Post-Operative Life Expectancy of Lung Cancer Patients Predicted by Bayesian Network Model

Kittipat Sriwong, Kittisak Kerdprasop, and Nittaya Kerdprasop

***Abstract*—Data mining is the computational process to find potentially useful knowledge from the stored information and then use the discovered knowledge to predict or classify the new data item that its target class was unknown. Among many available algorithms to do data classification or prediction, Bayesian network (BN) is one of the most accurate methods. BN is acyclic directed graph that models probabilistic dependencies among variables with a conditional probability distribution. In this research, we propose prediction model of the postoperative survival in the lung cancer patients using BN. Currently, cancer is one of the leading causes of morbidity and mortality worldwide. The top causes of cancer death is lung cancer. Lung cancer surgery is one of treatment methods, but this method is risky. Sometimes patients died after the surgery. We thus study the surgery risk using BN. We show the performance of the proposed BN technique through the specific application for predicting post-operative life expectancy in the lung cancer patients from the Wroclaw Thoracic Surgery Centre, Poland. The experimental results show that the BN with appropriate discretization and learning scheme can predict the one year survival after surgery as accurate as 91.28%.**

***Index Terms*—Data prediction, Bayesian network, Lung cancer, discretization, tree-augmented Na ïve Bayes.**

I. Introduction

Data mining is about finding hidden and useful knowledge from the stored information or existing database. The discovered knowledge can be in the form of new facts or novel condition of knowing something through the patterns or relationships among the stored data records and attributes [1]. There are many types of data mining tasks such as data classification, clustering, and other analysis tasks. Prediction and classification are kinds of data classification category. More than eighty percent of data mining tasks applied in real applications are data classification and it is also the focus of our work.

The popular techniques in the data classification include artificial neural network (ANN), Bayesian networks (BNs), decision tree, logistic regression, support vector machine (SVM), and many more. There is no single technique that is good for any application. However, BN is recently the most applicable technique for medical science to help predicting treatment outcome and disease diagnosis. The extensive use of BN is because of its overall high performance on data classification [2], [3]. The main concept of BNs is trying to create acyclic directed graphs modeling base on Bayes' theorem for computing the posterior probability distribution.

For medical science, BNs had been used to diagnose the liver disorders [4], to identify the fertile days of a woman's monthly cycle [5], to quantify risk for histo-pathologic cervical pre-cancer and cervical cancer [6], to predict short-term and long-term left ventricular assist device (LVAD) mortality [7], and to be used in the medical decision support system built from patient questionnaires [8].

In this research, we propose the techniques to build BN models for predicting the post-operative life expectancy in the lung cancer patients. In our work, the BN learning has been done through the tree augmented naive-Bayes (TAN) learning algorithm.

II. Background Theories

A. Bayesian Network

Bayesian networks (BNs), also known as Bayesian belief networks or Bayes nets for short, are acyclic directed graphs modeling probabilistic dependencies among variables with a conditional probability distribution (or CP table) for each node [9]. Each node in the graph represents a random variable, and edges represent direct correlations between the variables. BNs combine principles from graph theory, probability theory, computer science, and statistics.

Fig. 1(a) shows a simple BN model consisting of three nodes: rain, sprinkler, and grass wet. The node rain represents the presence/absence of rain. The node sprinkler captures the on/off of watering event. The grass wet represents the fact of grass being wet.

This network models the influencing situation among the three nodes such that the absence of rain causes the sprinkler to operate and the observed outcome is that the grass is wet. Relations like this can be quantified numerically by means of conditional probability distributions shown as the table accompanying each node. BNs are interesting because of their capability of what-if scenarios. The evidence of grass wet can be analyzed for its cause by adjusting probability of the BN. In Fig. 1(b), when it is observed that the grass is wet (grass wet = T is set to 100%), then it is highly probable (with probability = 0.591, or 59.1%) that this is because the sprinkler is on.

B. Performance Evaluation of the Model

To evaluate performance of prediction or classification model, we use the accuracy metric for assessing the model's performance. The computation of this metric is based on the

Manuscript received March 2, 2018; revised May 14, 2018. This work has been supported by grants from Suranaree University of Technology through the funding of Data Engineering and the Knowledge Engineering Research Units.

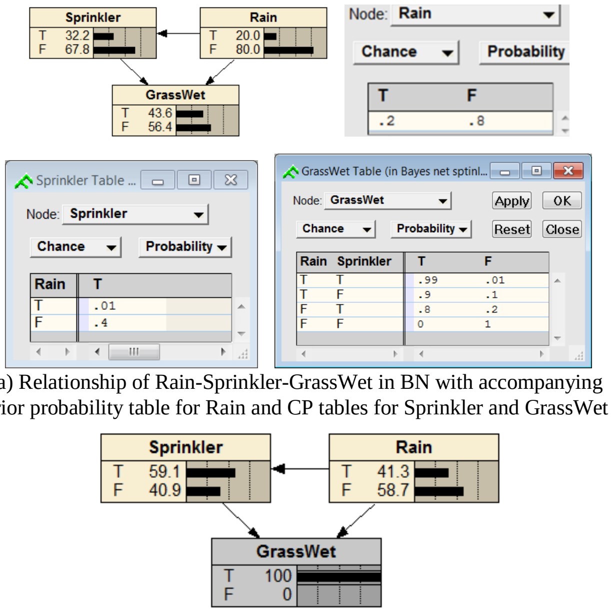
K. Sriwong is with the School of Computer Engineering, Suranaree University of Technology (SUT), 111 University Avenue, Muang, Nakhon Ratchasima 30000, Thailand (e-mail: kittipat.re@gmail.com).

K. Kerdprasop is with the School of Computer Engineering and Knowledge Engineering Research Unit, SUT, Thailand (e-mail: kerdpras@sut.ac.th).

N. Kerdprasop is with the School of Computer Engineering and Data Engineering Research Unit, SUT, Thailand (e-mail: nittaya@sut.ac.th).

doi: 10.18178/ijmlc.2018.8.3.700

values in confusion matrix [10] as shown in Table I.



T	F
.2	.8

Rain	T
T	.01
F	.4

Rain	Sprinkler	T	F
T	T	.99	.01
T	F	.9	.1
F	T	.8	.2
F	F	0	1

(a) Relationship of Rain-Sprinkler-GrassWet in BN with accompanying prior probability table for Rain and CP tables for Sprinkler and GrassWet

(b) Using BN to infer back the probable cause of grass being wet; it is because of the sprinkler is on = 59.1%, while because of the rain = 41.3%

Fig. 1. Example of Bayesian network capturing cause-effect model.

TABLE I: CONFUSION MATRIX FOR TWO CLASS CLASSIFICATION

		Actual Data	
		Positive	Negative
Predicted Data	Positive	TP	FP
	Negative	FN	TN

Accuracy is the proportion of the total number of predictions that were correctly made by of the classification model to all predictions. The computation is shown as in equation (1).

$$\text{ycaruccA} = \frac{(TP + TN)}{(TP + FN + FP + TN)} \qquad (1)$$

where:

TP is the number of correct predictions that an instance is positive class and actual class is positive class.

TN is the number of correct predictions that an instance is negative class and actual class is negative class

FP is the number of incorrect predictions that an instance is positive class but actual class is negative class

FN is the number of incorrect of predictions that an instance negative but actual class is positive class.

C. ROC Curve

In a Receiver Operating Characteristic (ROC) curve, the true positive rate (or sensitivity) is plotted in accordance to the false positive rate (or 1-specificity) for different cut-off points. Therefore, the closer the ROC curve is to the upper left corner, the higher the overall accuracy of the test [11]. Accuracy can also be measured by the area under the ROC curve (called the AUC). An AUC value equals to 1 represents a perfect test; whereas, an area of 0.5 represents a worthless test because it is just the same as random guess.

III. MODEL CREATION METHOD

A. Framework for Model Creation

In this research, we designed the process to create and assess the prediction model for the post-operative life expectancy in the lung cancer patients with Bayesian networks using the tree augmented naive-Bayes (TAN) learning algorithm. The model objective is to predict chance for the one year survival of patients after thoracic surgery. The steps in our model creation and evaluation is shown in Fig. 2.

From Fig. 2, we can describe our proposed research framework as follows. The data used in our study contain both categorical and continuous values. BN can handle only categorical values. We thus firstly discretize continuous variables to be discrete variables before applying the next step in our classification process. We experiment with three kinds of discretization: using hierarchical method for obtaining intervals, using uniform width, and using uniform counts.

To create model, we use train data after discretization for learning BN structure by applying the TAN learning algorithm. After completion, the algorithm creates BN model with appropriate structure of nodes. We then evaluate model with the rain data. Then, compute overall classification accuracy to assess the model's performance.

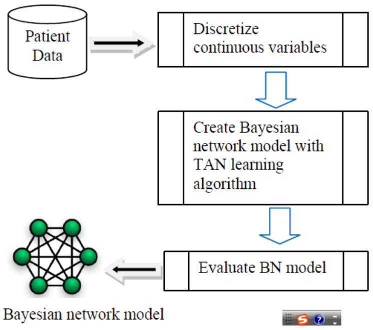

Patient Data
Discretize continuous variables
Create Bayesian network model with TAN learning algorithm
Evaluate BN model
Bayesian network model

Fig. 2. The research framework for building the BN model.

B. Dataset

To experiment our proposed prediction method, we used thoracic surgery dataset from the UCI Machine Learning Repository [12]. The dataset was collected retrospectively at Wroclaw Thoracic Surgery Centre for patients who underwent major lung resections for primary lung cancer in the years 2007 to 2011. The Centre is associated with the Department of Thoracic Surgery of the Medical University of Wroclaw and Lower-Silesian Centre for Pulmonary Diseases, Poland.

This dataset has 470 data instances and 17 attributes with 2 different classes of T (true) and F (false). The class T means the risk of not survive one year critical period after surgery, and the class F means the risk is false, that is, patient can survive after the one year critical period. From then total 470 patients' records, the class T (risk of not survive) contains 70 data instances; the other 400 instances are in class F (can survive the critical one-year period).

C. Discretizing Continuous Variables

It can be seen from Table II that there are three attributes (FVC, FEV1, Age) having numeric values. These three attributes have to be transformed to be categorical by means of discretization. To discretize continuous variables, we apply 3 discretization methods including hierarchical, uniform widths, and uniform counts. The states of FVC, FEV1, and Age after discretization with the 3 methods are summarized in Table III.

TABLE II: DETAILS AND MEANING OF DATA ATTRIBUTES

Attribute name	Meaning	Values
Diagnosis	Diagnosis codes for primary tumor, secondary tumor, or multiple tumors	{DGN1, DGN2, DGN3, DGN4, DGN5, DGN6, DGN8}
FVC	Forced vital capacity	Numeric
FEV1	Exhalation volume at the end of the first second of forced expiration	Numeric
Performance_status	Patients' performance based on Zubrod scale	{PRZ0, PRZ1, PRZ2}
Pain_before_surgery	Does the patient feel pain before surgery?	{T, F}
Haemoptysis_before_surgery	Is there haemoptysis before the surgery?	{T, F}
Dyspnoea_before_surgery	Is there dyspnoea before the surgery?	{T, F}
Cough_before_surgery	Does patient cough before the surgery?	{T, F}
Weakness_before_surgery	Does patient show weakness before surgery?	{T, F}
T_in_clinical_TNM	Size of the original tumor	{OC11, OC12, OC13, OC14}
Type_2_DM	Does patient have type 2 of diabetes mellitus?	{T, F}
MI_up_to_6_months	Does patient have myocardial infarction (heart attack) within 6 months?	{T, F}
PAD	Does patient have peripheral arterial diseases?	{T, F}
Smoking	Does patient smoke?	{T, F}
Asthma	Does patient have asthma symptom?	{T, F}
Age	Age of the patient	Numeric
One_year_survival_period	Chance of one year survival period after surgery	{T, F}

TABLE III: STATE OF VARIABLES (FVC, FEV1, AGE) AFTER DISCRETIZATION WITH 3 DIFFERENT METHODS

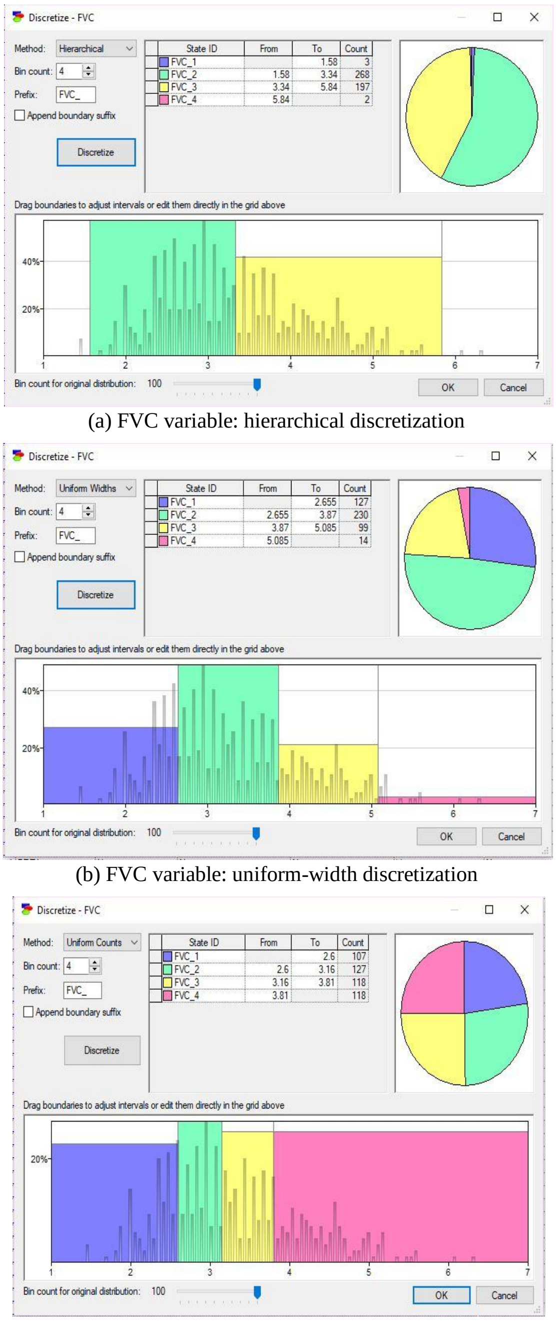
Variable	State	Discretization Method		
		Hierarchical	Uniform width	Uniform count
		Count	Count	Count
FVC	FVC_1	3	127	107
	FVC_2	268	230	127
	FVC_3	197	99	118
	FVC_4	2	14	118
FEV1	FEV1_1	455	456	67
	FEV1_2	1	0	67
	FEV1_3	1	0	67
	FEV1_4	2	0	61
	FEV1_5	6	2	73
	FEV1_6	4	7	67
	FEV1_7	1	5	68
AGE	AGE_1	1	1	57
	AGE_2	6	4	74
	AGE_3	17	19	70
	AGE_4	282	128	51
	AGE_5	110	191	76
	AGE_6	52	112	61
	AGE_7	2	15	81

From preliminary examination of appropriate ranges of intervals, we choose different number of intervals for different variables. FVC has been discretized to form four intervals. The variables FEV1 and AGE are discretized into seven intervals.

IV. EXPERIMENTAL RESULTS

A. Results of Discretization

The distribution as well as the range of values for each discretization method to transform numeric FVC variable to be categorical one are shown in Fig. 3. The distributions of FEV1 are illustrated in Fig 4, and distributions of AGE discretized values are demonstrated in Fig. 5.


Discretize - FVC

Method: Hierarchical

Bin count: 4

Prefix: FVC_

□ Append boundary suffix

State ID	From	To	Count
FVC_1		1.58	3
FVC_2	1.58	3.34	268
FVC_3	3.34	5.84	197
FVC_4	5.84		2

Drag boundaries to adjust intervals or edit them directly in the grid above

Bin count for original distribution: 100

(a) FVC variable: hierarchical discretization

Discretize - FVC

Method: Uniform Widths

Bin count: 4

Prefix: FVC_

□ Append boundary suffix

State ID	From	To	Count
FVC_1		2.655	127
FVC_2	2.655	3.87	230
FVC_3	3.87	5.085	99
FVC_4	5.085		14

Drag boundaries to adjust intervals or edit them directly in the grid above

Bin count for original distribution: 100

(b) FVC variable: uniform-width discretization

Discretize - FVC

Method: Uniform Counts

Bin count: 4

Prefix: FVC_

□ Append boundary suffix

State ID	From	To	Count
FVC_1		2.6	107
FVC_2	2.6	3.16	127
FVC_3	3.16	3.81	118
FVC_4	3.81		118

Drag boundaries to adjust intervals or edit them directly in the grid above

Bin count for original distribution: 100

(c) FVC variable: uniform-count discretization

Fig. 3. Results of FVC variable discretized with (a) hierarchical, (b) uniform-width, and (c) uniform-count methods.

It can be noticed from the experimental results that in FEV1 discretization with uniform-width method, 456 cases of patients (out of 470 cases) are in the category of FEV1 in the range 0-13.1514. There exist none of the patients in the next three ranges: 13.1514-25.3429, 25.3429-37.5343, and

37.5343-49.7257. We thus expect that the uniform-width method should not be the best discretization method, and the results (presented in model evaluation section) confirm our hypothesis.

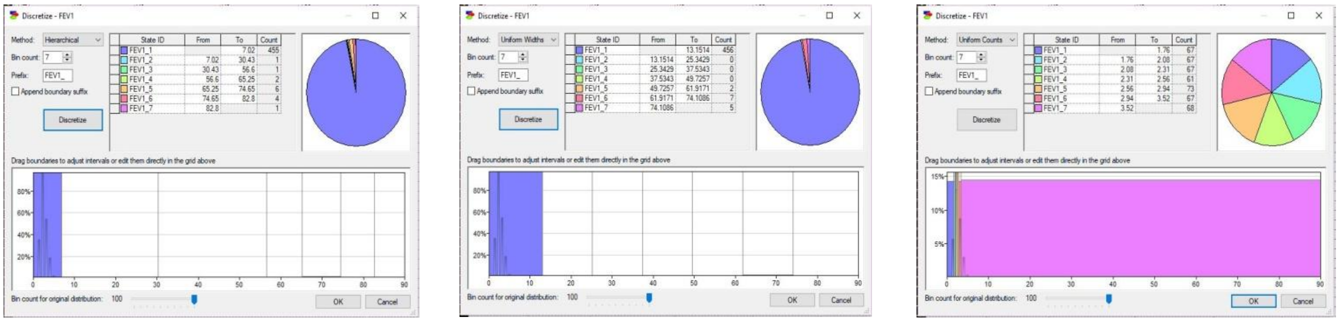
B. *Results of Bayesian Network Creation*

We build the BNs from the three kinds of discretization methods using the GeNIe Modeler academic version (available at https://www.bayesfusion.com). The models built from the hierarchical and uniform-width discretization methods are illustrated in Fig. 6. The model built from the uniform-count method is the most accurate one. We thus present separately the best BN model in Fig. 7. This Bayesian network model can capture the cause and effect relationships of the Thoracic Surgery based on the provided data attributes.

The built network model contains all 17 variables related to the post-operative life expectancy in the lung cancer patients. These predictive attributes are diagnosis, FVC, FEV1, performance status, pain before surgery, haemoptysis before surgery, dyspnoea before surgery, cough before surgery, weakness before surgery, T in clinical TNM (stage of malignant tumor), existence of type 2 Diabetes Mellitus, MI (heart attack) up to 6 months, PAD (peripheral arterial disease), smoking habit, asthma, age. The target of prediction is the chance for one year survival period.

In BN, arcs denote direct probabilistic relationships between pairs of nodes. If there exists the arc from node A to node B, the occurrence of event A can increase the probability of B's occurrence. Therefore, the arcs between pair of nodes in Fig. 7 can be summarized as follows:

- *Age* has impact on
  - *PAD* (peripheral arterial disease),
  - *MI_up_to_6_months* (myocardial infarction -- or heart attack -- within 6 months),
  - *T_in_clinical_TNM* (size of the original tumor), and
  - *Performance_status*.
- *FVC* (forced vital capacity) increases the likelihood of *FEV1* (exhalation volume of forced expiration).
- *FEV1* also mutually impacts *FVC*.
- *Performance_status* relates to
  - *Weakness_before_surgery*, and
  - *Cough_before_surgery*.
- *Pain_before_surgery* relates to *Diagnosis*.
- *Cough_before_surgery* relates to *Smoking*.



Discretize - FEV1

Method: Hierarchical
Bin count: 7
Prefix: FEV1_
□ Append boundary suffix

State ID	From	To	Count
FEV1_1		7.02	455
FEV1_2	7.02	30.43	1
FEV1_3	30.43	56.6	1
FEV1_4	56.6	65.25	2
FEV1_5	65.25	74.65	6
FEV1_6	74.65	82.8	4
FEV1_7	82.8		1

Drag boundaries to adjust intervals or edit them directly in the grid above

Bin count for original distribution: 100

Discretize - FEV1

Method: Uniform Widths
Bin count: 7
Prefix: FEV1_
□ Append boundary suffix

State ID	From	To	Count
FEV1_1		13.1514	456
FEV1_2	13.1514	25.3429	0
FEV1_3	25.3429	37.5343	0
FEV1_4	37.5343	49.7257	0
FEV1_5	49.7257	61.9171	2
FEV1_6	61.9171	74.1086	7
FEV1_7	74.1086		5

Drag boundaries to adjust intervals or edit them directly in the grid above

Bin count for original distribution: 100

Discretize - FEV1

Method: Uniform Counts
Bin count: 7
Prefix: FEV1_
□ Append boundary suffix

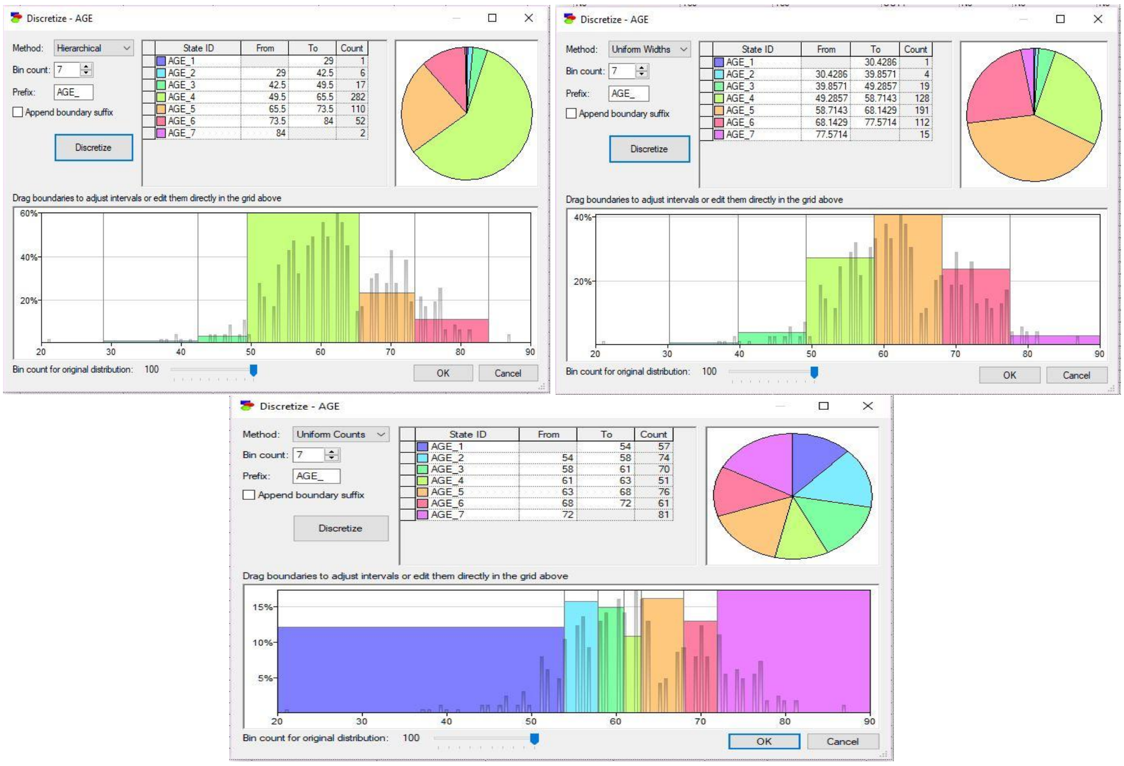
State ID	From	To	Count
FEV1_1		1.76	67
FEV1_2	1.76	2.08	67
FEV1_3	2.08	2.31	67
FEV1_4	2.31	2.56	61
FEV1_5	2.56	2.94	73
FEV1_6	2.94	3.52	67
FEV1_7	3.52		68

Drag boundaries to adjust intervals or edit them directly in the grid above

Bin count for original distribution: 100

(a) FEV1 variable: hierarchical discretization (b) FEV1 variable: uniform-width discretization (c) FEV1 variable: uniform-count discretization

Fig. 4. Result of FEV1 variable discretized with (a) hierarchical, (b) uniform-width, and (c) uniform-count methods.



Discretize - AGE

Method: Hierarchical
Bin count: 7
Prefix: AGE_
□ Append boundary suffix

State ID	From	To	Count
AGE_1		29	1
AGE_2	29	42.5	6
AGE_3	42.5	49.5	17
AGE_4	49.5	65.5	282
AGE_5	65.5	73.5	110
AGE_6	73.5	84	52
AGE_7	84		2

Drag boundaries to adjust intervals or edit them directly in the grid above

Bin count for original distribution: 100

Discretize - AGE

Method: Uniform Widths
Bin count: 7
Prefix: AGE_
□ Append boundary suffix

State ID	From	To	Count
AGE_1		30.4286	1
AGE_2	30.4286	39.8571	4
AGE_3	39.8571	49.2857	19
AGE_4	49.2857	58.7143	128
AGE_5	58.7143	68.1429	191
AGE_6	68.1429	77.5714	112
AGE_7	77.5714		15

Drag boundaries to adjust intervals or edit them directly in the grid above

Bin count for original distribution: 100

Discretize - AGE

Method: Uniform Counts
Bin count: 7
Prefix: AGE_
□ Append boundary suffix

State ID	From	To	Count
AGE_1		54	57
AGE_2	54	58	74
AGE_3	58	61	70
AGE_4	61	63	51
AGE_5	63	68	76
AGE_6	68	72	61
AGE_7	72		81

Drag boundaries to adjust intervals or edit them directly in the grid above

Bin count for original distribution: 100

Fig. 5. Discretization statistics over AGE variable using 3 methods: hierarchical (top-left), uniform-width (top-right), and uniform-count (bottom).

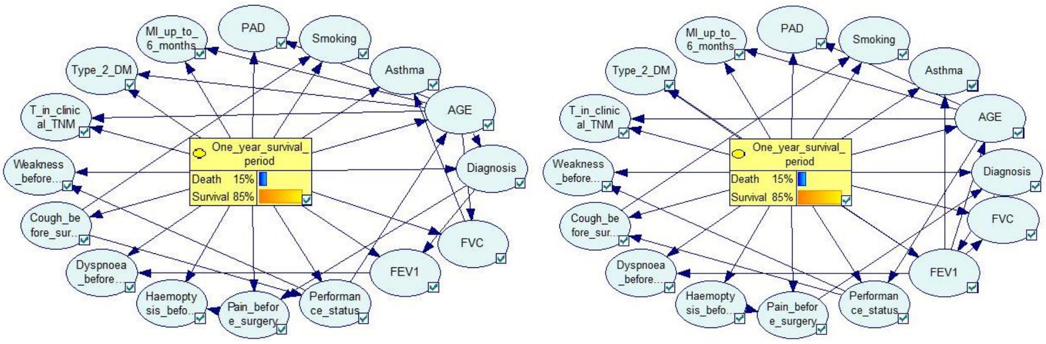


Fig. 6. Bayesian networks built from the hierarchical (top) and uniform-width (bottom) discretization methods.

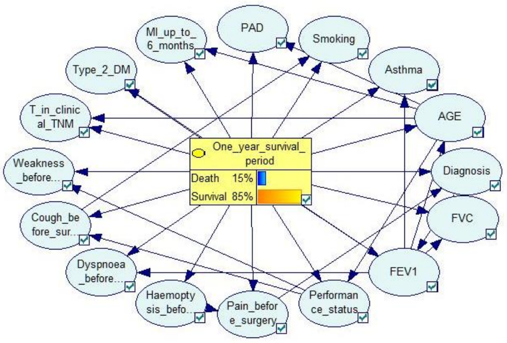


Fig.7. Bayesian network based on uniform-count discretization method capturing Thoracic Surgery Model.

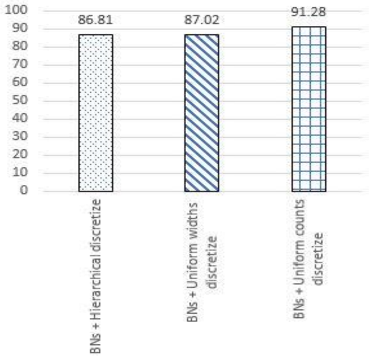


Fig. 8. Comparative chart showing the overall accuracy (in percentage) of BN + each discretization technique.

C. *Results of Model's Performance Evaluation*

We used overall accuracy for evaluating performance of the model to assess correctness on estimating the risk of patient's one year survival period. We compare the performance of BNs + uniform counts discretization to other two combinations of BN and discretization techniques that are known to be high accurate. The comparative results are summarized in Table IV, and graphically compared using a chart in Fig. 8.

From Table IV, it can be seen that the BNs + Uniform counts discretization can predict one year survival of patients more accurate than the BNs + Hierarchical discretization and the BNs + Uniform widths discretization techniques. For negative class (400 patients who did not survive after surgery), the three techniques can classify the negative class with similar performance. But for positive class (70 patients who can survive at least one year after the surgery), BNs + Uniform counts discretization model performs better than the BNs + Hierarchical discretization model and BNs + Uniform widths discretization model.

We also show ROC curves (Fig. 9) representing accuracy on classifying positive class (risk of death within one year) and negative class (survival the one-year critical period). Both classes have the same area under the ROC curve = 0.915839. This means the BN model can predict the positive class as accurate as the negative class.

TABLE IV: COMPARATIVE RESULTS FOR EACH CLASSIFICATION TECHNIQUE

Classification Techniques	Accuracy	Positive class	Negative class
BNs + Hierarchical discretize	86.81%	18/70	390/400
BNs + Uniform widths discretize	87.02%	17/70	392/400
BNs + Uniform counts discretize	91.28%	35/70	394/400

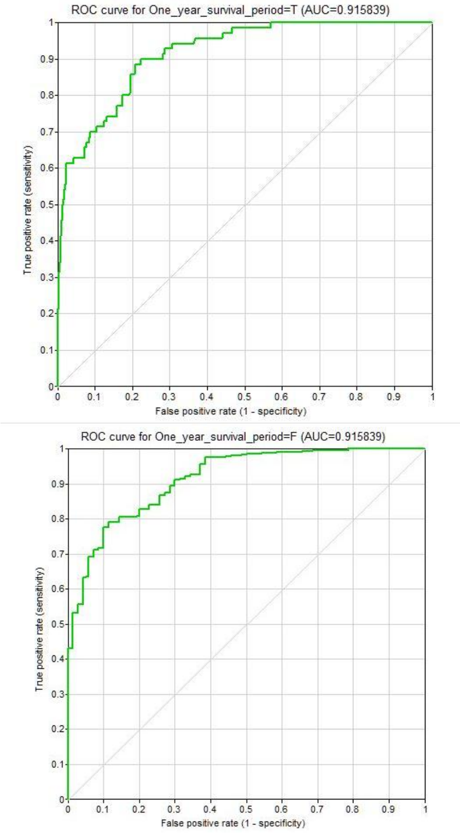


Fig. 9. ROC curves for risk of not survive the one-year period = **T** (top) and risky period = **F** (bottom) with the best technique: BNs + Uniform-count discretization.

V. CONCLUSION

We present in this work the high performance of data classification using Bayesian network (BN) modeling. The advantage of BN is its capability of modeling cause and effect relationships. We thus apply the BN modeling technique to study relationships of factors affecting one year survival chance of lung cancer patients after the surgery.

On applying BN technique, all the predicting factors have to take categorical values. We, therefore, use three kinds of discretization methods including hierarchical, uniform widths, and uniform counts. After discretization such that all data are categorical, we then create the BN model using the tree augmented naive-Bayes (TAN) learning algorithm. This TAN algorithm is available in the GeNIe Modeler software that we use to create the BN model.

We then compare the performance of BN models built from different discretization techniques. The experimental results show that one such suitable discretization technique appropriate for BN modeling is the uniform counts method. However, accuracy of the model depends on the number of discretized intervals and the number of data used for learning. Form our repeated experiments, we observe that uniform counts discretization technique shows constantly good performance on classifying one year survival period.

The repetitions with uniform counts discretization technique with different number of discretized intervals show the more number of intervals, the higher the predictive performance. Conversely, for some variables the more number of intervals, the lower on predictive performance. This observation, however, needs further investigation on the discretization characteristics.

REFERENCES

[1] J. Han, J. Pei, and M. Kamber*Data Mining: Concepts and Techniques*, Elsevier, 2011.

[2] J. Cheng and R. Greiner, “Comparing Bayesian network classifiers,” in *Proc. the Fifteenth Conference on Uncertainty in Artificial Intelligence*, Morgan Kaufmann, July 1999, pp. 101-108.

[3] N. Friedman, D. Geiger, and M. Goldszmidt, “Bayesian network classifiers,” *Machine Learning*, vol. 29, no. 2-3, pp. 131-163, 1997.

[4] H. Wasyluk, A. Onisko, and M. J. Druzdzel, “Support of diagnosis of liver disorders based on a causal Bayesian network model,” *Medical Science Monitor*, vol. 7, no. 1, pp. 327-332, 2001.

[5] A. Łupińska-Dubicka and M. J. Druzdzel, “A comparison of popular fertility awareness methods to a DBN model of the woman's monthly cycle,” in *Proc. the Sixth European Workshop on Probabilistic Graphical Models*, 2012, pp. 219-226.

[6] R. M. Austin, A. Onisko, and M. J. Druzdzel, “The Pittsburgh cervical cancer screening model: A risk assessment tool,” *Archives of Pathology and Laboratory Medicine*, vol. 134, no. 5, 744-750 2010.

[7] N. A. Loghmanpour, M. K. Kanwar, M. J. Druzdzel, R. L. Benza, S.vMurali, and J. F. Antaki, “A new Bayesian network-based risk stratification model for prediction of short-term and long-term LVAD mortality,” *ASAIO Journal*, vol. 61, no. 3, p. 313, 2015.

[8] A. C. Constantinou, N. Fenton, W. Marsh, and L. Radlinski, 2016, “From complex questionnaire and interviewing data to intelligent Bayesian network models for medical decision support,” *Artificial Intelligence in Medicine*, vol. 67, pp. 75-93.

[9] J. Pearl, *Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference*, Morgan Kaufmann, 1988.

[10] D. M. Powers, “Evaluation: From precision, recall and F-measure to ROC, informedness, markedness and correlation,” *Journal of Machine Learning Technologies*, vol. 2, no. 1, pp. 37-63, 2011.

[11] M. H. Zweig and G. Campbell, “Receiver-operating characteristic (ROC) plots: a fundamental evaluation tool in clinical medicine,” *Clinical Chemistry*, vol. 39, no. 4, pp. 561-577, 1993.

[12] UCI Dataset. (March 2017). *Thoracic Surgery Data Set*. [Online]. Available: https://archive.ics.uci.edu/ml/datasets/Thoracic+Surgery+Data

Kittipat Sriwong is currently a bachelor student with the School of Computer Engineering, Institute of Engineering, Suranaree University of Technology, Thailand. He has been fully financial supported by grant from Suranaree University of Technology throughout his bachelor study. He is a research assistant in the Knowledge Engineering Research Unit. His current research of interest includes data mining, support vector machine and Bayesian techniques, statistical data mining, and data mining applications in medical science.

Kittisak Kerdprasop is an associate professor at the School of Computer Engineering and Chair of the School. He is also the head of Knowledge Engineering Research Unit, Suranaree University of Technology, Thailand. He received his bachelor degree in Mathematics from Srinakarinwirot University, Thailand, in 1986, master degree in computer science from the Prince of Songkla University, Thailand, in 1991 and doctoral degree in computer science from Nova Southeastern University, U.S.A., in 1999. His current research includes machine learning, artificial intelligence and probabilistic knowledge bases.

Nittaya Kerdprasop is an associate professor and the head of Data Engineering Research Unit, School of Computer Engineering, Suranaree Uiversity of Technology. She received her B.S. in radiation techniques from Mahidol University, Thailand, in 1985, M.S. in computer science from the Prince of Songkla University, Thailand, in 1991 and Ph.D. in computer science from Nova Southeastern University, U.S.A., in 1999. Her research of interest includes knowledge discovery in databases, data mining, and artificial intelligence.