Naïve Bayes Model and Directed Graphical Model

Akash Choudhuri
Roll: 2019D014

M.Sc (2nd Year), Mathematics with Data Science
Institute of Mathematics & Applications, Bhubaneswar
akashchoudhuri.ima@iomaorissa.ac.in

July, 2021

Outline

- Conditional Independence and Bayes Theorem.
- The Naïve Bayes Model.
- Directed Graphical Models.
- Bayesian Networks.
- Application
- Programmed Example (if time permits).

Conditional Independence and Bayes Theorem

Axioms of Probability Theorem:

- For an event A, the probability of occurrence of that event A will be greater than or equal to zero.

$$p(A) \geq 0$$

- If there are disjoint events in a sample space, then the union of all events is the summation of individual probabilities.

$$P\left(\bigcup A_i\right) = \sum_i P(A_i)$$

- In case of an event involving the universal set has the probability of 1.

Important Concepts of Probability Theory

- **Random Variable:** A random variable is a measurable function which maps each outcome of the sample space to a Real value.
- **Joint Probability Distribution:** It finds the probability of many events occurring together by treating each event as a random variable. Eg, for 3 events X_1, X_2, X_3, Joint distribution is denoted by $P(X_1, X_2, X_3)$.
- **Marginal Probability Distribution:** Let X_1, X_2, X_3 be 3 random variables. Then the marginal distribution is:

$$P(x_1) = \sum \Sigma^p (x_1, x_2, x_3)$$

Introduction to Bayes Theorem

- **Conditional Independence:** We say an event X is conditionally independent of event Y given an event Z denoted as:

$$P(X|Y, Z) = P(X|Z).$$

- **Bayes Theorem:** Principled way of calculating a conditional probability without the joint probability.

In simpler terms, the result P(A|B) is referred to as the posterior probability and P(A) is referred to as the prior probability. Sometimes P(B|A) is referred to as the likelihood and P(B) is referred to as the evidence. This allows Bayes Theorem to be restated as:

$$\text{Posterior} = \text{Likelihood} * \text{Prior} / \text{Evidence}$$

The Naïve Bayes Model

Why ‘naïve’?

- This model uses Bayes Theorem with a small assumption that **there is independence among predictors**, ie, the presence of a particular feature in a class is unrelated to the presence of any other feature.

So, our Bayes Theorem formula is re-written by omitting the denominator (a littler bit of maths can show that and it reduces to:

By Bayes Theorem, $P(B|A)=(P(A|B)*P(B))/P(A)$

$=P(A|B)*P(B)$

Generalising the Equation,

$$P(c|X)=P(x_1|c)*P(x_2|c)*P(x_3|c)\ldots\ldots P(x_n|c)*P(c)$$

Naïve Bayes Classifier Algorithm for Discrete Data

- **Step 1:** Given a set of features D containing target variable T, calculate $P(X_i \mid Y_i)$ where

$$X_i, Y_i \in D \text{ and } X_i \neq Y_i$$

- **Step 2:** Calculate the Class Probabilities of Y given as P(Y).
- **Step 3:** Train the Model by finding the probabilities.
- **Step 4:** For a new set of features which is a subset of D, find the corresponding T.

Directed Graphical Models

Kinds of Graphical Models

- Undirected Graphical Models also known as Markov Random Fields.
- Directed graphical models also known as Bayesian (belief) networks. The important Characteristics of Bayesian Networks are:
  - Bayesian Networks require that the graph is a DAG (directed acyclic graphs).
  - No directed cycles allowed.

Bayesian Networks

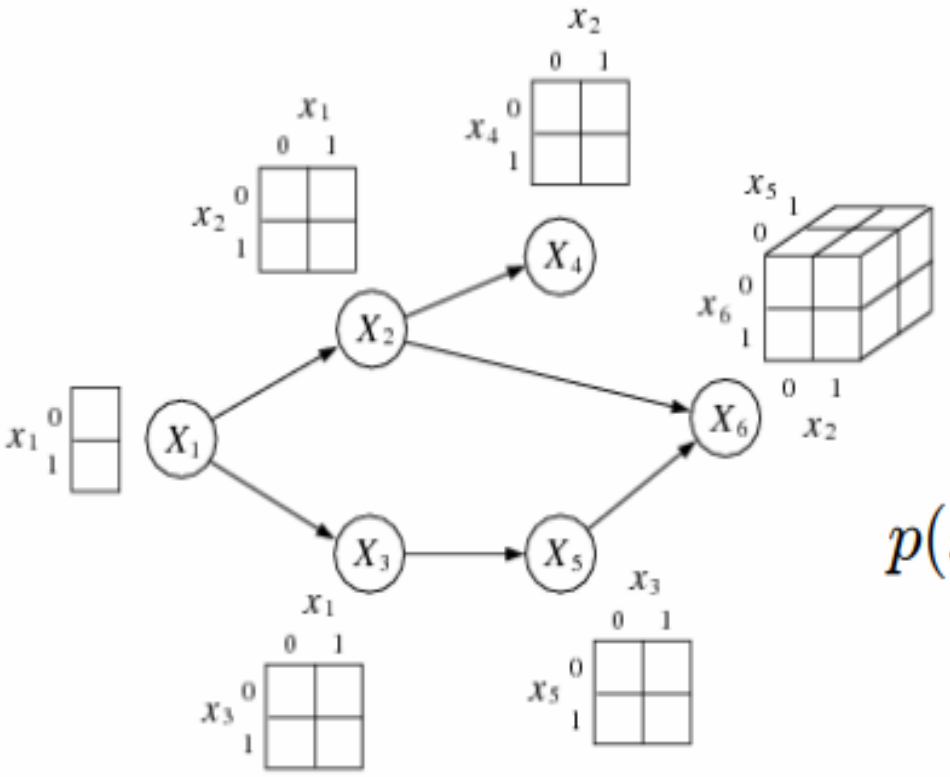
Bayesian Networks

- Judea Paul, who is credited with the invention of Bayesian Networks, won the Turing Award in 2011 for this discovery.
- A probability distribution factorizes according to a DAG if it can be written as:

$$P(x) = \prod_{j=1}^{d} P\left(x_j \middle| x_{\pi j}\right)$$

Where Π_j are the parents of j , and the nodes are ordered topologically (parents before children).

Continued….



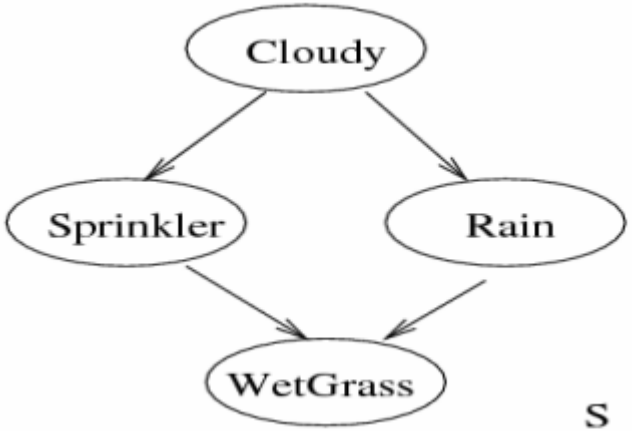
Each row of the conditional probability table (CPT) defines the distribution over the child's values given its parents values. The model is locally normalized.

$$p(x_{1:6}) = p(x_1)p(x_2|x_1)p(x_3|x_1)p(x_4|x_3)$$
$$p(x_5|x_2,x_3)p(x_6|x_2,x_5)$$

Example Bayesian Network

P(C=F)	P(C=T)
0.5	0.5

C	P(S=F)	P(S=T)
F	0.5	0.5
T	0.9	0.1



C	P(R=F)	P(R=T)
F	0.8	0.2
T	0.2	0.8

S	R	P(W=F)	P(W=T)
F	F	1.0	0.0
T	F	0.1	0.9
F	T	0.1	0.9
T	T	0.01	0.99

Continued

c	s	r	w	prob
0	0	0	0	0.200
0	0	0	1	0.000
0	0	1	0	0.005
0	0	1	1	0.045
0	1	0	0	0.020
0	1	0	1	0.180
0	1	1	0	0.001
0	1	1	1	0.050
1	0	0	0	0.090
1	0	0	1	0.000
1	0	1	0	0.036
1	0	1	1	0.324
1	1	0	0	0.001
1	1	0	1	0.009
1	1	1	0	0.000
1	1	1	1	0.040

- The joint distribution is computed using Naïve Bayes Model as:

$$p(C, S, R, W) = p(C)\, p(S|C)\, p(R|C)\, p(W|S, R)$$

- Prior that sprinkler is on:

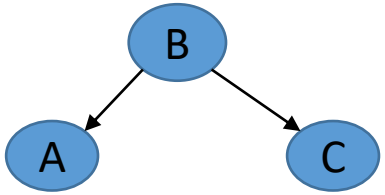
$$p(S=1) = \sum_{c=0}^{1} \sum_{r=0}^{1} \sum_{w=0}^{1} p(C=c, S=1, R=r, W=w) = 0.3$$

- Posterior that sprinkler is on given that grass is wet:

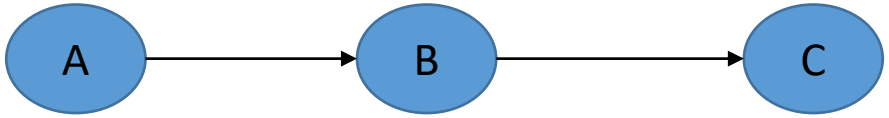
$$p(S=1|W=1) = \frac{p(S=1, W=1)}{p(W=1)} = 0.43$$

Conditional Independencies Implied from Bayesian Networks

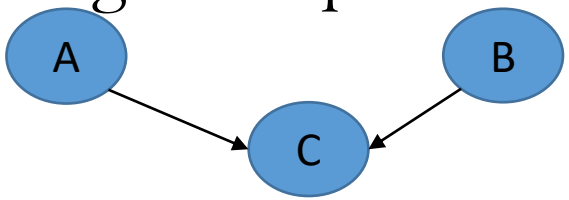
- **Common Parent:** Fixing B, A and C are decoupled in this network ($A \perp C \mid B$).



- **Cascade Structure:** In this network, $A \perp C \mid B$.



- **V- Structure:** Knowing C couples A & B.



D- Separation

Let A,B &C be non-overlapping sets of nodes (vertices) of a graph G. To ascertain $(A \perp B \mid C)$, consider all paths from any node in A to any node in B. Any such path is said to be block if it includes a node such that:

- The arrows on the path meet either head-to-tail or tail-to-tail and the node is in the set C.

OR

- The arrows meet head-to-head at the nodes and neither the node nor any of its descendants is in the set C.

Fact: If A is d-separated from B by C, then $(A \perp B \mid C)$ holds in the graph.

Application

A Bayesian Network Model for Predicting Post stroke Outcomes With Available Risk Factors

- An inference engine was constructed for post-stroke outcomes based on Bayesian network classifiers.
- The prediction system that was trained on data of 3,605 patients with acute stroke forecasts the functional independence at 3 months and the mortality 1 year after stroke.
- Feature selection methods were applied to eliminate less relevant and redundant features from 76 risk variables.
- Bayesian network with selected features by wrapper-type feature selection can predict 3-month functional independence with an AUC of 0.889 using only 19 risk variables and 1-year mortality with an AUC of 0.893 using 24 variables.

Dataset

- During admission, all patients were thoroughly investigated for medical history, clinical manifestations, and the presence of vascular risk factors.
- All registered patients underwent brain imaging studies including brain computed tomography (CT) and/or MRI.
- Stroke classification was determined during weekly conferences based on the consensus of stroke neurologists. Data including clinical information, risk factors, imaging study findings, laboratory analyses, and other special evaluations were collected. Along with these data, prognosis during hospitalization and long-term outcomes were also determined.

Methodology

- 76 Random variables were extracted from the data. Then a Bayesian Network was constructed using the formulae given before.
- Given a data set D with variable V_i, the observed distribution P_D is described as a joint probability distribution over D. The learning process now measures and compares the quality of Bayesian networks to evaluate how well the represented distribution explains the given data set. The log-likelihood is the basic common value used for measuring the quality of a Bayesian network as follows:

$$LL(B|D) = \sum_{V_i} log(P(V_i | \pi_B(V_i))),$$

Methodology cont.

- The algorithm searched the best Bayesian network based on the Bayesian information criterion. In this case maximum description length (MDL) score was used as evaluator. The MDL score is described as:

$$\mathrm{MDL} = -\mathrm{LL}(\mathrm{B}|\mathrm{D}) + \frac{\log N}{2} |B|$$

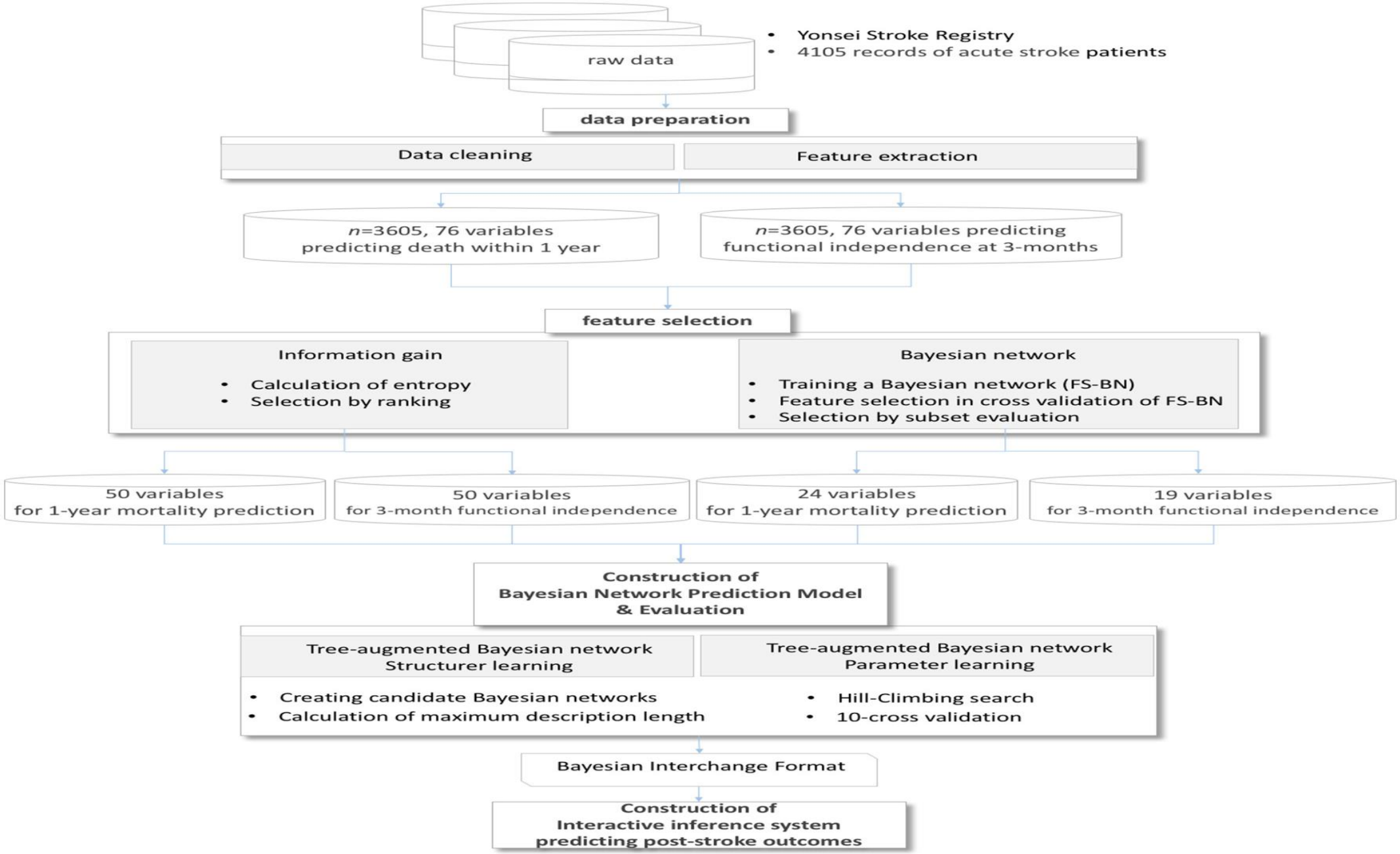
Where,

N is the number of instances in D,

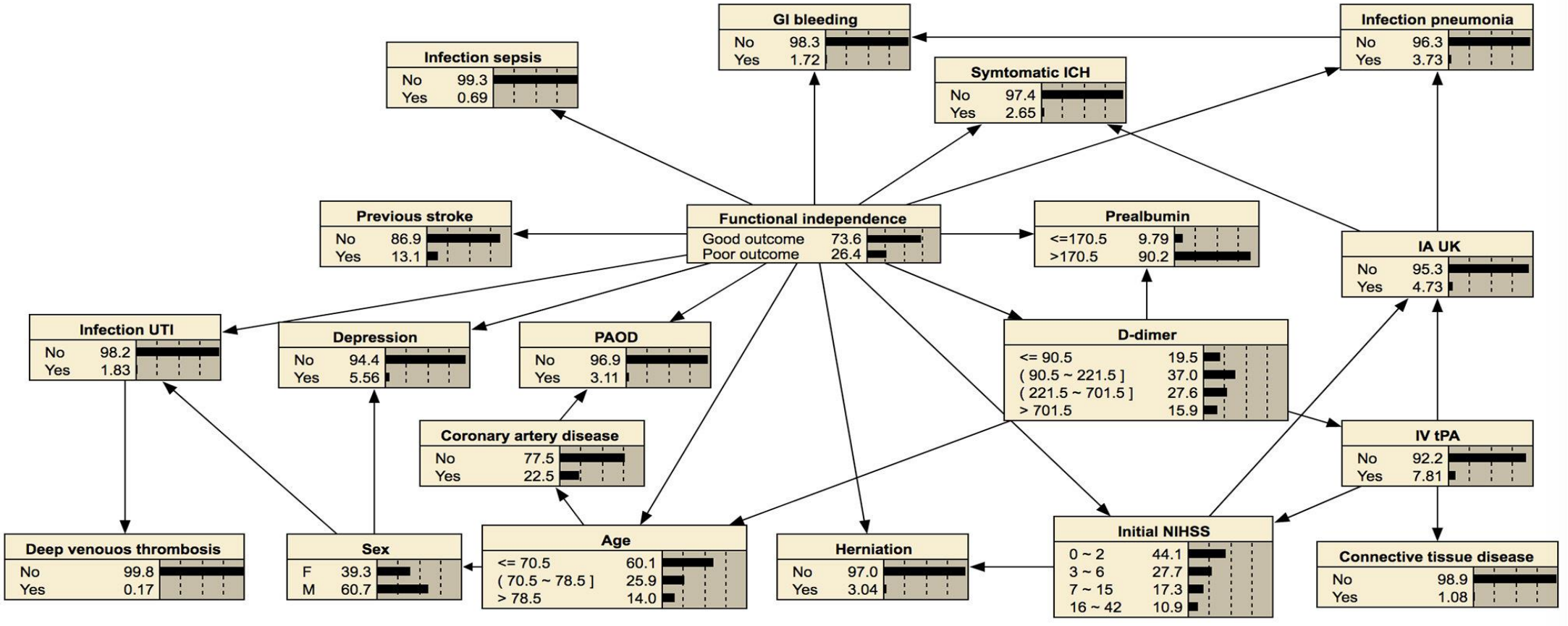
and |B| is the number of parameters in B.

The smaller the MDL score, the better the network.

- For the type of Bayesian network structure, tree-augmented network (TAN) structures were constructed that restrict the number of parents to two nodes

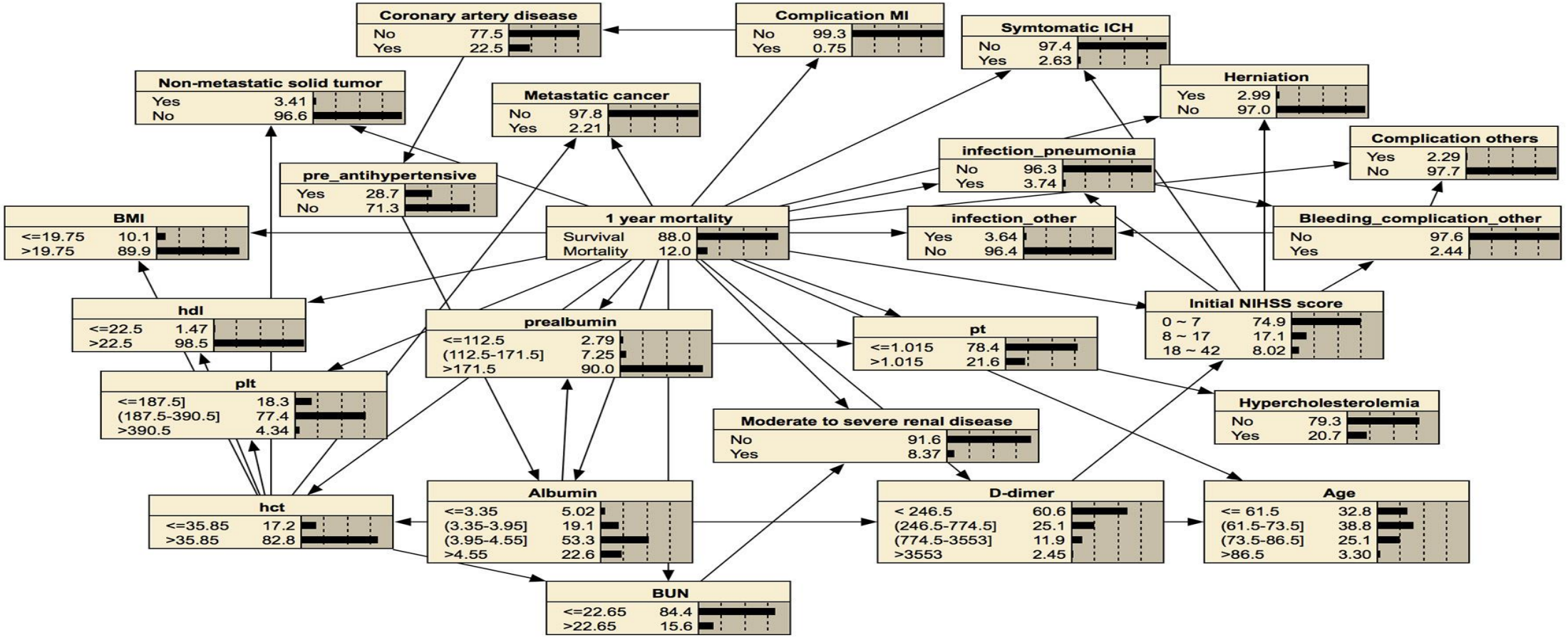

raw data
Yonsei Stroke Registry
4105 records of acute stroke patients
data preparation
Data cleaning
Feature extraction
n=3605, 76 variables predicting death within 1 year
n=3605, 76 variables predicting functional independence at 3-months
feature selection
Information gain
Calculation of entropy
Selection by ranking
Bayesian network
Training a Bayesian network (FS-BN)
Feature selection in cross validation of FS-BN
Selection by subset evaluation
50 variables for 1-year mortality prediction
50 variables for 3-month functional independence
24 variables for 1-year mortality prediction
19 variables for 3-month functional independence
Construction of Bayesian Network Prediction Model & Evaluation
Tree-augmented Bayesian network Structurer learning
Creating candidate Bayesian networks
Calculation of maximum description length
Tree-augmented Bayesian network Parameter learning
Hill-Climbing search
10-cross validation
Bayesian Interchange Format
Construction of Interactive inference system predicting post-stroke outcomes

Results



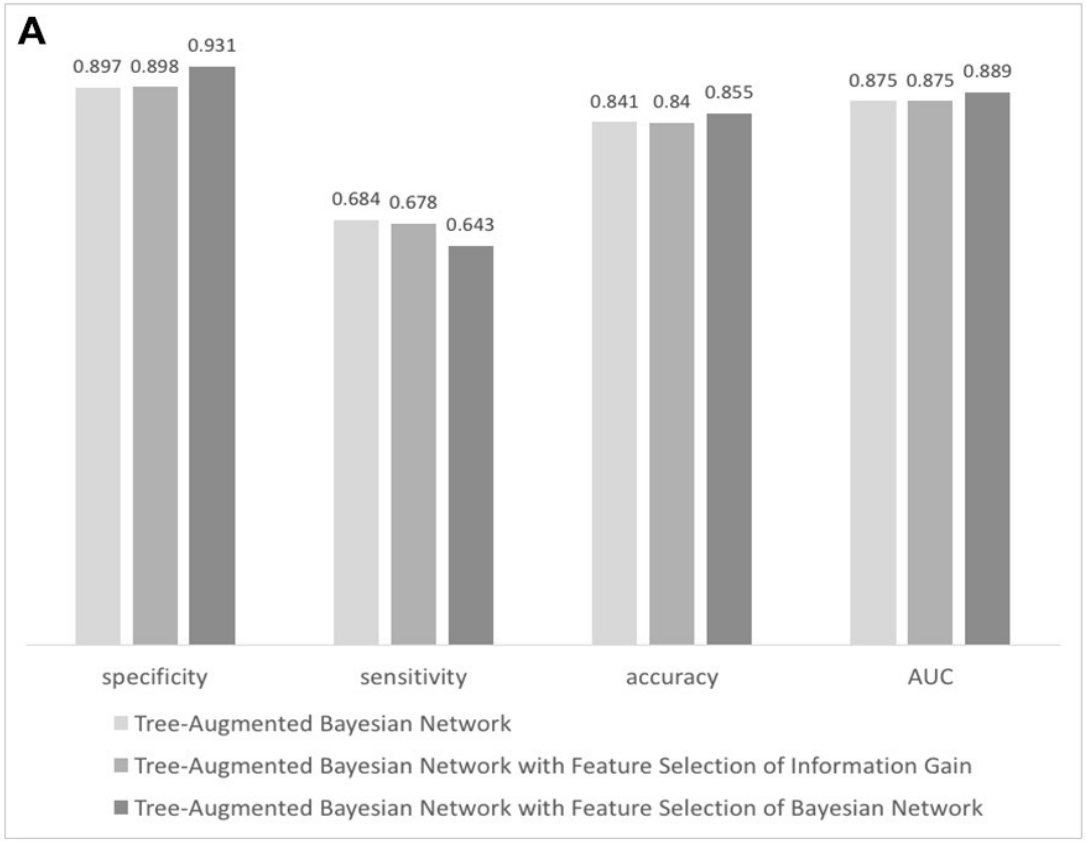
Bayesian network for predicting functional independence at 3 months. The tree-augmented Bayesian network used 19 variables selected by the wrapper of the Bayesian network for prediction.

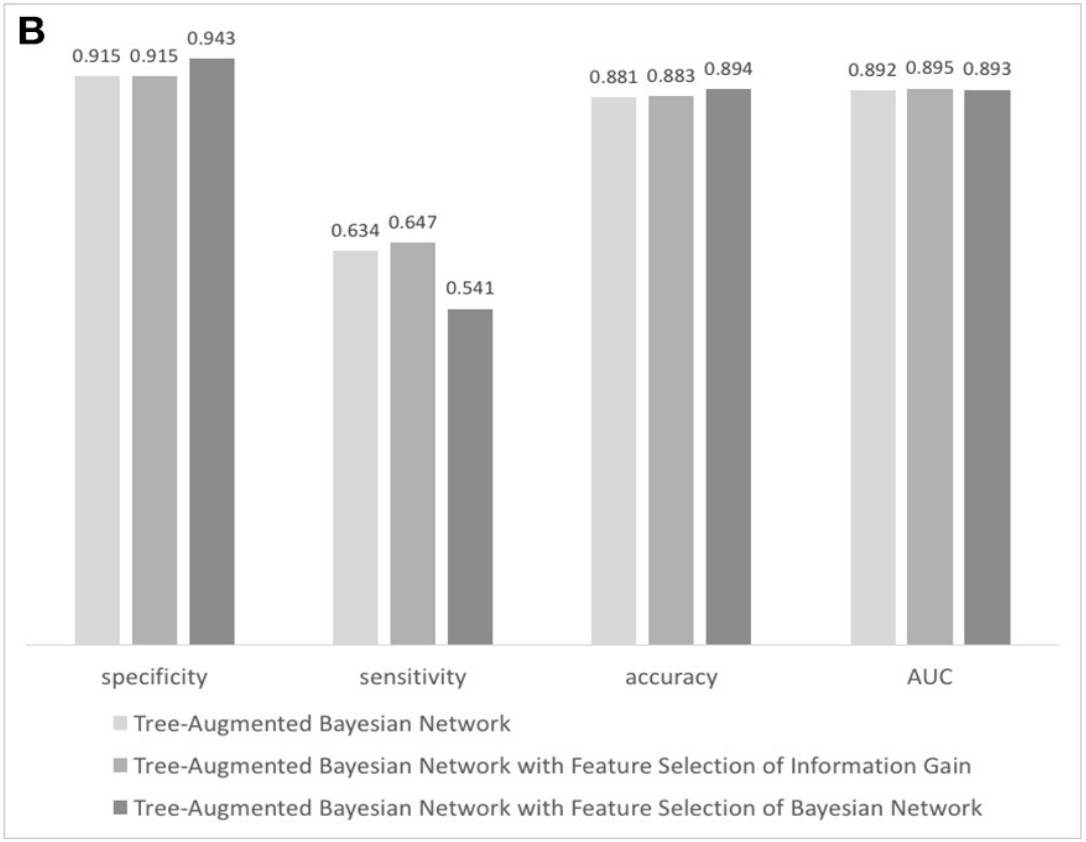
Results Cont.



Bayesian network for predicting 1-year mortality. The tree-augmented Bayesian network used 24 variables selected by the wrapper of the Bayesian network for prediction.

Performance Evaluations





Performance evaluation of Bayesian network-based classifiers: (A) performance of classifiers forecasting 90-day functional independence and (B) performance of classifiers for 1-year mortality prediction.

Programmed Example

References

- https://www.kaggle.com/johnoliverjones/naive-bayesian-network-with-7-features/comments
- https://www.kaggle.com/c/porto-seguro-safe-driver-prediction
- https://www.cs.ubc.ca/~murphyk/Teaching/CS540-Fall08/L15DGM.pdf
- https://www.frontiersin.org/articles/10.3389/fneur.2018.00699/full

Thank You

Questions/ Queries? Do reach out to me!