Computers in Biology and Medicine 47 (2014) 147–160

Contents lists available at ScienceDirect

Computers in Biology and Medicine

journal homepage: www.elsevier.com/locate/cbm

Modeling and predicting the occurrence of brain metastasis from lung cancer by Bayesian network: A case study of Taiwan

Kung-Jeng Wang [a,*], Bunjira Makond [a,b], Kung-Min Wang [c]

[a] Department of Industrial Management, National Taiwan University of Science and Technology, Taipei 106, Taiwan, ROC
[b] Faculty of Commerce and Management, Prince of Songkla University, Trang, Thailand
[c] Department of Surgery, Shin-Kong Wu Ho-Su Memorial Hospital, Taipei, Taiwan, ROC

ARTICLE INFO

Article history:
Received 4 October 2013
Accepted 5 February 2014

Keywords:
Bayesian network
Brain metastasis
Lung cancer

ABSTRACT

The Bayesian network (BN) is a promising method for modeling cancer metastasis under uncertainty. BN is graphically represented using bioinformatics variables and can be used to support an informative medical decision/observation by using probabilistic reasoning. In this study, we propose such a BN to describe and predict the occurrence of brain metastasis from lung cancer. A nationwide database containing more than 50,000 cases of cancer patients from 1996 to 2010 in Taiwan was used in this study. The BN topology for studying brain metastasis from lung cancer was rigorously examined by domain experts/doctors. We used three statistical measures, namely, the accuracy, sensitivity, and specificity, to evaluate the performances of the proposed BN model and to compare it with three competitive approaches, namely, naive Bayes (NB), logistic regression (LR) and support vector machine (SVM). Experimental results show that no significant differences are observed in accuracy or specificity among the four models, while the proposed BN outperforms the others in terms of sampled average sensitivity. Moreover the proposed BN has advantages compared with the other approaches in interpreting how brain metastasis develops from lung cancer. It is shown to be easily understood by physicians, to be efficient in modeling non-linear situations, capable of solving stochastic medical problems, and handling situations wherein information are missing in the context of the occurrence of brain metastasis from lung cancer.

1. Introduction

Twenty to forty percent of cancer patients develop brain metastases during their illness [39,38]. Lung cancer is a leading cause of death worldwide and often spreads to the brain given that 65% of patients diagnosed with a primary tumor in their lungs will have brain metastases [7]. The survival time of patients is longer when lung cancer is detected early because it still localized and can be effectively treated. By contrast, survival time decreases and quality of life deteriorates once lung cancer metastasizes to the brain [13]. Modeling and predicting the development of brain metastasis from lung cancer become necessary in the early detection of brain metastasis. A model with good predictive ability can help physicians distinguish between lung cancer patients who will most likely develop brain metastasis, and those who will only suffer from lung cancer. Such information will help physicians decide on the most suitable treatment for lung cancer patients as well as determine appropriate management treatments to reduce or prevent brain metastasis. Therefore, the clinical outcome for these patients can be improved [14].

Previous studies have proposed a number of models to predict cancer outcomes. For example, Bajard et al. [3] used multivariate analysis to predict factors of brain metastases in a group of patients with stages I to III non-small cell lung cancer (NSCLC). The variables include age at the time of diagnosis, gender, performance status, weight-loss, stage, T-status, N-status, histological type, type of treatment, administration of chemotherapy, use of cisplatin, and response to initial treatment. A statistical method and conditional probability analysis were performed to analyze the metastatic patterns of lung cancer cases. Patient characteristics included in this study were age, gender, histology, lung cancer, and metastatic sites [37]. Hierarchical logistic regression (LR) was used to determine the predicted probability of metastatic disease to the brain as a function of age, sex, tumor size, cell type, locations of the tumors, and lymph node stage of the primary NSCLC patients [32]. Although traditional statistical and machine learning models, such as LR and support vector machine (SVM), are popularly used for cancer prediction,

* Corresponding author. Tel.: +886 2 27376769.
E-mail addresses: kjwang@mail.ntust.edu.tw (K.-J. Wang), bunjira_m@hotmail.com (B. Makond), albert.hua@msa.hinet.net (K.-M. Wang).

http://dx.doi.org/10.1016/j.compbiomed.2014.02.002
0010-4825

these models are not as promising as the Bayesian network (BN) given that the BN can use reasoning under uncertainty whereas both LR and SVM cannot.

The BN is a powerful tool for representing stochastic events and conducting prediction tasks. Lowd and Domingos [27] proposed a naive Bayes (NB) as an alternative model for the BN for general probability estimation task, the results showed that both NB and BN have the same computational time and accuracy performance; however, NB is unable to apply in relation domain because of its independency assumption while BN is widely used. Zhang et al. [49] concluded in their study that SVM and BN are the best two algorithms for predicting overweight and obesity from the Wirral database; however, [10] and Jayasurya et al. [20] concluded that BN is better to handle missing data than SVM; therefore, BN is more suitable for the medical domain. Oh et al. [36] stated that this tool can approximate complex multivariable probability distributions of heterogeneous variables as interpretable local probabilities to incorporate prior clinical and biological knowledge as well as to visualize and interpret the interactions among variables of interest for clinical use. Mancini et al. [29] stated that the traditional statistical methods are ineffective to describe the relationship between variables in biomedical domain because of their limitation of independency whereas the BN can overcome this limitation and become a popular method for analyzing in biomedical data. Lalande et al. [24] built a BN to identify older adults with high risk of falls; this tool would integrate numerous risk factors based on literature data, in order to obtain a fall risk assessment, giving robust results whatever the settings; in addition, the BN is interesting model because it concerns knowledge from experts and also knowledge contained in data. The BN can also be used as a classifier based on a learned network structure. As a result, each node can compute for the posterior probability distribution, which is useful for decision-makers.

Given the attractive characteristics of the BN model, researchers have used it in various medical problems. For example, BN is used in mammographic diagnosis for breast cancer [21]. Hoot and Aronsky [17] created a BN model that included 29 variables for predicting a 90-day graft survival. The predictive performance measured by area under the receiver operation characteristic curve was 0.674. Morales et al. [31] applied Bayesian classifiers to estimate the implantation probability of embryos in artificial insemination treatments from embryo images and to predict the successfully suitability implantation of an embryo chosen for being transferred. In perspective of the receiver operating characteristic, they concluded that the tree augmented naive Bayes, *k*-dependence Bayesian, and naive Bayes classifiers performed almost as well as the semi naive Bayes and selective naive Bayes classifiers. Visscher et al. [46] utilized BN for the diagnosis and treatment of ventilator-associated pneumonia. Oh et al. [36] developed a BN model to predict local failure in lung cancer. Corani et al. [9] presented a BN for predicting the outcome of in vitro fertilization (IVF); they concluded that BN is equally or more predictive than well-recognized classification algorithms with the further advantage of being biologically interpretable.

The remainder of this paper is organized as follows: in Section 2, we briefly review BN along with other methods used in this study. Variables, data, graphical model construction, including evaluation criteria, are described in Section 3. The experimental results are presented in Section 4. We provide the discussion and conclusions in Section 5.

2. Materials and methods

This section we thoroughly describe the proposed BN. In addition, the benchmark models: NB, LR, SVM are discussed as well as re-sampling techniques and model evaluation indexes used in this study.

2.1. Bayesian network

The BN is a probability graphical model that encodes a joint probability distribution over a set of random variables $\boldsymbol{X} = \{X_1, \ldots, X_n\}$, and these variables can be either discrete or continuous. Formally, a BN forms a directed-acyclic graph (DAG) by using a set of nodes representing the variables and a set of directed edges representing the relationships between the variables [36]. Each variable X_i is independent of its non-descendants given its parents in the graph. The joint probability distribution over $\boldsymbol{X}$ is given by $P(X_1, \ldots, X_n) = \prod_{i=1}^{n} P(X_i|P_a(X_i))$, where $P_a(X_i)$ is the set of parents of X_i.

2.1.1. Structure and parameter learning

The three-step construction of a BN is as follows [44,26]: (i) indentify the set of relevant variables and their possible values, as well as details for variable identification. (ii) Find the network structure by connecting nodes that represent variables with arcs into a DAG, and construct the graph by using expert knowledge or by applying the algorithm obtained from the data. (iii) Define the conditional probability table (CPT) for each node in the graph.

In this study, the network structure was rigorously examined by domain experts/doctors and constructed using data from related medical literature. After the structure of a BN was known, we quantified the relationship between connected nodes by CPT for discrete variables. We used the maximum likelihood (ML) method to obtain the values for CPT.

The ML estimators were obtained as follows. Given that D represents a set of random variables $\{X_1, X_2, \ldots, X_n\}$ where X_i is an element of the BN variables, the goal of parameter learning is to find the most probable values for vector θ, which can be quantified by the log likelihood function $L_D(\theta) = \log(p(D|\theta))$. Assuming that the samples are independently selected from the underlying distribution, we need to maximize $L_D(\theta) = \log \prod_{ijk} \theta_{ijk}^{n_{ijk}}$, where n_{ijk} indicates the number of elements of D containing both x_i^k and pa_i^j, where x_i^k is a value or state of X_i and pa_i^j is a set of states of parents of X_i. Maximum likelihood estimation has its optimum at $\theta_{ijk} = \frac{n_{ijk}}{\sum_k n_{ijk}}$ [50].

2.1.2. Inference in Bayesian network

Inference in a BN is an important task, which the posterior probability of the query variables is computed given a set of evidence variables as the knowledge to the network. Two algorithms are used for inference in a BN, namely, exact and approximate. The most popular exact inference algorithm converts a BN into a junction tree and then performs an exact inference on the junction tree. The computation of exact inference on a junction tree consists of three steps [48]: (i) in local evidence absorption, the clique potential tables that include evidence variables are updated. (ii) In evidence propagation, the evidence is propagated throughout the entire junction tree. And (iii) in the query, in this step the probability distribution of the query variables is computed from the potential table of the cliques that include query variables. Such inference algorithm can be implemented in Bayes Net of Matlab Toolbox [33] and was used in our experiments.

The advantages of using a BN can be stated as follows: (1) its ability to represent the causal relationships among variables in the visual form of graph, which makes the dependence and independence as well as interaction relationships of specifics variables easy to recognize and interpret [8,10], (2) While SVM, LR, and NB are linear classifiers [43], Bayesian network is efficient in modeling both linear and non-linear situations [16,45,35,4], (3) its ability

to solve stochastic decision-making problems by finding the probability distribution of children nodes given the evidence values of their parents and vice versa [16,45,28], and (4) its ability to handle situations wherein data/information are missing.

2.2. Benchmark methods

2.2.1. Naive Bayes

The naive Bayes is the simplest BN classifier, in which each variable node X_i has the class node C_i , but does not have any other parent. All variables node X_i are mutually independent given the class variable. This classifier learns the conditional probability of each variable X_i given the label C_i in the training data. Classification is then done by applying Bayes rule to compute the probability of C given the particular instantiation of $X_i, \ldots, X_n$. Surprisingly, the performance of a NB is somewhat competitive given that this is clearly an unrealistic assumption. More details for the NB can refer to Lowd and Domingos [27] and Su and Zhang [43].

2.2.2. Logistic regression

The LR is well known as the *gold standard* method in prediction tasks [2]. It is a statistical method that describes the relation between predictor variables denoted by $\boldsymbol{x}' = (x_1, x_2, \ldots, x_p)$ and a response variable, which is a categorical variable with two values [18], that is, the "occurrence of brain metastasis" or "nonoccurrence of brain metastasis" in this study.

The conditional probability of the occurrence of brain metastasis for lung cancer patients can be written as $P(Y=1|\boldsymbol{x}) = \pi(\boldsymbol{x})$; then, the LR model for p predictor variables can be written as

$$\pi(\boldsymbol{x}) = \frac{e^{(\beta_0+\beta_1 x_1+\beta_2 x_2+\ldots+\beta_p x_p)}}{1+e^{(\beta_0+\beta_1 x_1+\beta_2 x_2+\ldots+\beta_p x_p)}} \quad (1)$$

where $0 \le \pi(\boldsymbol{x}) \le 1$.

A useful transformation of a LR is the logit transformation, defined as

$$g(\boldsymbol{x}) = \ln\left(\frac{\pi(\boldsymbol{x})}{1-\pi(\boldsymbol{x})}\right) = \beta_0+\beta_1 x_1+\beta_2 x_2+\cdots+\beta_p x_p \quad (2)$$

The odds ratio (OR) associated with one unit change in x_j is represented with $e^{(\beta_j)}$.

2.2.3. Support vector machine

The SVM is a widely recognized machine learning method used for classification tasks. This method is performed by finding the hyperplane that classifies the space of all possible instances into two classes and maximally separates these two classes. The SVM differs from other machine learning methods and maps inseparable data into a higher dimensional space where these data can be linearly separated [15].

2.3. Re-sampling techniques

In this study, we used legal records of 36,043 lung cancer patients from the National Health Insurance of Taiwan (NHI) database that is significantly imbalanced with 99% in majority and 1% of minority. Given the imbalanced data set of this study, the classification model can be made ineffective by the instances in majority class, which results in high accuracy for the majority class but poor accuracy for the minority class. To overcome such problem, the imbalanced data set can be processed using re-sampling techniques, such as under-sampling and oversampling. Random under-sampling is a simple technique for balancing class distribution. The instances in the majority class in the training set are randomly eliminated until the ratio between the number of instances in the minority and majority classes is at the target level. Although the random under-sampling technique is simple, it is empirically proven as one of the most effective re-sampling techniques. In particular, few of the more sophisticated under-sampling methods have outperformed random under-sampling in empirical studies [25]. Furthermore, random oversampling is a simple technique that balances class distribution by randomly replicating minority class instances. Similar to random under-sampling, random oversampling is a simple yet effective re-sampling approach [25].

3. Material and modeling

3.1. Variables

Epidemiological studies have reported that age, gender, and residence are risk factors that may increase a person's chance to develop lung cancer [30,22,23,42,11]. Given that lung cancer metastasis occurs at the time of diagnosis or after undergoing treatment, treatment was also used as a factor for occurrence of brain metastasis [3,19]. Accordingly, in the present study, six variables were used to construct the proposed BN model: (1) age; (2) gender; (3) region of residence, environment of patients; (4) location of lung cancer within human body; (5) treatment (primary treatment described in the medical database of patients); and (6) occurrence, which represents the development of brain metastasis from lung cancer. This occurrence is designated as "yes" if the second or latter diagnosis has brain metastasis; otherwise, it is designated as "no." Occurrence likewise functions as a response variable. Moreover, the possible values of these variables are presented in Table 1.

3.2. Data

This study used the NHI database from 1996 to 2010 collected by the Bureau of NHI, Taiwan (BNHI). We obtained the permission of the Institutional Review Board (IRB) to use this database. The BNHI collects comprehensive data of cancer patients each year. The two groups of files are registration and original claim data for reimbursement [34]. In this study, we retrieved data from the "Ambulatory care expenditures by visits file" (CD file), which is in the "Original claim data for reimbursement" file.

This study focuses on patients primarily diagnosed with lung cancer from 1996 to 2010. The cancers patients selected using the International Classification of Diseases, Ninth Revision, Clinical Modification (ICD-9-CM) were as follows: code 162 for lung cancer patients and code 1983 for cancer patients with brain metastasis. For precision, this study excluded the data of lung cancer patients who were recorded with A-codes because the information regarding the site of lung cancer is incomplete in the database.

We obtained the legal records of 36,043 lung cancer patients from the database. As shown in Table 1, 35,605 patients were diagnosed with only lung cancer and only 438 patients developed brain metastasis. Sixty-five percent of these patients were men and majority of these patients were older than 70 years. Lung cancer rarely occurs (12%) in people who are younger than 50 years old. The data showed that half of all patients were from northern Taiwan and 3% of patients were from eastern Taiwan. Most patients received drug medication after they were diagnosed with lung cancer. Seventy-five percent of lung cancer patients were diagnosed with site in bronchus and lungs, or unspecified site (1629). A total 438 patients (1%) had lung cancer that metastasized to the brain, and the rest of the patients (99%) only had lung cancer. Notably, the data set was highly imbalanced.

Given the imbalanced data set, a classification model can be made ineffective by the instances in the majority class, which results in high accuracy for the majority class but poor accuracy for the minority class. To deal with this problem, we used both

Table 1
Data profile of lung cancer patients.

Characteristics	Number of patients	(%)
Gender		
Female (F)	12,473	35
Male (M)	23,570	65
Age		
Less than 50 yr (<50)	4396	12
5–0–60 yr (50–60)	6184	17
60–70 yr (60–70)	9340	26
More than 70 yr (>70)	16,123	45
Region (in Taiwan)		
Central branch (C)	6687	19
Eastern branch (E)	1250	3
Northern branch (N)	18,174	50
Southern branch (S)	9932	28
Site of lung cancer		
Trachea, bronchus and lung (162)	3404	9
Trachea (1620)	1344	4
Bronchus (1622)	262	1
Upper lobe, bronchus or lung (1623)	472	4
Middle lobe, bronchus or lung (1624)	243	1
Lower lobe, bronchus or lung (1625)	872	2
Other parts of bronchus or lung (1628)	1449	4
Bronchus and lung, unspecified (1629)	26,997	75
Treatment		
Radiotherapy (Ra)	2045	6
Chemotherapy (Ch)	1201	3
Drugs (Dr)	32,797	91
Occurrence (brain metastasis)		
Yes (Y)	438	1
No (N)	35,605	99

random under-sampling and random over-sampling method to adjust our data distribution. Random under-sampling was applied to eliminate instances in the majority class until the size of the reminding data was decreased to 2365 instances. Given that many duplicated instances exist in the majority class, applying random under-sampling to eliminate instances did not lead to significant information loss. In the minority class, random over-sampling was utilized to select instances from the class randomly, and this process was then repeated until the number of instances was increased up to 2190. The new training set then comprised 4555 (2365 instances in the majority class and 2190 from the minority class) and was almost balanced.

3.3. Graphical model construction

The graphical BN model with six variables was constructed and is shown in Fig. 1. The network topology is derived on the basis of correlation analysis of variables and evidences from literatures, and then reviewed by domain experts. The correlation analysis in Table A1 of Appendix A indicates some variables do significantly correlate with others. For instance, gender is correlated with site of lung cancer and treatment. It is also confirmed by Ko et al. [23], Samet et al. [42], and de Groot and Munden [11] that gender is a determinant of lung cancer. In addition, WHO Department of Gender, Women and Health [47] stated that gender is a cause of treatment.

Further, age is confirmed by Brown et al. [5] to be a determinant of treatment. Although age is not correlated with site of lung cancer in our correlation analysis, Samet et al. [42,11] have justified that age is a major determinant of lung cancer; hence we decide to build the link from age to site.

The correlation analysis shows that region is correlated with the site of lung cancer, treatment, and the occurrence of brain metastasis. Such finding is aligned with the study of American Cancer Society. [1],

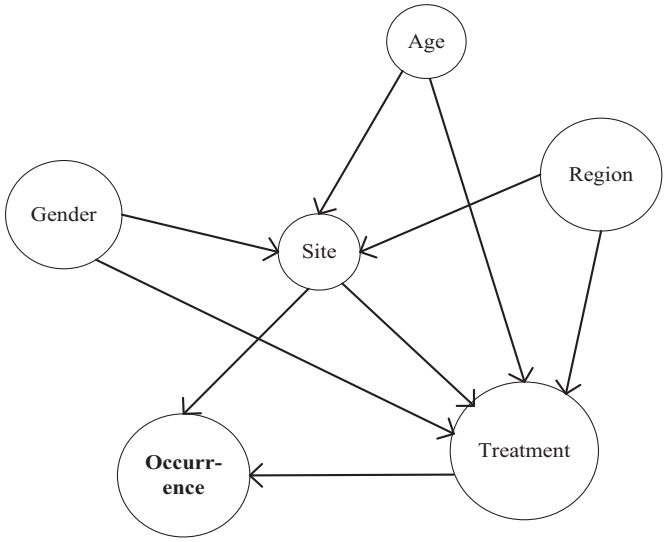


Fig. 1. The BN topology of brain metastasis occurrence from lung cancer.

Samet et al. [42,11] which claimed that region is a determinant of site of lung cancer. In addition, Cartman et al. [6] and Forrest et al. [12] confirmed that region affects treatment. The correlation analysis results show that region is also related to the occurrence of brain metastasis. However, because we already have had the relation from treatment to the occurrence of brain metastasis, the direct link between region and the occurrence of brain metastasis is removed to reduce the complexity of the model. Moreover, the treatment affecting on the occurrence is confirmed by Hubbs et al. [19].

The correlation results also show that site of lung cancer is correlated to the treatment and the occurrence of brain metastasis

which is in coincidence with the study of Bajard et al. [3] and American Cancer Society [1]. concluding that the occurrence of brain metastasis is associated with the site of lung cancer. The resulting BN topology for studying brain metastasis from lung cancer was rigorously examined by domain experts/doctors.

3.4. Model evaluation criteria

To evaluate the prediction performance of brain metastasis, we used 80% of the data for the training set and 20% as a testing set. The performance of the model was determined from three indexes, namely, accuracy index, which calculates if the overall prediction is correct, as well as sensitivity and specificity indexes, which measure the positive and negative predictive performances, respectively.

4. Experiments and results

In this section, we discuss our experimental results in two parts: explanatory graphical model and inference and predictive performance.

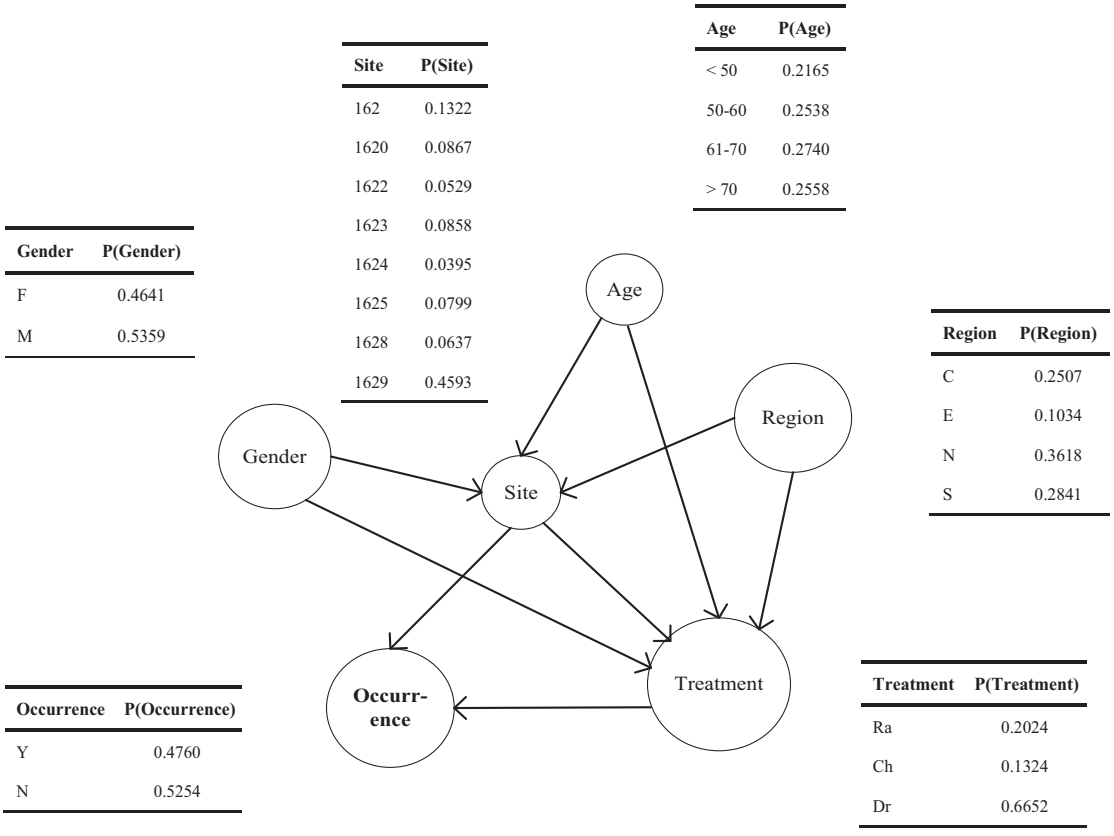
4.1. Explanatory graphical model and inference

The data are used to calculate the conditional probabilities which explain the changes of each variable affected by the evidences would provide the useful information for determining the possibility of brain metastasis. Fig. 2 shows the graphical model with six variables together with the prior probability of triggering nodes.

Table 2 shows the resulting conditional probability for **Site** prediction as well as the findings from the changes in lung cancer site after the presentation of gender, age, and region. Examples are as follows. (1) Lung cancer in the trachea, bronchus and lung (162) mostly occurs in male patients aged of 50 to 60 and living in the east, which accounted for 27.87%. (2) Lung cancer in the trachea (1620) mostly occurs in male patients over 70 years old and living in the east, which accounted for 16.44%. (3) Lung cancer in the bronchus (1622) mostly occurs in female patients over 70 years old and living in the east, which accounted for 13.51%. (4) Lung cancer in the upper lobe, bronchus, and lung (1623) mostly occurs in male patients aged 50 to 60 and living in the south, which accounted for 19.23%. (5) Lung cancer in the middle lobe, bronchus, and lung (1624) mostly occurs in female patients aged of 50 to 60 and living in the center, which accounted for 10.42%. (6) Lung cancer in the lower lobe, bronchus and lung (1625) mostly occurs in male patients over 70 years old and living in the south, which accounted for 14%. (7) Lung cancer in other parts of the bronchus or lung (1628) mostly occurs in male patients aged of 61 to 70 and living in the center, which accounted for 24.35%. (8) Lung cancer in the bronchus and lung or an unspecified site (1629) mostly occurs in male patients aged of 50 to 60 and living in the east, which accounted for 24.35%.

The resulting conditional probabilities of treatment when presenting gender, age, region, and site of lung cancer are listed in Table B1 (Appendix B), which offers informative findings on **Treatment** prediction. For instance, patient who are male, aged of 50 to 60, living in central Taiwan, and have lung cancer in bronchus (1622), are highly likely to undergo radiotherapy. Female patients below 50 living in the south, and having cancer in the middle lobe, bronchus, and lung (1624) are very likely to undergo chemotherapy (83.33%). Likewise, the appearance of specific characteristics in patients results in the certainty of undergoing treatment with drugs. However, most patients with lung cancer in bronchus and lung, or an unspecified site (1629) rarely undergo drug treatment.

The resulting conditional probabilities of the **occurrence** of brain metastasis after presenting the site of lung cancer and treatment are listed in Table 3, which reveals several interesting



Gender	P(Gender)
F	0.4641
M	0.5359

Site	P(Site)
162	0.1322
1620	0.0867
1622	0.0529
1623	0.0858
1624	0.0395
1625	0.0799
1628	0.0637
1629	0.4593

Age	P(Age)
< 50	0.2165
50-60	0.2538
61-70	0.2740
> 70	0.2558

Region	P(Region)
C	0.2507
E	0.1034
N	0.3618
S	0.2841

Occurrence	P(Occurrence)
Y	0.4760
N	0.5254

Treatment	P(Treatment)
Ra	0.2024
Ch	0.1324
Dr	0.6652

Fig. 2. The BN topology of brain metastasis occurrence from lung cancer with the prior probability of each node; the detailed CPTs are presented in Appendix.

Table 2
Resulting conditional probability for **Site** prediction.

G	*A*	*R*	*P*(*S*|*G*, *A*, *R*)							
			162	1620	1622	1623	1624	1625	1628	1629
F	< 50	C	0.1221	0.0916	0.0992	0.0992	0.0153	0.0992	0.0687	0.4046
M	< 50	C	0.1791	0.0373	0.0373	0.0522	0.0597	0.0821	0.1269	0.4254
F	50–60	C	0.1736	0.0764	0.0278	0.0208	0.1042	0.0903	0.1875	0.3194
M	50–60	C	0.1067	0.1533	0.0600	0.0867	0.0600	0.1067	0.0667	0.3600
F	61–70	C	0.0348	0.1304	0.0609	0.0609	0.0087	0.1043	0.2435	0.3565
M	61–70	C	0.1968	0.0745	0.1011	0.0479	0.0106	0.1170	0.0851	0.3670
F	> 70	C	0.2039	0.1456	0.0097	0.1068	0.0388	0.0388	0.1165	0.3398
M	> 70	C	0.2034	0.1356	0.0565	0.1469	0.0226	0.1073	0.0960	0.2316
F	< 50	E	0.2128	0.0851	0.0000	0.0000	0.0851	0.0000	0.1277	0.4894
M	< 50	E	0.0962	0.0962	0.0000	0.1154	0.0000	0.1346	0.0385	0.5192
F	50–60	E	0.1111	0.0000	0.0000	0.1111	0.0556	0.0556	0.0278	0.6389
M	50–60	E	0.2787	0.0984	0.0000	0.0984	0.0984	0.0820	0.0164	0.3279
F	61–70	E	0.1231	0.0308	0.0769	0.1385	0.0000	0.0615	0.0000	0.5692
M	61–70	E	0.0635	0.1111	0.0794	0.1111	0.0794	0.0000	0.0159	0.5397
F	> 70	E	0.1622	0.0405	0.1351	0.0405	0.0000	0.0270	0.0811	0.5135
M	> 70	E	0.2603	0.1644	0.0000	0.0685	0.0411	0.0548	0.0548	0.3562
F	< 50	N	0.1171	0.0829	0.0732	0.0537	0.0390	0.1024	0.0439	0.4878
M	< 50	N	0.1898	0.1095	0.0292	0.1095	0.0511	0.0438	0.0292	0.4380
F	50–60	N	0.1587	0.1058	0.0212	0.0741	0.0212	0.0794	0.0159	0.5238
M	50–60	N	0.1167	0.0625	0.0750	0.0417	0.0042	0.0792	0.0583	0.5625
F	61–70	N	0.0920	0.1166	0.0307	0.0368	0.0245	0.0920	0.0123	0.5951
M	61–70	N	0.1722	0.0769	0.0183	0.0513	0.0733	0.0733	0.0293	0.5055
F	> 70	N	0.0977	0.0837	0.0465	0.0698	0.0186	0.0837	0.0977	0.5023
M	> 70	N	0.1681	0.0575	0.0265	0.0752	0.0398	0.0575	0.0619	0.5133
F	< 50	S	0.0621	0.1034	0.0345	0.1793	0.0828	0.0828	0.0345	0.4207
M	< 50	S	0.0741	0.0222	0.0296	0.1630	0.0444	0.0963	0.0222	0.5481
F	50–60	S	0.0680	0.0194	0.0874	0.1456	0.0243	0.0922	0.0680	0.4951
M	50–60	S	0.0692	0.1308	0.0231	0.1923	0.0692	0.0308	0.0692	0.4154
F	61–70	S	0.0682	0.1023	0.1080	0.0511	0.0398	0.0568	0.0284	0.5455
M	61–70	S	0.1122	0.1220	0.0683	0.1073	0.0244	0.0585	0.0439	0.4634
F	> 70	S	0.1900	0.0800	0.0900	0.0900	0.0200	0.1400	0.0400	0.3500
M	> 70	S	0.0964	0.0457	0.0711	0.0863	0.0609	0.0964	0.0457	0.4975

