

SCIENTIFIC REPORTS



OPEN

Application of tabu search-based Bayesian networks in exploring related factors of liver cirrhosis complicated with hepatic encephalopathy and disease identification

Zhuang Zhang¹, Jie Zhang², Zhen Wei³, Hao Ren¹, Weimei Song¹, Jinhua Pan¹, Jinchun Liu⁴, Yanbo Zhang¹ & Lixia Qiu¹

This study aimed to explore the related factors and strengths of hepatic cirrhosis complicated with hepatic encephalopathy (HE) by multivariate logistic regression analysis and tabu search-based Bayesian networks (BNs), and to deduce the probability of HE in patients with cirrhosis under different conditions through BN reasoning. Multivariate logistic regression analysis indicated that electrolyte disorders, infections, poor spirits, hepatorenal syndrome, hepatic diabetes, prothrombin time, and total bilirubin are associated with HE. Inferences by BNs found that infection, electrolyte disorder and hepatorenal syndrome are closely related to HE. Those three variables are also related to each other, indicating that the occurrence of any of those three complications may induce the other two complications. When those three complications occur simultaneously, the probability of HE may reach 0.90 or more. The BN constructed by the tabu search algorithm can analyze not only how the correlative factors affect HE but also their interrelationships. Reasoning using BNs can describe how HE is induced on the basis of the order in which doctors acquire patient information, which is consistent with the sequential process of clinical diagnosis and treatment.

Hepatic encephalopathy (HE) is a very serious complication of liver disease. It is a comprehensive symptom of central nervous system dysfunction based on metabolic disorders, caused by acute or chronic severe liver function disorder, or abnormal portal-systematic circulation¹. Most patients with cirrhosis have different degrees of HE at varying stages of the disease, the mortality rate of which is extremely high^{2,3}. The 1-year survival rate of patients who have chronic liver disease complicated with HE is 0.42, and their 3-year survival rate is 0.23⁴. HE is divided into dominant and occult types¹: dominant HE has obvious clinical manifestations, such as disturbance of consciousness, behavioral disorders, and coma⁵; whereas the clinical signs of occult HE are difficult to detect. Thus, prevention and early identification of HE could be a reasonable and effective way to reduce HE harm. At present, there is no “gold standard” for HE diagnosis. Clinicians can only conduct auxiliary diagnostic evaluation of HE from different angles such as liver function tests, imaging examinations, and neuropsychiatric abnormalities. Occult HE often has no clinical symptoms, and only through careful neuropsychological or neurophysiological examination can patients be found to have cognitive dysfunction. However, the neuropsychological

¹Department of Health Statistics, School of Public Health, Shanxi Medical University, No.56 XinJian South Road, Taiyuan, Shanxi, 030001, China. ²Department of Basis, Jinci College of Shanxi Medical University, No.2 Jinliao Road, Jinci Town, Jinyuan District, Taiyuan, Shanxi, 030025, China. ³Puruisheng (Beijing) Pharmaceutical Technology Development Co., Ltd., No. 188 South Fourth Ring Road, Fengtai District, Beijing, 100071, China. ⁴Department of Gastroenterology, First Hospital of Shanxi Medical University, No.85 JieFang South Road, Taiyuan, Shanxi, 030001, China. Zhuang Zhang, Jie Zhang and Zhen Wei contributed equally. Correspondence and requests for materials should be addressed to L.Q. (email: qlx_1126@163.com)

testing process is cumbersome, can consume too much time, and is easily susceptible to the patient's condition, educational level, and cultural background. Neurophysiological tests are inaccurate and require complex instruments, and their routine clinical applicability is relatively limited⁶. In short, the pathogenesis of HE is complex and diverse, and its clinical symptoms are variable and sometimes difficult to find. There are no clear diagnosis or treatment guides for it at home or abroad. Therefore, search methods for clinical diagnosis and treatment of patients with cirrhosis use patients' clinical manifestations and factors that affect the occurrence of HE to identify HE, which has important clinical diagnostic value and is worthy of our research.

In recent years, with the development of hospital information technology, hospital information systems contain large amounts of patient and medical data such as patient sociodemographic characteristics, clinical symptoms, physiological and biochemical indicators, and auxiliary examinations, which have formed large datasets^{7,8}. Full and reasonable use of these data is an important basis for current big data analysis. We can use traditional logistic regression on the big data generated during clinical diagnosis and treatment of patients with cirrhosis to establish a prediction model for the incidence of HE in patients with cirrhosis⁹, and finally make judgments based on clinical expertise. Logistic regression requires that the independent variables be independent⁹, but the clinical manifestations of HE and the factors affecting the occurrence of HE are not independent, and there may be interactions that form complex networks of relationships. Therefore, it is difficult for logistic regression to accurately reveal the potential overall information in the data, and it is impossible for logistic regression to infer which factors play a role in the occurrence and development of HE. In addition, any logistic regression model that is applied must predict the probability of developing HE when the state of each variable is known. In clinical practice, the factors used for model prediction may be missing: patients may conceal certain information, or some items may not be detected in clinical diagnosis, which could lead to the failure of logistic prediction.

Bayesian networks (BNs), proposed by Judea Pearl in the 1980s, reflect the potential relationships and relationship strengths between factors by constructing directed acyclic graphs (DAGs) and conditional probability distribution tables¹⁰. BNs do not have strict requirements for statistical assumptions in the modeling process¹¹. Although BNs do not extract variables' main effects and interaction effects, they comprehensively reflect the complex relationships between variables according to the overall structure of the network and can accurately reveal potential overall information in the data¹². BNs can use missing data to identify patients¹³ in a manner that reasonably reflects the sequentiality of clinical diagnosis and treatment^{14,15}. The advantages of BNs in clinical applications have also been demonstrated in our previous studies. Because there were too many initial variables in this study's initial cirrhosis data, some variables were only weakly related to the occurrence and development of HE. The direct use of a BN to establish the model leads to the selection of all variables, which is not conducive to accurate analysis. To simplify the network's relationships and improve network performance, in this study we employed univariate and multivariate logistic regression analysis to screen for related factors of cirrhosis with HE¹⁶, and then constructed a BN to explore the related factors of cirrhosis complicated with HE, revealing the complex network relationships among related factors¹⁷. During the care process of patients with cirrhosis, we can continuously update the network probabilities according to the patient information obtained by doctors. We also adopt BNs to conduct risk reasoning about the probability of HE in patients with cirrhosis^{18,19}.

Materials and Methods

Data collection. We collected information about patients with cirrhosis who were hospitalized in the Department of Gastroenterology, First Hospital of Shanxi Medical University from January 2006 to December 2015 and who had complete medical records. We consulted a large number of documents and carefully analyzed the cases' structure and content and selected indicators that may be related to the progression of cirrhosis. Biochemical indicators were taken as the results of the first test within 24 hours after admission. After the data were sorted, 950 cases were effective, of which 68 were complicated by HE.

There are 23 factors in the survey: demographic characteristics (age, gender), lifestyle (smoking, drinking), past medical history (e.g., hypertension and hepatic diabetes), clinical manifestations (infection, electrolyte disorder, hepatorenal syndrome, spontaneous peritonitis, upper gastrointestinal bleeding, poor spirits, liver disease face, spider nevus, liver palm, jaundice, abdominal varicose veins, splenomegaly, hepatomegaly, ascites), and biochemical indicators (serum albumin [g/L], serum total bilirubin [mmol/L], and prothrombin time [s]). All of this study's methods were carried out in accordance with relevant guidelines and regulations.

Ethical statements. This study was approved by the First Hospital of Shanxi Medical University Ethics Committee. Informed consent was signed by all study participants or their agents. The present study was conducted according to the Declaration of Helsinki.

Quality control. This study's investigators were clinicians and graduate students. Before the start of the survey, the clinicians instructed the graduate students to familiarize themselves with clinical records and to determine the location of the information to be collected in the medical records. Finally, the clinicians normalized the course of looking up medical records. Data entry was performed by double entry, and a person reviewed both sets of entries. If anomalous values were discovered during the review process, the original case was extracted and the erroneous information was corrected.

Bayesian networks. BNs are increasingly popular data-mining techniques that are used with categorical data and effective theoretical models to represent uncertain knowledge and reasoning²⁰. As a reasoning method that employs probabilistic uncertainty, BNs have served in intelligent systems and have been successfully applied in medical diagnosis, expert systems, statistical decision making, learning, and prediction²¹.

BNs represent a set of random variables and their conditional interdependencies via a Directed Acyclic Graph (DAG)²². The DAG's nodes represent random variables $U = \{X_1, \dots, X_n\}$, and the directed edges E represent the

probabilistic dependency relations between variables^{23,24}. If there is a directional arc from X_i to X_j , we can informally regard that X_i causes X_j , so that X_i and X_j are usually said to be parent and child, respectively. Each node has a conditional probability distribution table that represents the conditional probability distribution of each node in the state of a given parent node. The BN is a representation of the joint probability distribution of a set of random variables $X = \{X_1, \dots, X_n\}$, so a probability expression can be obtained:

$$P(X_1, \dots, X_n) = P(X_1)P(X_2|X_1)\dots P(X_n|X_1, X_2, \dots, X_{n-1}) = \prod_1^n P(X_i|\pi(X_i)). \quad (1)$$

$\pi(X_i)$ is the collection of the parent of X_i , $\pi(X_i) \subseteq \{X_1, \dots, X_{i-1}\}$.

The first step of constructing a BN is to determine the nodes. To avoid including too many nodes and introducing excessive complexity to the network structure, we should initially filter the nodes in combination with prior expert knowledge or traditional statistical methods. Then, we should find the optimal model based on the tabu search algorithm. Finally, on the basis of structural learning, we should calculate the network node conditions using the maximum likelihood estimation method.

Tabu search algorithm. The tabu search algorithm is a sub-heuristic algorithm that simulates human memory function. It has the characteristics of few parameters, simple structure, and robust global optimization ability. For a given current network structure, the algorithm uses three operations to generate neighborhoods without generating a network loop: add edges, reduce edges, and reverse edges, and then it searches for a local optimal solution in the neighborhood and puts it into a tabu table. The tabu table is then used to search for the local optimal solution, so that the next search is as far as possible from the current one to avoid duplication of the search process. In conjunction with the use of contempt criteria to amnesty some of the optimal solutions in the tabu table, the tabu restrictions are ignored. When these two steps iterate and search in a circular way, it results in a taboo table that is updated constantly, and ultimately, we obtain the global optimal solution²⁵.

Definitions. (1) Cirrhosis: comply with the diagnostic criteria jointly revised by the Chinese Medical Association's Infectious Diseases and Parasitic Diseases Branch and the Hepatology Branch in 2000¹; (2) HE: refers to the "definition of HE definition, diagnosis, and quantification" standard issued by the 11th World Conference on Gastroenterology Work Group in March 2003, with the diagnosis mainly based on cirrhosis history, neurological abnormalities, liver function tests, image examination, and other auxiliary examination, excluding nerve abnormalities caused by other diseases²⁶; (3) Hypertension: according to the World Health Organization (WHO) standards, systolic blood pressure SBP ≥ 140 mmHg and/or diastolic blood pressure DBP ≥ 90 mmHg, or having a previous history of hypertension and currently taking antihypertensive drugs²⁷; (4) Diabetes mellitus: fasting blood glucose (FBG) ≥ 7.0 mmol/L or 2-hour postprandial plasma glucose ≥ 11.1 mmol/L or reporting diagnosed diabetes mellitus²⁸; (5) Smoker: smoking ≥ 1 cigarette daily on average over the previous 6 months; (6) Drinker: drinking at least once per week regularly for 6 consecutive months with average intake quantity of more than 50 g²⁹.

Establish database and statistical analysis. In our study, we used Epidata to build a database and enter data. Statistical Analysis System (SAS) version 9.2 software (SAS Institute Inc., