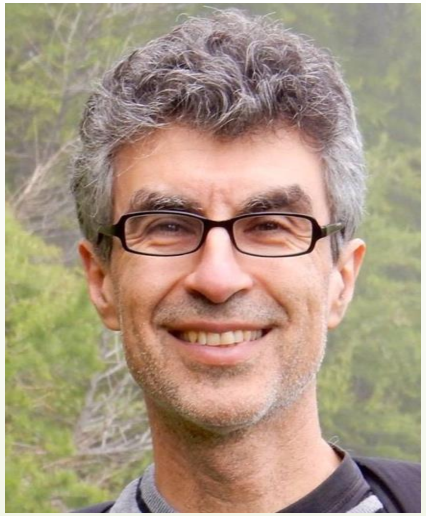
Since the proposal of a fast learning algorithm for deep belief networks in 2006, the deep learning techniques have drawn ever-increasing research interests.

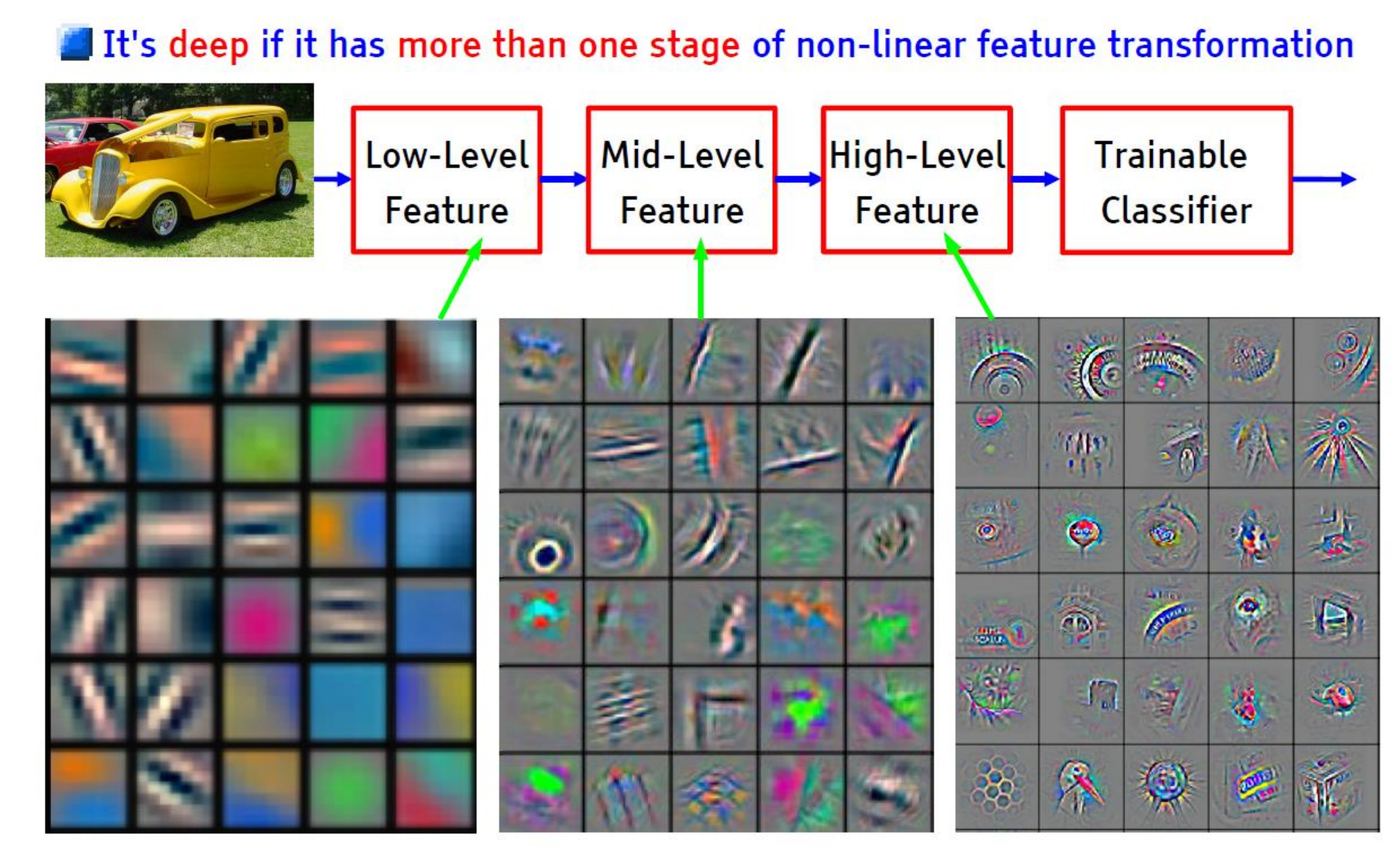


(Source: Andrew Ng, "Deep Learning, Self-Taught Learning and Unsupervised Feature Learning", 2013)

The Breakthrough in 2006:

1. Geoffrey E. Hinton, “A fast learning algorithm for deep belief nets”, University of Toronto.
2. Yoshua Bengio, “Greedy layer-wise training of deep networks”, University of Montreal.
3. Yann LeCun, “Efficient learning of sparse representations with an energy-based model”, New York University.