Table 3
Resulting conditional probability for the prediction of the **occurrence** of brain metastasis.

S	*T*	*P*(*P*|*S*,*T*)	
		Y	N
162	Ra	0.1567	0.8433
1620	Ra	0.0247	0.9753
1622	Ra	0.0000	1.0000
1623	Ra	0.0208	0.9792
1624	Ra	0.0000	1.0000
1625	Ra	0.0345	0.9655
1628	Ra	0.1481	0.8519
1629	Ra	0.6333	0.3667
162	Ch	0.0482	0.9518
1620	Ch	0.0000	1.0000
1622	Ch	0.0000	1.0000
1623	Ch	0.0702	0.9298
1624	Ch	0.0000	1.0000
1625	Ch	0.0000	1.0000
1628	Ch	0.0000	1.0000
1629	Ch	0.1981	0.8019
162	Dr	0.6052	0.3948
1620	Dr	0.2715	0.7285
1622	Dr	0.0197	0.9803
1623	Dr	0.3361	0.6639
1624	Dr	0.0574	0.9426
1625	Dr	0.2805	0.7195
1628	Dr	0.3727	0.6273
1629	Dr	0.8797	0.1203

points. The chance of brain metastasis greatly increases when a patient has lung cancer in the bronchus and lung or at an unspecified site (1629) and undergoes drug treatment as the primary treatment. Conversely, any patient with lung cancer in the trachea (1620), bronchus (1622), middle lobe, bronchus, and lung (1624), lower lobe, bronchus, and lung (1625), or other parts of the bronchus or lung (1628) and undergoes radiotherapy or chemotherapy does not experience brain metastasis.

Moreover, when the proposed BN is used, we can compute for the posterior probabilities for any query/prediction given the appropriate evidence. The BN allows the model to find the probability despite incomplete evidence. For example, brain metastasis is difficult to diagnose. Thus, physicians can use the proposed BN to predict the likelihood of the patient experiencing brain metastasis given particular information regarding the patient. If the patient is a 55-year-old man diagnosed with lung cancer in the trachea (1620) and living in the eastern part of Taiwan (without any information concerning treatment), the probability of the occurrence of brain metastasis is 27.15% (the probability computation is expressed in Appendix C). Similarly, the probability of the non-occurrence of brain metastasis is the difference between 1 and the probability of the occurrence of brain metastasis is 72.85% (the probability computation is expressed in Appendix C).

The type of treatment is also informative and important for both physicians and patients. The proposed BN will help physicians anticipate the probability of the most likely treatment for patients when the physician lacks information regarding the patient's status. For example, for a female patient with lung cancer in the trachea, bronchus, and lung (162), the probability that the patient should undergo radiotherapy is 13.35% (the probability computation is expressed in Appendix C).

The resulting probabilities of the occurrence of brain metastasis and treatment are presented in Table 4(a) and (b), respectively. This information can help the physician decide on appropriate

Table 4
The posterior probability.

(a) The probability of the occurrence of brain metastasis given gender, age, region, and site

G	A	R	S	$P(O\|G, A, R, S)$	
				Y	N
M	50–60	E	1620	0.2715	0.7285

(b) The probability of treatment given gender and site

G	S	$P(T\|G,S)$		
		Ra	Ch	Dr
F	162	0.1335	0.1289	0.7376

Table 5
The posterior probabilities.

(a) The probability of the region of a patient given site, treatment, and occurrence

S	T	O	$P(R\|S,T,O)$			
			Center	East	North	South
162	Ch	Y	0.3440	0.1593	0.4338	0.0629

(b) The probability of the site of lung cancer of a patient given gender, age, and occurrence

G	A	O	$P(S\|G,A,O)$							
			162	1620	1622	1623	1624	1625	1628	1629
F	<50	Y	0.1319	0.0292	0.0018	0.0500	0.0025	0.0409	0.0359	0.7078

treatments for the patient according to the prediction of brain metastasis. The updated information can also help patients better understand their health conditions and prepare them physically and mentally for subsequent treatments.

The likelihood of the development of brain metastasis is important information for medical management. The proposed BN with predictive reasoning can help the healthcare policymakers gain a clear picture of the possible development of brain metastasis. For example, given a patient undergoing chemotherapy for lung cancer in the trachea, bronchus, and lung (162) and having brain metastasis, the probability of this illness occurring in a patient living in any part of Taiwan is expressed as $P(R|S=162,\ T=Ch, O=Y)$ and presented in Table 5(a). For example, the results show that the probability of this illness occurring in a patient living in central Taiwan is 34.40% (the probability computation is expressed in Appendix C).

Having lung cancer in different sites results in the different treatment types and costs. The proposed BN allows insurance policymakers to gauge quickly the probability of lung cancer in patients in specific demographics. In addition to the likelihood of lung cancer, the relationship between patients' demographic information and the occurrence of lung cancer is presented in Table 5(b). For instance, the probability of a patient with lung cancer in the middle lobe, bronchus, or lung (1624) given female patient aged lower 50 years old and having brain metastasis is 0.25% (the probability computation is expressed in Appendix C).

4.2. Assessment on prediction performance

We compare the performance of the proposed BN model with those of standard machine learning methods (LR, NB, and SVM) by using Weka at the default settings.

The 5 replications of the accuracy, sensitivity, and specificity values for the proposed BN, LR, NB, and SVM are presented in Table 6. In terms of sampled average, the proposed Bayesian network performs the best in sensitivity; LR performs the best in accuracy; and NB performs the best in specificity. Note that sampled average requires further statistical test to justify their differences, as stated in the following.

One-way ANOVA was used to test whether the performances of the four models are statistically significant different from each other. The model is treated as *factor*. The diagnostics of ANOVA assumptions are done and shown in Appendix D (Table D1–D3). We used Kolmogorov–Smirnov test for normality and Durbin–Watson test for independency. The results in Tables D1 and D2 show that the data have normality and independency properties. However, the results of Levene test in Table D3 show that the homoscedasticity of the data is failed. Because of the existence of the normality and independency properties and the failure of the homoscedasticity property, we further used Games–Howell statistics for multiple comparison tests. The ANOVA results in Table 7 show no significant differences are observed in accuracy, specificity among the four models, while certain significant difference might occur in sensitivity given a confidence level at 95% ($\alpha=0.05$). Games–Howell test is further conducted. The different means of accuracy, sensitivity, and specificity of models validated by Games–Howell test would be listed in different columns while the in different means are listed in the same column. The results in Table 8 which list the means in the same column confirm that no significant differences are observed in accuracy, sensitivity, and specificity for all models.

5. Conclusion

The prognosis in patients with brain metastases from lung cancer was usually poor in the past [40,41]. In this study, we

Table 6
The performance of models in terms of accuracy, sensitivity, and specificity value.

Model	Bayesian network			Naive Bayes			Support vector machine			Logistic regression		
Replication	Accuracy	Sensitivity	Specificity	Accuracy	Sensitivity	Specificity	Accuracy	Sensitivity	Specificity	Accuracy	Sensitivity	Specificity
1	0.8134	0.8213	0.8060	0.8233	0.7964	0.8486	0.7806	0.6920	0.8678	0.8211	0.8145	0.8273
2	0.7980	0.8095	0.7872	0.8024	0.7914	0.8128	0.7794	0.8209	0.7404	0.8123	0.8141	0.8106
3	0.8386	0.8575	0.8224	0.8452	0.8266	0.8612	0.8266	0.7672	0.8776	0.8485	0.8527	0.8449
4	0.8441	0.8488	0.8403	0.8463	0.8171	0.8703	0.8288	0.7537	0.8902	0.8540	0.8415	0.8643
5	0.8057	0.8270	0.7873	0.8244	0.8104	0.8364	0.7827	0.8341	0.7382	0.8244	0.8270	0.8221
Mean	0.8200	**0.8328**	0.8087	0.8283	0.8084	**0.8459**	0.7996	0.7736	0.8228	**0.8321**	0.8299	0.8338
Std	0.0204	0.0198	0.0230	0.0182	0.0145	0.0225	0.0257	0.0570	0.0767	0.0182	0.0170	0.0210

Table 7
ANOVA analysis (α=0.05).

Index	Source of variance	Sum of squares	*df*	Mean square	*F*	Sig.
Accuracy	Between groups	0.003	3	0.001	2.422	0.104
	Within groups	0.007	16	0.000		
	Total	0.010	19			
Sensitivity	Between groups	0.011	3	0.004	3.611	0.037
	Within groups	0.017	16	0.001		
	Total	0.028	19			
Specificity	Between groups	0.004	3	0.001	0.684	0.575
	Within groups	0.029	16	0.002		
	Total	0.033	19			

Table 8
The results using Games–Howell test (α=0.05).

Index	Model	Mean
Accuracy	Support vector machine	0.7996
	Bayesian network	0.8200
	Naive Bayes	0.8283
	Logistic regression	0.8321
Sensitivity	Support vector machine	0.7736
	Naive Bayes	0.8084
	Logistic regression	0.8300
	Bayesian network	0.8328
Specificity	Bayesian network	0.8086
	Support vector machine	0.8228
	Logistic regression	0.8338
	Naive Bayes	0.8459

resolved this issue by presenting a Bayesian network model. To our knowledge, this is the first time the Bayesian network is used to predict the occurrence of brain metastasis from lung cancer. In the present study, we identified six variables by a rigorous correlation analysis and constructed the Bayesian network model examined by domain experts/doctors, including: (1) age; (2) gender; (3) region of residence, environment of patients; (4) location of lung cancer within human body; (5) treatment; and (6) occurrence, which represents the development of brain metastasis from lung cancer. A nationwide database containing more than 50,000 cases of cancer patients from 1996 to 2010 in Taiwan was used. The profile of lung cancer patients was categorized and the explanatory graphical model and a variety of inferences were presented.

We compared the performance of the proposed Bayesian network model with those of standard machine learning methods (LR, NB, and SVM). The 5 replications of the accuracy, sensitivity, and specificity values for the proposed model, LR, and SVM are analyzed. In terms of sampled average, the proposed model performs the best in sensitivity; LR performs the best in accuracy; and

Table A1
Correlation analysis (Kendall's tau_b).

	Gender	Age	Region	Site	Treatment	Occurrence
Gender	1.000	0.084**	−0.034*	−0.032*	−0.041**	−0.006
Age		1.000	−0.004	−0.022	−0.031*	−0.001
Region			1.000	0.069**	−0.036**	−0.104**
Site				1.000	0.065**	−0.402**
Treatment					1.000	−0.312**
Occurrence						1.000

* Correlation is significant at the 0.05 level (2-tailed).
** Correlation is significant at the 0.01 level (2-tailed).

NB performs the best in specificity. Although a further statistical analysis showed that the proposed model does not significantly overwhelm the benchmark methods, the advantages of using the BN over the benchmark models can be concluded as follows: (1) the proposed model is easily understood by physicians, (2) the proposed model is efficient in modeling both linear and non-linear situations, (3) the ability of the proposed BN to solve stochastic medical decision-making problems, such as the occurrence of brain metastasis from lung cancer, by finding the probability distribution of children nodes given the evidence values of their parents and vice versa, (4) the ability the proposed model to handle situations wherein information are missing, as indicated in a variety of medical related queries/decisions given partial conditions (refer to Appendix C), and (5) the sensitivity is critical to cancer patient identification and the sensitivity of the proposed Bayesian network is the highest among all benchmark methods.

A suggested future study involves extending the proposed Bayesian network model to predict the cost of treatment for lung cancer patients that develop brain metastasis. Some clinical investigation using the proposed model is also suggested.

Conflict of interest statement

None declared.

Acknowledgments

The authors gratefully acknowledge the comments and suggestions of the editor and the anonymous referees. This work is partially supported by the National Science Council, the top-research-university project and the model-of-vocational-university project of Ministry of Education (Taiwan), and National Taiwan University of Science and Technology – Taipei Medical University Joint Research Program.

This study is based in part on data from the National Health Insurance Research Database provided by the Bureau of National

Table B1
The conditional probability table (CPT) for the treatment node.

G	*A*	*R*	*S*	*P*(*T*|*G*,*A*,*R*,*S*)		
				Ra	Ch	Dr
F	< 50	C	162	0.0000	0.0000	1.0000
M	< 50	C	162	0.0000	0.3333	0.6667
F	50–60	C	162	0.3600	0.2000	0.4400
M	50–60	C	162	0.3125	0.0000	0.6875
F	61–70	C	162	0.0000	0.0000	1.0000
M	61–70	C	162	0.3243	0.2432	0.4324
F	> 70	C	162	0.4762	0.1905	0.3333
M	> 70	C	162	0.1389	0.1111	0.7500
F	< 50	E	162	0.0000	0.0000	1.0000
M	< 50	E	162	0.0000	0.0000	1.0000
F	50–60	E	162	0.0000	0.0000	1.0000
M	50–60	E	162	0.0000	0.4118	0.5882
F	61–70	E	162	0.0000	0.2500	0.7500
M	61–70	E	162	0.0000	0.0000	1.0000
F	> 70	E	162	0.0000	0.0000	1.0000
M	> 70	E	162	0.4211	0.2632	0.3158
F	< 50	N	162	0.0000	0.0417	0.9583
M	< 50	N	162	0.1538	0.3462	0.5000
F	50–60	N	162	0.1333	0.2000	0.6667
M	50–60	N	162	0.1071	0.1429	0.7500
F	61–70	N	162	0.3333	0.3333	0.3333
M	61–70	N	162	0.2553	0.0638	0.6809
F	> 70	N	162	0.2381	0.1429	0.6190
M	> 70	N	162	0.1053	0.0526	0.8421
F	< 50	S	162	0.0000	0.6667	0.3333
M	< 50	S	162	0.5000	0.0000	0.5000
F	50–60	S	162	0.5714	0.0000	0.4286
M	50–60	S	162	0.3333	0.0000	0.6667
F	61–70	S	162	0.2500	0.0000	0.7500
M	61–70	S	162	0.4783	0.0000	0.5217
F	> 70	S	162	0.5789	0.0000	0.4211
M	> 70	S	162	0.3684	0.0000	0.6316
F	< 50	C	1620	0.0000	0.3333	0.6667
M	< 50	C	1620	0.0000	0.0000	1.0000
F	50–60	C	1620	0.0000	0.2727	0.7273
M	50–60	C	1620	0.2174	0.0000	0.7826
F	61–70	C	1620	0.6000	0.0000	0.4000
M	61–70	C	1620	0.0714	0.5000	0.4286
F	> 70	C	1620	0.2000	0.4667	0.3333
M	> 70	C	1620	0.1667	0.3750	0.4583
F	< 50	E	1620	0.0000	0.0000	1.0000
M	< 50	E	1620	0.0000	0.0000	1.0000
F	50–60	E	1620	0.0000	0.0000	0.0000
M	50–60	E	1620	0.0000	0.0000	1.0000
F	61–70	E	1620	0.0000	0.0000	1.0000
M	61–70	E	1620	0.0000	0.0000	1.0000
F	> 70	E	1620	0.0000	0.0000	1.0000
M	> 70	E	1620	0.2500	0.2500	0.5000
F	< 50	N	1620	0.1176	0.5882	0.2941
M	< 50	N	1620	0.2667	0.4000	0.3333
F	50–60	N	1620	0.2000	0.2000	0.6000
M	50–60	N	1620	0.4667	0.1333	0.4000
F	61–70	N	1620	0.4211	0.2632	0.3158
M	61–70	N	1620	0.1905	0.4286	0.3810
F	> 70	N	1620	0.2778	0.2778	0.4444
M	> 70	N	1620	0.3846	0.3077	0.3077
F	< 50	S	1620	0.2000	0.2667	0.5333
M	< 50	S	1620	0.0000	0.0000	1.0000
F	50–60	S	1620	0.0000	0.0000	1.0000
M	50–60	S	1620	0.0000	0.2941	0.7059
F	61–70	S	1620	0.4444	0.0000	0.5556
M	61–70	S	1620	0.1600	0.0000	0.8400
F	> 70	S	1620	0.0000	0.7500	0.2500
M	> 70	S	1620	0.2222	0.0000	0.7778
F	< 50	C	1622	0.3077	0.0000	0.6923
M	< 50	C	1622	0.0000	0.0000	1.0000
F	50–60	C	1622	0.0000	0.0000	1.0000
M	50–60	C	1622	0.7778	0.0000	0.2222
F	61–70	C	1622	0.0000	0.0000	1.0000
M	61–70	C	1622	0.3684	0.0000	0.6316
F	> 70	C	1622	0.0000	0.0000	1.0000
M	> 70	C	1622	0.6000	0.0000	0.4000
F	< 50	E	1622	0.0000	0.0000	0.0000

Table B1 (*continued*)

G	*A*	*R*	*S*	*P*(T|*G*,*A*,*R*,*S*)		
				Ra	Ch	Dr
M	< 50	E	1622	0.0000	0.0000	0.0000
F	50–60	E	1622	0.0000	0.0000	0.0000
M	50–60	E	1622	0.0000	0.0000	0.0000
F	61–70	E	1622	0.0000	0.0000	1.0000
M	61–70	E	1622	0.0000	0.0000	1.0000
F	> 70	E	1622	0.0000	0.0000	1.0000
M	> 70	E	1622	0.0000	0.0000	0.0000
F	< 50	N	1622	0.4000	0.0000	0.6000
M	< 50	N	1622	0.0000	0.0000	1.0000
F	50–60	N	1622	0.0000	0.0000	1.0000
M	50–60	N	1622	0.5556	0.0000	0.4444
F	61–70	N	1622	0.0000	0.0000	1.0000
M	61–70	N	1622	0.0000	0.0000	1.0000
F	> 70	N	1622	0.0000	0.0000	1.0000
M	> 70	N	1622	0.0000	0.0000	1.0000
F	< 50	S	1622	0.0000	0.0000	1.0000
M	< 50	S	1622	0.0000	0.0000	1.0000
F	50–60	S	1622	0.0000	0.6111	0.3889
M	50–60	S	1622	0.0000	0.0000	1.0000
F	61–70	S	1622	0.4211	0.3158	0.2632
M	61–70	S	1622	0.5000	0.3571	0.1429
F	> 70	S	1622	0.0000	0.1111	0.8889
M	> 70	S	1622	0.5000	0.2857	0.2143
F	< 50	C	1623	0.4615	0.0000	0.5385
M	< 50	C	1623	0.4286	0.0000	0.5714
F	50–60	C	1623	0.0000	0.0000	1.0000
M	50–60	C	1623	0.6154	0.0000	0.3846
F	61–70	C	1623	0.0000	0.0000	1.0000
M	61–70	C	1623	0.5556	0.1111	0.3333
F	> 70	C	1623	0.0000	0.0000	1.0000
M	> 70	C	1623	0.3077	0.1538	0.5385
F	< 50	E	1623	0.0000	0.0000	0.0000
M	< 50	E	1623	0.0000	0.0000	1.0000
F	50–60	E	1623	0.0000	0.0000	1.0000
M	50–60	E	1623	0.0000	0.0000	1.0000
F	61–70	E	1623	0.0000	0.0000	1.0000
M	61–70	E	1623	0.0000	0.0000	1.0000
F	> 70	E	1623	0.0000	0.0000	1.0000
M	> 70	E	1623	0.0000	0.0000	1.0000
F	< 50	N	1623	0.0000	0.5455	0.4545
M	< 50	N	1623	0.4000	0.0000	0.6000
F	50–60	N	1623	0.3571	0.4286	0.2143
M	50–60	N	1623	0.7000	0.0000	0.3000
F	61–70	N	1623	0.1667	0.0000	0.8333
M	61–70	N	1623	0.3571	0.3571	0.2857
F	> 70	N	1623	0.4000	0.2000	0.4000
M	> 70	N	1623	0.1176	0.5294	0.3529
F	< 50	S	1623	0.0000	0.1923	0.8077
M	< 50	S	1623	0.0909	0.0000	0.9091
F	50–60	S	1623	0.2333	0.0000	0.7667
M	50–60	S	1623	0.1200	0.3200	0.5600
F	61–70	S	1623	0.6667	0.0000	0.3333
M	61–70	S	1623	0.2727	0.0909	0.6364
F	> 70	S	1623	0.5556	0.0000	0.4444
M	> 70	S	1623	0.2941	0.4706	0.2353
F	< 50	C	1624	0.0000	0.0000	1.0000
M	< 50	C	1624	0.0000	0.0000	1.0000
F	50–60	C	1624	0.0000	0.5333	0.4667
M	50–60	C	1624	0.0000	0.0000	1.0000
F	61–70	C	1624	0.0000	0.0000	1.0000
M	61–70	C	1624	0.0000	0.0000	1.0000
F	> 70	C	1624	0.0000	0.0000	1.0000
M	> 70	C	1624	0.0000	0.0000	1.0000
F	< 50	E	1624	0.0000	0.0000	1.0000
M	< 50	E	1624	0.0000	0.0000	0.0000
F	50–60	E	1624	0.0000	0.0000	1.0000
M	50–60	E	1624	0.0000	0.0000	1.0000
F	61–70	E	1624	0.0000	0.0000	0.0000
M	61–70	E	1624	0.0000	0.0000	1.0000
F	> 70	E	1624	0.0000	0.0000	0.0000
M	> 70	E	1624	0.0000	0.0000	1.0000
F	< 50	N	1624	0.7500	0.0000	0.2500
M	< 50	N	1624	0.0000	0.0000	1.0000
F	50–60	N	1624	0.0000	0.0000	1.0000

Table B1 (*continued*)

G	*A*	*R*	*S*	*P*(T|*G*,*A*,*R*,*S*)		
				Ra	Ch	Dr
M	50–60	N	1624	0.0000	0.0000	1.0000
F	61–70	N	1624	0.0000	0.0000	1.0000
M	61–70	N	1624	0.2000	0.5000	0.3000
F	> 70	N	1624	0.0000	0.0000	1.0000
M	> 70	N	1624	0.6667	0.0000	0.3333
F	< 50	S	1624	0.0000	0.8333	0.1667
M	< 50	S	1624	0.0000	0.0000	1.0000
F	50–60	S	1624	0.2000	0.0000	0.8000
M	50–60	S	1624	0.0000	0.2222	0.7778
F	61–70	S	1624	0.7143	0.0000	0.2857
M	61–70	S	1624	0.0000	0.0000	1.0000
F	> 70	S	1624	0.0000	0.0000	1.0000
M	> 70	S	1624	0.3333	0.1667	0.5000
F	< 50	C	1625	0.0000	0.0000	1.0000
M	< 50	C	1625	0.4545	0.0000	0.5455
F	50–60	C	1625	0.4615	0.0000	0.5385
M	50–60	C	1625	0.3125	0.0000	0.6875
F	61–70	C	1625	0.2500	0.0000	0.7500
M	61–70	C	1625	0.3182	0.0000	0.6818
F	> 70	C	1625	0.0000	0.0000	1.0000
M	> 70	C	1625	0.2632	0.0000	0.7368
F	< 50	E	1625	0.0000	0.0000	0.0000
M	< 50	E	1625	0.0000	0.0000	1.0000
F	50–60	E	1625	0.0000	0.0000	1.0000
M	50–60	E	1625	0.0000	0.0000	1.0000
F	61–70	E	1625	0.0000	0.0000	1.0000
M	61–70	E	1625	0.0000	0.0000	0.0000
F	> 70	E	1625	0.0000	0.0000	1.0000
M	> 70	E	1625	0.0000	0.0000	1.0000
F	< 50	N	1625	0.2857	0.0000	0.7143
M	< 50	N	1625	0.6667	0.0000	0.3333
F	50–60	N	1625	0.0000	0.6000	0.4000
M	50–60	N	1625	0.2632	0.3684	0.3684
F	61–70	N	1625	0.2667	0.2000	0.5333
M	61–70	N	1625	0.4500	0.1500	0.4000
F	> 70	N	1625	0.3333	0.0000	0.6667
M	> 70	N	1625	0.1538	0.4615	0.3846
F	< 50	S	1625	0.2500	0.1667	0.5833
M	< 50	S	1625	0.0000	0.0000	1.0000
F	50–60	S	1625	0.2105	0.4211	0.3684
M	50–60	S	1625	0.0000	0.0000	1.0000
F	61–70	S	1625	0.3000	0.0000	0.7000
M	61–70	S	1625	0.4167	0.2500	0.3333
F	> 70	S	1625	0.0000	0.5000	0.5000
M	> 70	S	1625	0.2632	0.4211	0.3158
F	< 50	C	1628	0.2222	0.5556	0.2222
M	< 50	C	1628	0.0000	0.0000	1.0000
F	50–60	C	1628	0.0000	0.3333	0.6667
M	50–60	C	1628	0.0000	0.5000	0.5000
F	61–70	C	1628	0.0714	0.1429	0.7857
M	61–70	C	1628	0.1250	0.1875	0.6875
F	> 70	C	1628	0.1667	0.3333	0.5000
M	> 70	C	1628	0.0000	0.2353	0.7647
F	< 50	E	1628	0.0000	0.0000	1.0000
M	< 50	E	1628	0.0000	0.0000	1.0000
F	50–60	E	1628	0.0000	0.0000	1.0000
M	50–60	E	1628	0.0000	0.0000	1.0000
F	61–70	E	1628	0.0000	0.0000	0.0000
M	61–70	E	1628	0.0000	0.0000	1.0000
F	> 70	E	1628	0.0000	0.0000	1.0000
M	> 70	E	1628	0.0000	0.0000	1.0000
F	< 50	N	1628	0.0000	0.0000	1.0000
M	< 50	N	1628	0.0000	0.0000	1.0000
F	50–60	N	1628	0.0000	0.0000	1.0000
M	50–60	N	1628	0.5714	0.0000	0.4286
F	61–70	N	1628	0.0000	0.0000	1.0000
M	61–70	N	1628	0.0000	0.3750	0.6250
F	> 70	N	1628	0.5238	0.0000	0.4762
M	> 70	N	1628	0.0000	0.4286	0.5714
F	< 50	S	1628	0.0000	0.0000	1.0000
M	< 50	S	1628	0.0000	0.0000	1.0000
F	50–60	S	1628	0.0000	0.0000	1.0000
M	50–60	S	1628	0.0000	0.0000	1.0000
F	61–70	S	1628	0.0000	0.0000	1.0000

Table B1 (*continued*)

G	A	R	S	P(T|G,A,R,S)		
				Ra	Ch	Dr
M	61–70	S	1628	0.0000	0.0000	1.0000
F	> 70	S	1628	0.0000	0.0000	1.0000
M	> 70	S	1628	0.0000	0.0000	1.0000
F	< 50	C	1629	0.2264	0.2075	0.5660
M	< 50	C	1629	0.2982	0.1228	0.5789
F	50–60	C	1629	0.2609	0.1957	0.5435
M	50–60	C	1629	0.1481	0.0926	0.7593
F	61–70	C	1629	0.0732	0.1463	0.7805
M	61–70	C	1629	0.2319	0.0145	0.7536
F	> 70	C	1629	0.0286	0.1714	0.8000
M	> 70	C	1629	0.4146	0.0244	0.5610
F	< 50	E	1629	0.1739	0.3478	0.4783
M	< 50	E	1629	0.1481	0.3333	0.5185
F	50–60	E	1629	0.3043	0.2174	0.4783
M	50–60	E	1629	0.1500	0.4500	0.4000
F	61–70	E	1629	0.0541	0.0000	0.9459
M	61–70	E	1629	0.1765	0.2353	0.5882
F	> 70	E	1629	0.1842	0.0789	0.7368
M	> 70	E	1629	0.1154	0.2308	0.6538
F	< 50	N	1629	0.1700	0.1100	0.7200
M	< 50	N	1629	0.0833	0.0833	0.8333
F	50–60	N	1629	0.0606	0.0707	0.8687
M	50–60	N	1629	0.2222	0.0667	0.7111
F	61–70	N	1629	0.1134	0.0928	0.7938
M	61–70	N	1629	0.2246	0.0870	0.6884
F	> 70	N	1629	0.1019	0.0648	0.8333
M	> 70	N	1629	0.1121	0.0345	0.8534
F	< 50	S	1629	0.2459	0.0820	0.6721
M	< 50	S	1629	0.2162	0.0946	0.6892
F	50–60	S	1629	0.0588	0.1078	0.8333
M	50–60	S	1629	0.2593	0.2037	0.5370
F	61–70	S	1629	0.3854	0.0938	0.5208
M	61–70	S	1629	0.3895	0.0737	0.5368
F	> 70	S	1629	0.2571	0.0571	0.6857
M	> 70	S	1629	0.2959	0.0204	0.6837

Health Insurance, Department of Health and managed by National Health Research Institutes, Taiwan, ROC. The interpretation and conclusions contained herein do not represent those of Bureau of National Health Insurance, Department of Health or National Health Research Institutes.

Appendix A. The correlation analysis

See Table A1.

Appendix B. The conditional probability table (CPT)

See Table B1.

Appendix C

The probability computation

The probability of the occurrence of brain metastasis if the patient is a 55-year-old man diagnosed with lung cancer in the trachea (1620) and living in the eastern part of Taiwan (without any information concerning treatment) is expressed as follows:

$$P(O=Y|G=M,A=50-60\ ,R=E,S=1620)$$
$$=\frac{P(O\ =Y,G=M,A=50-60,R=E,S=1620)}{P(G=M,A=50-60,R=E,S=1620)}$$
$$=\frac{\sum_T P(G=M)P(A=50-60)P(R=E)P(S=1620|G=M,A=50-60,R=E)}{\sum_{T,O}P(G=M)P(A=50-60)P(R=E)P(S=1620|G=M,A=50-60,R=E)}\frac{P(T|G=M,A=50-60,R=E,S=1620)P(O=Y|S=1620,T)}{P(T|G=M,A=50-60,R=E,S=1620)P(O|S=1620,T)}$$
$$=\frac{P(G=M)P(A=50-60)P(R=E)P(S=1620|G=M,A=50-60,R=E)}{P(G=M)P(A=50-60)P(R=E)P(S=1620|G=M,A=50-60,R=E)}\frac{\sum_T P(T|G=M,A=50-60\ ,R=E,S=1620)P(O=Y|S=1620,T)}{\sum_T P(T|G=M,A=50-60\ ,R=E,S=1620)\sum_O P(O|S=1620,T)}$$
$$=\frac{0.2715}{1}$$
$$=0.2715.$$

The probability of the non-occurrence of brain metastasis if the patient is a 55-year-old man diagnosed with lung cancer in the trachea (1620) and living in the eastern part of Taiwan (without any information concerning treatment) is expressed as follows:

$$P(O=N|G=M,A=50-60\ ,R=E,S=1620)$$
$$=1-P(O=Y|G=M,A=50-60\ ,R=E,S=1620)$$
$$=0.7285.$$

The probability for the patient who is a female patient with lung cancer in the trachea, bronchus, and lung (162) should undergo radiotherapy is expressed as follows:

$$P\left(T=Ra\middle|G=F,S=162\right)=\frac{P(T\ =Ra,\ G=F,S=162)}{P(G=F,S=162)}$$

$$= \frac{\sum_{A,R,O} P(G=F)P(A)P(R)P(S=162|G=F,A,R)P(T=Ra|G=F,A,R,S=162)P(O|S=162,T=Ra)}{\sum_{A,R,T,O} P(G=F)P(A)P(R)P(S=162|G=F,A,R)P(T|G=F,A,R,S=162)P(O|S=162,T)}$$

$$= \frac{P(G=F)\sum_O P(O|S=162,T=Ra)\sum_{A,R} P(A)P(R)P(S=162|G=F,A,R)P(T=1|G=F,A,R,S=162)}{P(G=F)\sum_A P(A)\sum_R P(R)P(S=162|G=F,A,R)\sum_T P(T|G=F,A,R,S=162)\sum_O P(O|S=162,T)}$$

$$= 0.1335.$$

The probability of the occurrence of brain metastasis in a patient living in central Taiwan given a patient undergoing chemotherapy for lung cancer in the trachea, bronchus, and lung (162) and having brain metastasis is expressed as follows:

$$P\left(R=C \middle| S=162,\ T=Ch, O=Y\right) = \frac{P(R=C,\ S=162,\ T=Ch, O=Y)}{P(S=162,\ T=Ch, O=Y)}$$

$$= \frac{\sum_{G,A} P(G)P(A)P(R=C)P(S=162|G,A,R=C)P(T=Ch|G,A,R=C,S=162)P(O=Y|S=162,T=Ch)}{\sum_{G,A,R} P(G)P(A)P(R)P(S=162|G,A,R=C)P(T=Ch|G,A,R,S=162)P(O=Y|S=162,T=Ch)}$$

$$= \frac{P(R=C)P(O=Y|S=162,T=Ch)\sum_{G,A} P(G)P(A)P(S=162|G,A,R=C)P(T=Ch|G,A,R=C,S=162)}{P(O=Y|S=162,T=Ch)\sum_{G,A,R} P(G)P(A)P(R)P(S=162|G,A,R)P(T=Ch|G,A,R=C,S=162)}$$

$$= \frac{(0.2507)(0.0482)(0.0258)}{(0.0482)(0.0188)}$$

$$= \frac{0.0003117}{0.000906} = 0.3440.$$

The probability of a patient with lung cancer in the middle lobe, bronchus, or lung (1624) given female patient aged lower 50 years old and having brain metastasis is expressed as follows:

$$P\left(S=1624 \middle| G=F, A=\ <50, O=Y\right) = \frac{\sum_{R,T} P(G=F, A=<50, S=1624,\ O=Y)}{\sum_{R,T,S} P(G=F, A=<50,\ O=Y)}$$

Table D1
Kolmogorov–Smirnov test for normality (α=0.05).

Index	Model	p Value
Accuracy	Bayesian network	0.2000
	Support vector machine	0.0520
	Logistic regression	0.2000
	Naive Bayes	0.2000
Sensitivity	Bayesian network	0.2000
	Support vector machine	0.2000
	Logistic regression	0.2000
	Naive Bayes	0.2000
Specificity	Bayesian network	0.2000
	Support vector machine	0.1010
	Logistic regression	0.2000
	Naive Bayes	0.2000

Table D2
Durbin–Watson test for independency (α=0.05).

Index	Statistic
Accuracy	$d=1.8537 > d_U$
Sensitivity	$d=2.6542 > d_U$
Specificity	$d=2.7940 > d_U$

Table D3
Levene test for homoscedasticity (α=0.05).

Index	p Value
Accuracy	0.281
Sensitivity	**0.026**
Specificity	**0.000**

$$= \frac{\sum_{R,T} P(G=F)P(A=<50)P(R)P(S=1624|G=F,A=<50,R)P(T|G=F,A=<50,R,S=1624)P(O=Y|S=1624,T)}{\sum_{R,T,S} P(G=F)P(A=<50)P(R)P(S|G=F,A=<50,R)P(T|G=F,A=<50,R,S)P(O=Y|S,T)}$$

$$= \frac{P(G=F)P(A=<50)\sum_R P(R)P(S=1624|G=F,A=<50,R)\sum_T P(T|G=F,A=<50,R,S=1624)P(O=Y|S=1624,T)}{P(G=F)P(A=<50)\sum_{R,T,S} P(R)P(S|G=F,A=<50,R)P(T|G=F,A=<50,R,S)P(O=1|S,T)}$$

$$= 0.0025.$$

Appendix D

The diagnosis of ANOVA assumptions
See Tables D1–D3.

References

[1] American Cancer Society, Global Cancer Facts and Figures, 2nd edition, American Cancer Society, Atlanta, 2011.
[2] T. Badriyah, J.S. Briggs, D.R. Prytherch, Decision trees for predicting risk of mortality using routinely collected data, Int. J. Soc. Hum. Sci. 6 (2012) 303–306.
[3] A. Bajard, V. Westeel, A. Dubiez, P. Jacoulet, D. Pernet, J.C. Dalphin, A. Depierre, Multivariate analysis of factors predictive of brain metastases in localised non-small cell lung carcinoma, Lung Cancer 45 (2004) 317–323.
[4] S. Bozkurt, A. Uyar, Comparison of Bayesian network and binary logistic regression methods for prediction of prostate cancer, in: Proceedings of the 2011 4th International Conference on Biomedical Engineering and Informatics (BMEI), 2011, pp. 1689–1691.
[5] J.S. Brown, D. Eraut, C. Trask, A.G. Davison, Age and the treatment of lung cancer, Thorax 51 (1996) 564–568.
[6] M.L. Cartman, A.C. Hatfield, M.F. Muers, M.D. Peake, R.A. Haward, D. Forman, Lung cancer: district active treatment rates affect survival, J. Epidemiol. Community Health 56 (2002) 424–429.
[7] A. Chi, R. Komaki, Treatment of brain metastasis from lung cancer, Cancers 2 (2010) 2100–2137, http://dx.doi.org/10.3390/cancers2042100.
[8] N. Cruz-Ramírez, H.G. Acosta-Mesa, H. Carrillo-Calvet, L.A. Nava-Fernández, R.E. Barrientos-Martínez, Diagnosis of breast cancer using Bayesian networks: a case study, Comput. Biol. Med. 37 (2007) 1553–1564.
[9] G. Corani, C. Magli, A. Giusti, L. Gianaroli, L.M. Gambardella, A Bayesian network model for predicting pregnancy after in vitro fertilization, Comput. Biol. Med. 43 (2013) 1783–1792.
[10] A. Dekker, C. Dehing-Oberije, D. De uysscher, P. Lambin, A. Hope, K. Komati, G. Fung, Y.U. Shipeng, W. De Neve, Y. Lievens, Survival prediction in lung cancer treated with radiotherapy bayesian networks vs. support vector machines in handling missing data, in: Proceeding of International Conference on Machine Learning and Applications, IEEE, 2009, pp. 494–497.
[11] P. de Groot, R.F. Munden, Lung cancer epidemiology, risk factors, and prevention, Radiol. Clin. North Am. 50 (5) (2012) 863–876.
[12] L.F. Forrest, J. Adams, H. Wareham, G. Rubin, M. White, Socioeconomic inequalities in lung cancer treatment: systematic review and meta – analysis, PLOS Med. 10 (2) (2013) 1–25.
[13] I.T. Gavrilovic, J.B. Posner, Brain metastases: epidemiology and pathophysiology, J. Neuro-Oncol. 75 (2005) 5–14, http://dx.doi.org/10.1007/s11060-004-8093-6.
[14] O. Graesslin, Nomogram to predict subsequent brain metastasis in patients with metastatic breast cancer, J. Clin. Oncol. 28 (12) (2010) 2032–2037.
[15] X. Guo, T. Sun, H. Wu, W. He, Z. Liang, M. Zhang, A. Guo, W. Wang, Support vector machine prediction model of early-stage lung cancer based on curvelet transform to extract texture features of CT image, World Acad. Sci. Eng. Technol. 71 (2010) 333–337.
[16] W. Hämäläinen, M. Vinni, Comparison of machine learning methods for intelligent tutoring systems, in: Proceedings of the Eighth International Conference in Intelligent Tutoring Systems, Taiwan, 2006, pp. 525–534.
[17] N. Hoot, D. Aronsky, Using Bayesian networks to predict survival of liver transplant patients, in: Proceedings of AMIA 2005 Symposium Proceedings, Washington, DC, 2005, pp. 345–349.

[18] D.W. Hosmer, S. Lemeshow, Applied Logistic Regression, 2nd ed., A Wiley-Interscience Publication, John Wiley & Sons Inc., New York, New York, USA, 2000.
[19] J.L. Hubbs, J.A. Boyd, D. Hollis, J.P. Chino, M. Saynak, C.R. Kelsey, Factors associated with the development of brain metastases, Cancer (2010) 5038–5046, http://dx.doi.org/10.1002/cncr.25254.
[20] K. Jayasurya, G. Fung, S. Yu, C. Dehing-Oberije, D. De Ruysscher, A. Hope, W. De Neve, Y. Lievens, P. Lambin, A.L.A.J. Lambin, Comparison of Bayesian network and support vector machine models for two-year survival prediction in lung cancer patients treated with radiotherapy, Med. Phys. 37 (4) (2010) 1401–1407.
[21] C.E. Kahn , L.M. Roberts, K.A. Shaffer, P. Haddawy, Construction of a Bayesian network for mammographic diagnosis of breast cancer, Comput. Biol. Med. 27 (1997) 19–29.
[22] Y.-C. Ko, C.-H. Lee, M.-J. Chen, C.-C. Huang, W.-Y. Chang, H.-J. Lin, H.-Z. Wang, P.-Y. Chang, Risk factors for primary lung cancer among non-smoking women in Taiwan, Int. J. Epidemiol. 26 (1) (1997) 24–31.
[23] Y.-C. Ko, J.-L. Wang, C.-C. Wu, W.-T. Huang, M.-C. Lin, Lung cancer at a medical center in southern Taiwan, Chang Gung Med. J. 28 (2005) 387–395.
[24] L. Lalande, L. Bourguignon, C. Carlier, M. Ducher, Bayesian networks: a new method for the modeling of bibliographic knowledge application to fall risk assessment in geriatric patients, Med. Biol. Eng. Comput. 51 (2013) 657–664.
[25] A. Y-C. Liu, The Effect of Oversampling and Undersampling on Classifying Imbalanced Text Datasets (Master thesis), University of Texas, USA, 2004University of Texas, USA, 2004.
[26] A. Lorensuhewa, B. Pham, S. Geva, Inferencing design styles using Bayesian Networks, Ruhuna J. Sci. 1 (2006) 113–124.
[27] D. Lowd, P. Domingos, Naive Bayes models for probability estimation, in: Proceedings of the Twentysecond International Conference on Machine Learning, ACM Press, 2005, pp. 529–536.
[28] P.J. Lucas, L.C. van der Gaag, A. Abu-Hanna, Bayesian networks in biomedicine and health-care, Artif. Intell. Med. 30 (3) (2004) 201–214.
[29] F. Mancini, F.S. Sousa, A.D. Hummel, A.E.J. Falcão, L.C. Yi, C.F. Ortolani, D. Sigulem, I.T. Pisa, Classification of postural profiles among mouth-breathing children by learning vector quantization, Methods Inf. Med. 50 (4) (2011) 349–357.
[30] F.M. Medina, R.R. Barrera, J.F. Morales, R.C. Echegoyen, J.G. Chavarria, F.T. Rebora, Primary lung cancer in Mexico city: a report of 1019 cases, Lung Cancer 14 (1996) 185–193.
[31] D.A. Morales, E. Bengoetxea, P. Larrañaga, Selection of human embryos for transfer by Bayesian classifiers, Comput. Biol. Med. 38 (2008) 1177–1186.
[32] A. Mujoomdar, J.H.M. Austin, R. Malhotra, C.A. Powell, G.D.N. Pearson, M.C. Shiau, H. Raftopoulos, Clinical predictors of metastatic disease to the brain from non-small cell lung carcinoma: primary tumor size, cell type, and lymph node metastases, Radiology 242 (3) (2007) 882–888.
[33] K. Murphy, Bayes Net Toolbox, Retrieved ⟨http://www.ai.mit.edu/murphyk/Software/BNT/bnt.html⟩, 2002.
[34] National Health Insurance Research Database, Taiwan. Bureau of National Health Insurance, Department of Health and managed by National Health Research Institutes, Available from URL: ⟨http://www.nhri.org.tw/nhird/en/index.htm⟩ (Accessed 2013 Jan.).
[35] O. Nee, A. Hein, Clinical Decision Support with Guidelines and Bayesian Networks, Decision Support Systems Advances in. Retrieved from ⟨http://www.intechopen.com/books/decision-support-systems-advances-in⟩, 2010.
[36] J.H. Oh, J. Craft, R.A. Lozi, M. Vaidya, Y. Meng, J.O. Deasy, J.D. Bradley, I.E. Naqa, A Bayesian network approach for modeling local failure in lung cancer, Phys. Med. Biol. 56 (6) (2011) 1635–1651, http://dx.doi.org/10.1088/0031-9155/56/6/008.
[37] A. Oikawa, H. Takahashi, H. Ishikawa, K. Kurishima, K. Kagohashi, H. Satoh, Application of conditional probability analysis to distant metastases from lung cancer, Oncol. Lett. 3 (2012) 629–634.
[38] N. Penel, A. Brichet, B. Prevost, A. Duhamel, R. Assaker, F. Dubois, J.-J. Lafitte, Prognostic factors of synchronous brain metastases from lung cancer, Lung Cancer 33 (2001) 143–154.
[39] K. Pietzner, G. Oskay-Oezcelik, K. Khalfaoui, D. Boehmer, W. Lichtenegger, J. Sehouli, Brain metastases from epithelial ovarian cancer: overview and optimal management, Anticancer Res. *29* (2009) 2793–2798.
[40] P.E. Postmus, H. Haaxma-Reiche, E.F. Smit, H.J.M. Groen, H. Karnicka, T. Lewinski, J.V. Meerbeeck, M. Clerico, A. Gregor, D. Curran, T. Sahmoud, A. Kirkpatrick, G. Giaccone, Treatment of brain metastases of small-cell lung cancer: comparing teniposide and teniposide with whole-brain radio therapy – a phase III study of the european organization for the research and treatment of cancer lung cancer cooperative group, J. Clin. Oncol. 18 (19) (2000) 3400–3408.
[41] P. Rodrigus, P. de Brouwer, E. Raaymakers, Brain metastases and non-small cell lung cancer: prognostic factors and correlation with survival after irradiation, Lung Cancer 32 (2001) 129–136.
[42] J.M. Samet, E. Avila-Tang, P. Boffetta, L.M. Hannan, S. Olivo-Marston, M.J. Thun, C.M. Rudin, Lung cancer in never smokers: clinical epidemiology and environmental risk factors, Clin. Cancer Res. 15 (18) (2011) 5626–5645.
[43] J. Su, H. Zhang, Full Bayesian network classifiers, in: Proceedings of the 23rd International Conference on Machine Learning (ICML '06), NY, USA, ACM, 2006, pp. 1–8.
[44] C.R. Twardy, A.E. Nicholson, K.B. Korb, J. McNeil, Epidemiological data mining of cardiovascular Bayesian networks, Electron. J. Health Inf. 1 (1) (2006) 1–13 (e3).
[45] L. Uusitalo, Advantages and challenges of Bayesian networks in environmental modeling, Ecol. Model. 203 (2007) 312–318.
[46] S. Visscher, P.J.F. Lucas, C.A.M. Schurink, M.J.M. Bonten, Modelling treatment effects in a clinical Bayesian network using Boolean threshold functions, Artif. Intell. Med. 46 (2009) 251–266.
[47] WHO Department of Gender, Women and Health, Gender in Lung Cancer and Smoking Research, Department of Gender, Women and Health Family and Community Health World Health Organization, Geneva, Switzerland, 2004.
[48] Y. Xia, V.K. Prasanna, Junction tree decomposition for parallel exact inference, in: Proceedings of Parallel and Distributed Processing, Los Angeles, CA, 2008, pp. 1–12.
[49] S. Zhang, C. Tjortjis, X. Zeng, H. Qiao, I. Buchan, J. Keane, Comparing data mining methods with logistic regression in childhood obesity prediction, Inf. Syst. Front. 11 (2009) 449–460.
[50] C.P. de Campos, Q. Ji, Improving bayesian network parameter learning using constraints, Pattern Recognit. (2008) 1–4.

Kung-Jeng Wang is a Professor of the department of Industrial Management of National Taiwan University of Science and Technology (Taiwan Tech). He received his PhD in industrial engineering from University of Wisconsin at Madison. Dr. Wang works closely with industries for research on manufacturing management and resource portfolio planning. He has published about 60 academic articles in international academic journals, such as in IEEE Transactions on Systems, Man, and Cybernetics, IIE Transactions, European Journal of Operational Research, International Journal of Production Research, and Journal of Robotics and CIM. His current research interests are in the areas of intelligent systems, biomedical informatics, and supply chain management.

Bunjira Makond is PhD candidate in the Department of Industrial Management of National Taiwan University of Science and Technology. Her research interest is in the area of biomedical informatics.

Kung-Min Wang is a surgery doctor in Shin-Kong Wu Ho-Su Memorial Hospital, Taiwan. His research interests are in the areas of biomedical informatics, clinical medicine and cancer treatment.