Feature visualization of convolutional net trained on ImageNet from [Zeiler & Fergus 2013]

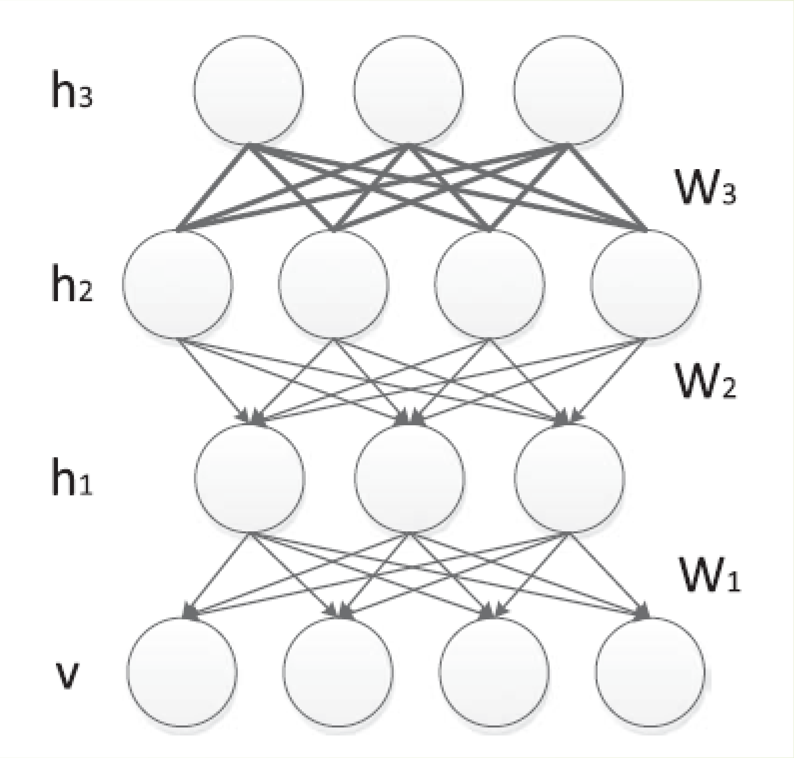
(Source: Yann LeCun, Deep Learning Tutorial, ICML, Atlanta, 2013-06-16)

Three deep learning architectures

1. Deep Belief Net

- The DBNs are composed of multiple layers of stochastic and latent variables and can be regarded as a special form of the Bayesian probabilistic generative model.

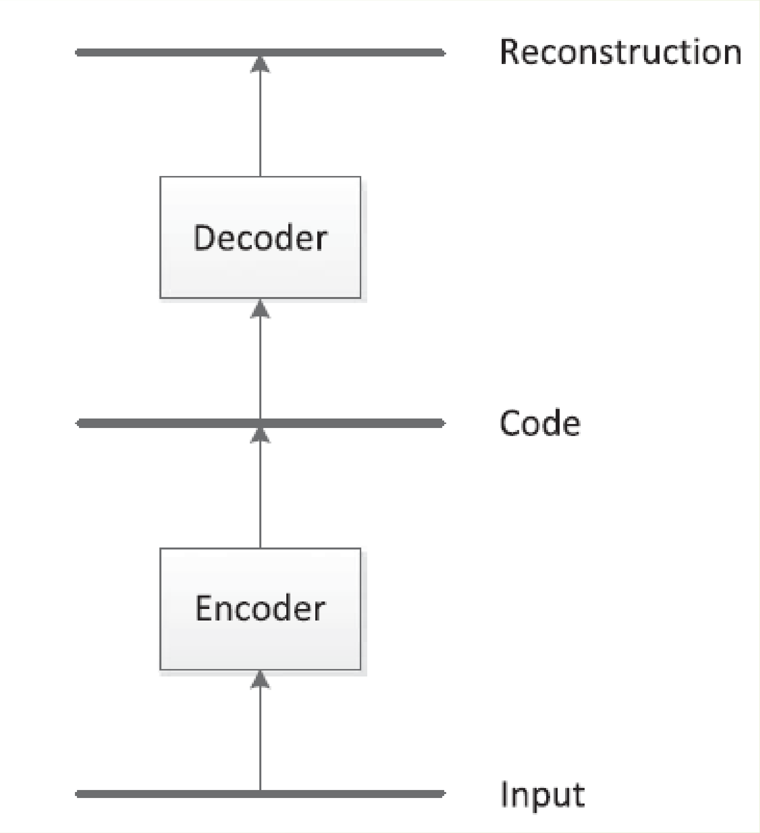
- Compared with ANNs, DBNs are more effective, especially when applied to problems with unlabeled data.



Three deep learning architectures

2. Autoencoder

- AE is an unsupervised learning algorithm used to efficiently code the dataset for the purpose of dimensionality reduction.
- The AE is a one-hidden-layer feed-forward neural network similar to the multilayer perceptron.



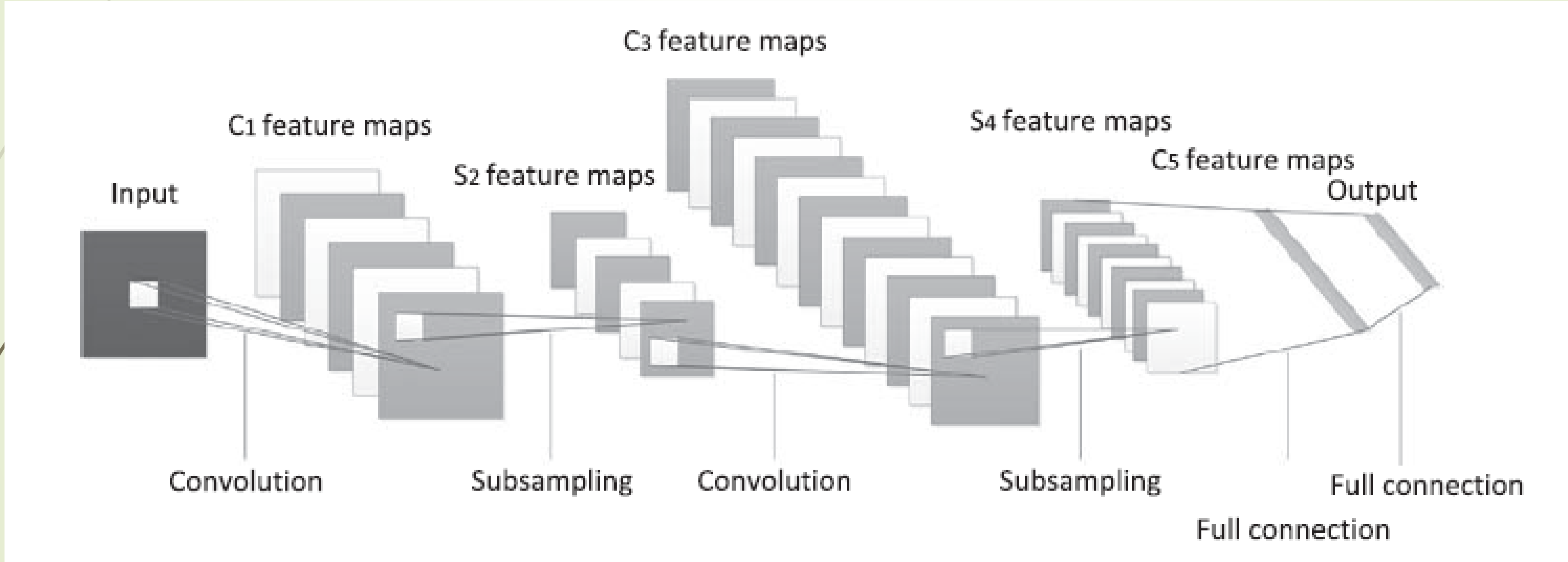
Three deep learning architectures

3. Deep Convolutional Neural Network

- CNNs are a subtype of the discriminative deep architecture and have shown satisfactory performance in processing two-dimensional data with grid-like topology, such as images and videos. The architecture of CNNs is inspired by the animal visual cortex organization.
- In CNNs, the convolution has replaced the general matrix multiplication in standard NNs. As such, the number of weights is decreased, thereby reducing the complexity of the network.

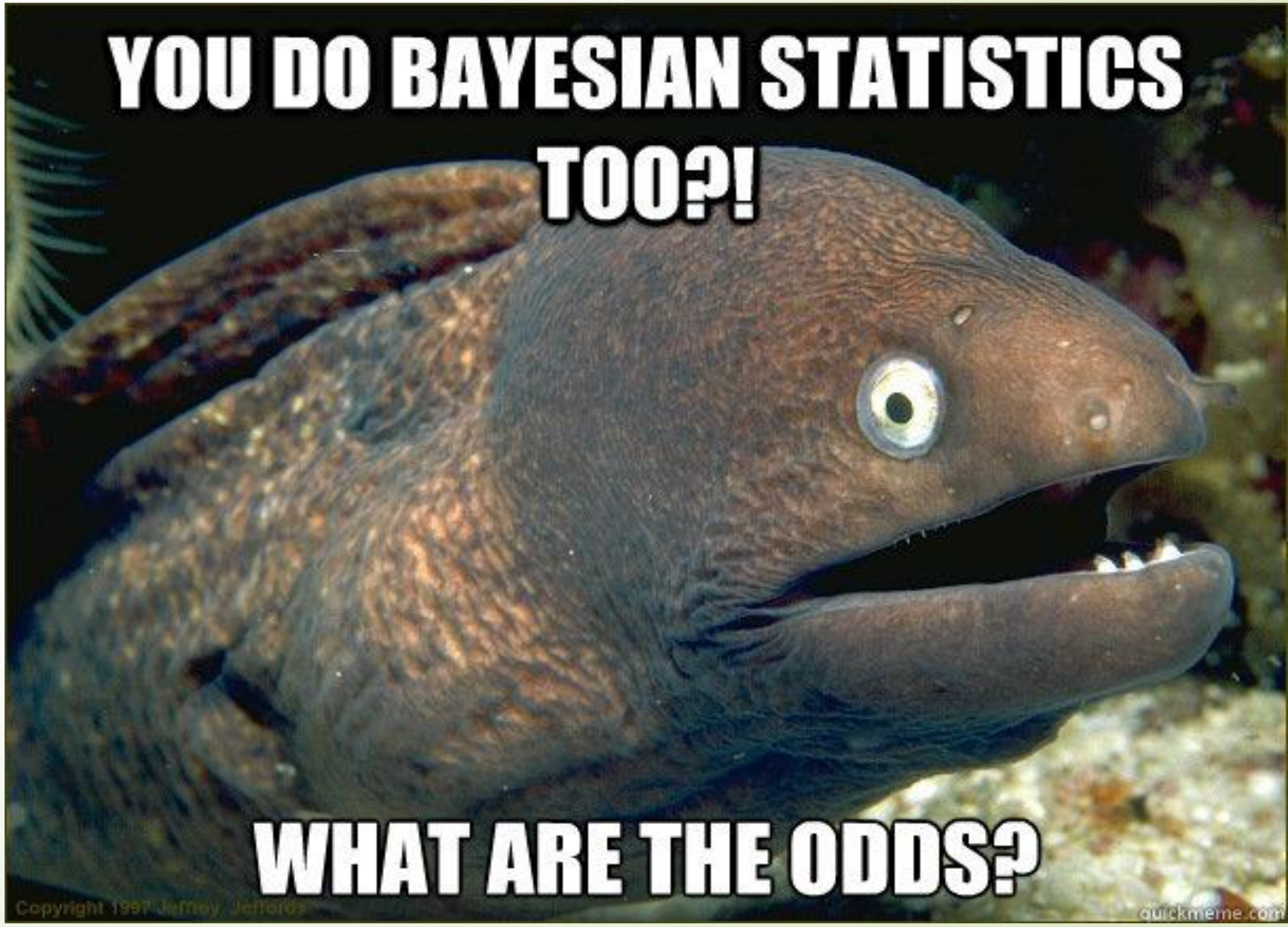
Three deep learning architectures

- Deep Convolutional Neural Network



Predicting Disease Complications

Using Hidden Variable Discovery for Learning Dynamic Bayesian Networks and Bayesian Deep learning



IDA Research Group

History of Bayesian Neural Network (BNN)

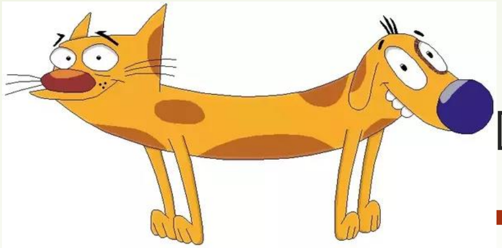
Approximate Inference in Bayesian Neural Networks

- Laplace Approximation (David MacKay-1992)
- Minimal Description Length (Hinton and Van Kamp-1993)
- Hamiltonian Monte Carlo (Radford Neal-1995)
- Ensemble Learning (Barber and Bishop-1998)
- Gal and Ghahramani :
  - Approximate Dropout NN and reparameterised posterior ad normal priors over network weights.
  - Optimising any Neural Network with dropout is equivalent to a form of approximate Bayesian Inference.
  - A network trained with dropout already is a Bayesian Neural Network!

Dropout is a regularization technique for reducing overfitting in NN by co-adaptations on training data. (Dropping out Hidden and observed units)

We need to approximate the weight of posterior in BNN

Bayesian Deep Learning



Bayesian Reasoning

Pros.

- A framework for inference and

decision making

- Unified framework for building,

Inference, prediction and decision making

- Explicit accounting for uncertainty and

variability outcomes.

- Robust to overfitting, tools for model

selection and composition.

Cons.

- Many coupled and linear models
- Potentially intractable inference, computationally expensive or long simulation time.

Deep Learning

- Rich non-linear models for classification and sequence prediction.
- Scalable learning using stochastic approximation and conceptually simple.
- Easily compostable with other gradient-based methods.
- Only estimates the points over confidently.
- Require large amount of labelled data.
- Hard to score models,

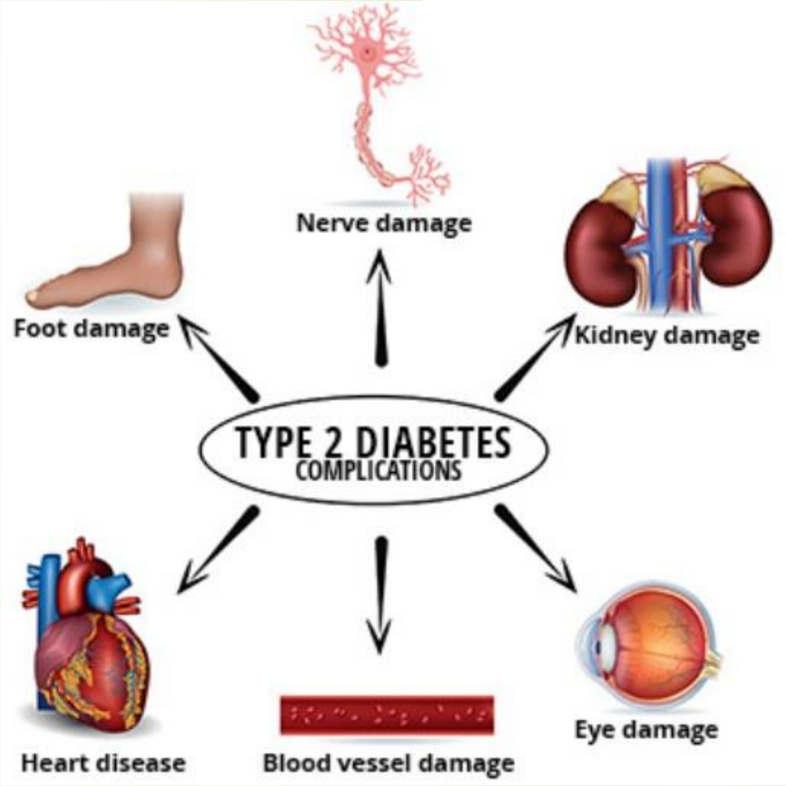
do selection and complexity

penalisation.

Natural to marry these approaches

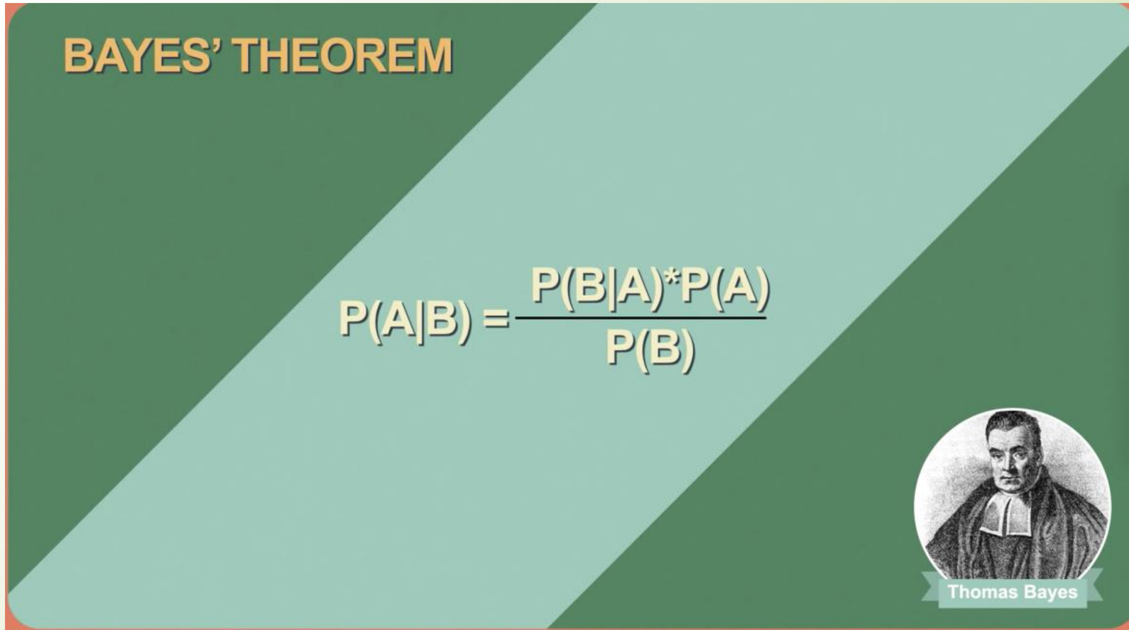
Diabetes & Its Complications

- Type 2 Diabetes Mellitus (T2DM), which is a non-insulin dependent diabetes or adult-onset diabetes.
- Type 2 Diabetes Mellitus (T2DM) most common form.
- Accounting for at least 90% of all cases.
- The World Health Organization (WHO) estimates that by 2030 ~550 million people suffering.
- Complications such as Eye and Liver Disease are common in Diabetes.
- Predicting these earlier very valuable but difficult.
- Using Latent variable to capture.



Introduction to Bayesian Modeling

- The theorem was invented by an english reverend Thomas Bayes (1701-1761) and published posthumously (1763).
- Given a prior state of knowledge or belief, it tells how to update beliefs based upon observations (current data).



Bayesian and what is Odds?

Likelihood ratio = the probability that a test is correct divided by the probability that it is incorrect

Positive Likelihood

$$LR- = \frac{1 - \text{sensitivity}}{\text{specificity}}$$

Positive Likelihood

$$LR+ = \frac{\text{sensitivity}}{1-\text{specificity}}$$

	Disease Positive	Disease Negative
Test Positive	A (True Positive)	B (False Positive)
Test Negative	C (False Negative)	D (True Negative)

	Disease Positive	Disease Negative
Test Positive	610	13,212
Test Negative	118	127,344
	Sensitivity = 83.8%	Specificity = 90.6%

From the study it was determined:
0.5% of women have breast cancer
99.5% of women do not have breast cancer

$$\text{Odds} = \frac{\text{P(something is true)}}{\text{P(something is false)}}$$

$$= \frac{0.005}{0.995}$$

$$= 0.005$$

BAYES' THEOREM
(applied to mammograms)

Post-test odds = Pre-test odds * LR+
= 0.43 * 8.9
= 3.8

Probability = Odds / (1+Odds)
= 3.8 / (1+3.8)
= 79%

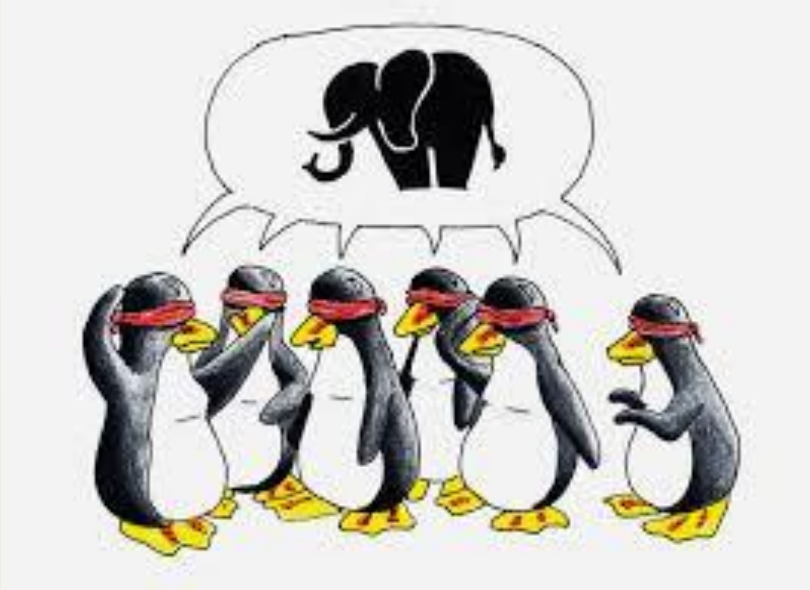
IC* (inductive causation) algorithm

- In the framework of Pearl's causality, algorithms IC and IC* provide a procedure to determine which causal connections among nodes in a network can be inferred from empirical observations.
- Even in the presence of latent variables, indicating the limits of what can be learned without active manipulation of the system.
- Established to analyze causal influences (effective connectivity) among T2DM features.
- Learn a partially oriented DAG (pattern) with latent variables
- The output P is an adjacency matrix, in which:
- P(i,j) = -1 if there is either a latent variable L such that i <-L-> j OR there is a directed edge from i->j.
- P(i,j) = -2 if there is a marked directed i-*>j edge.
- P(i,j) = P(j,i) = 1 if there is and undirected edge i—j.
- P(i,j) = P(j,i) = 2 if there is a latent variable L such that i<-L->j.

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

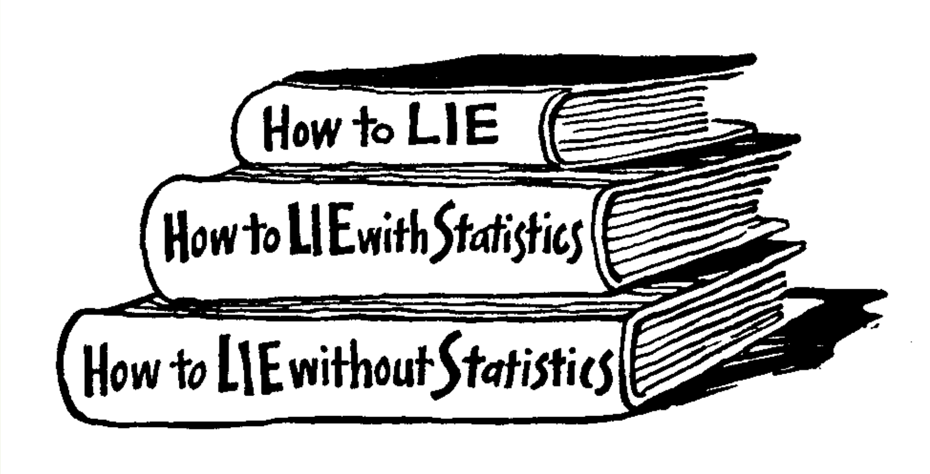
Latent Variable discovery

- IC* algorithm is a constraint based methods with an informative graph, which applies conditional independence analysis to infer casual structures.
- The learned DAG will not be unique
- Latent variable:
  - Some variables are unmeasured, called hidden or latent variables
  - The space of possible structures with latent variables is unbounded.

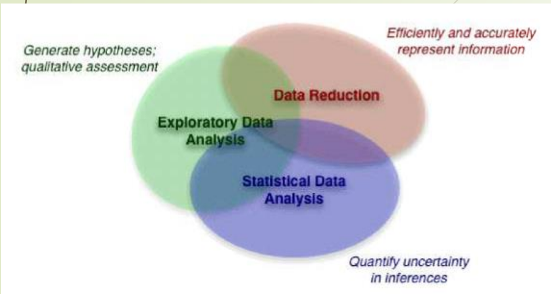


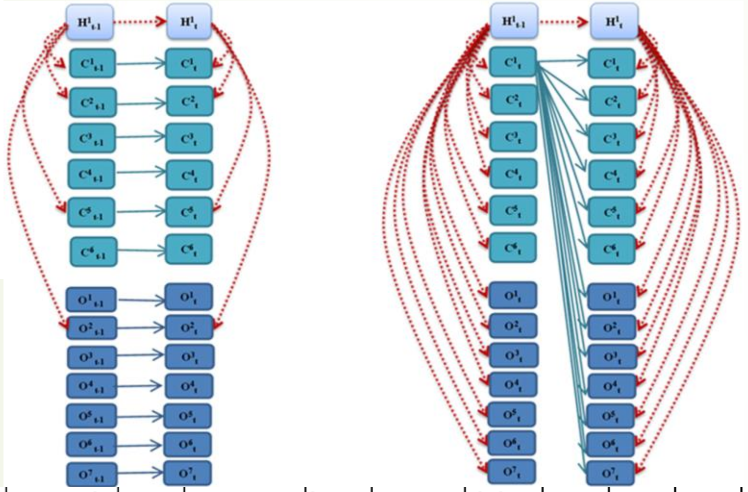
Step-wise Approach

- For all undiagnosed patients for a specific comorbidity Randomly chosen two consecutive time points of data [0 0].
- For all patients diagnosed with a specific comorbidity selected the two consecutive time points that represent the switch from no comorbidity to comorbidity [0 1].
- Randomly resample from the undiagnosed patients so that the same number of pairs appear as for diagnosed patients.

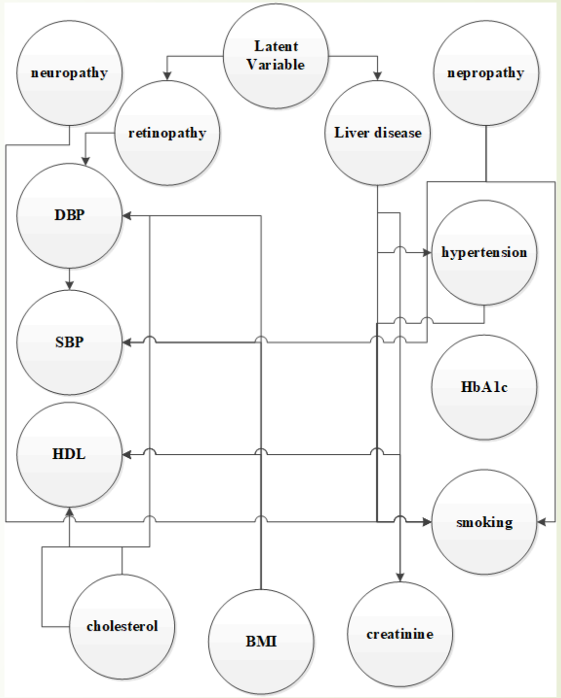


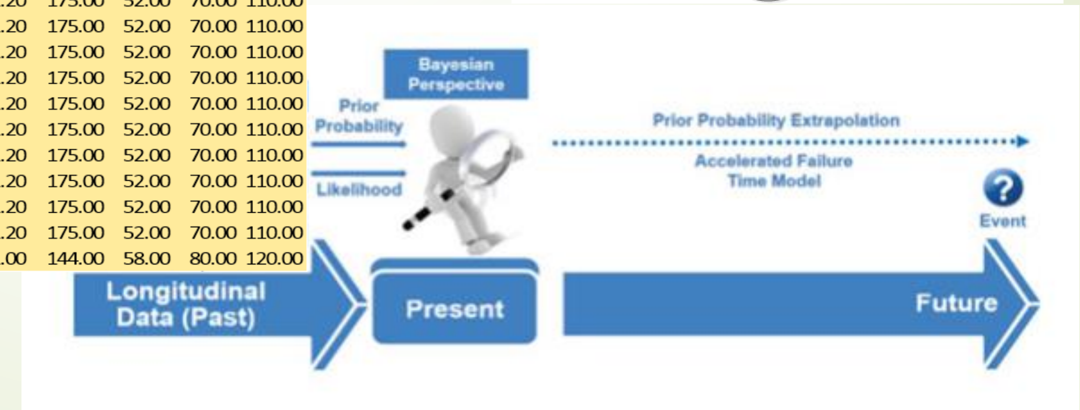
Dynamic Bayesian Network and Latent variable Data and structure





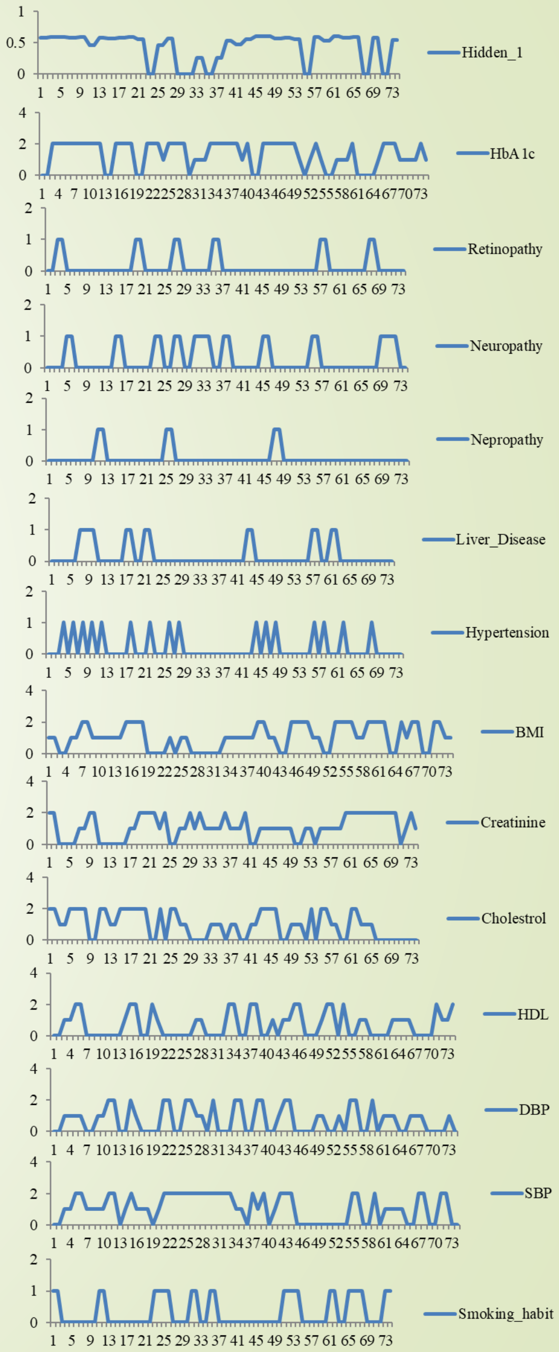
patient_id	HbA1c	Retinopathy	LIVER	Hypertention	bmi	creatinine	chol_tot	HDL	DBP	SBP
182.00	9.60	0	0	0	34.60	0.90	190.00	59.00	70.00	120.00
182.00	9.60	0	0	1	34.60	0.90	190.00	59.00	70.00	120.00
182.00	9.60	1	0	1	34.60	0.90	190.00	59.00	70.00	120.00
182.00	9.60	1	0	1	34.60	0.90	190.00	59.00	70.00	120.00
182.00	9.70	1	0	1	34.60	1.20	175.00	52.00	70.00	110.00
182.00	9.70	1	0	1	34.60	1.20	175.00	52.00	70.00	110.00
182.00	9.70	1	0	1	34.60	1.20	175.00	52.00	70.00	110.00
182.00	9.70	1	0	1	34.60	1.20	175.00	52.00	70.00	110.00
182.00	9.70	1	0	1	34.60	1.20	175.00	52.00	70.00	110.00
182.00	9.70	1	0	1	34.60	1.20	175.00	52.00	70.00	110.00
182.00	9.70	1	0	1	34.60	1.20	175.00	52.00	70.00	110.00
182.00	9.70	1	0	1	34.60	1.20	175.00	52.00	70.00	110.00
182.00	9.70	1	0	1	34.60	1.20	175.00	52.00	70.00	110.00
182.00	9.70	1	0	1	34.60	1.20	175.00	52.00	70.00	110.00
182.00	9.70	1	0	1	34.60	1.20	175.00	52.00	70.00	110.00
182.00	9.70	1	0	1	34.60	1.20	175.00	52.00	70.00	110.00
182.00	9.40	1	0	1	35.88	1.00	144.00	58.00	80.00	120.00





prediction Results for Liver Disease

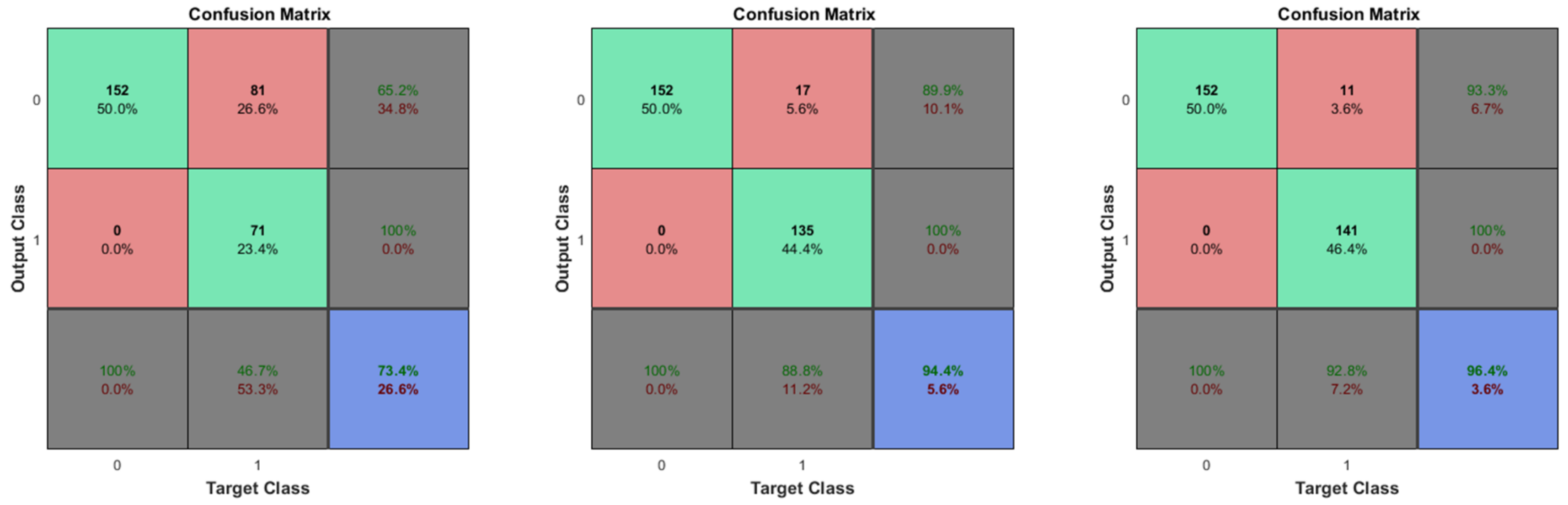
Predicted Latent variable
pattern VS T2DM
complication and Features



No Latent variable

First Step using IC*

Second Step Using IC*



Thank you for listening!

Any Question?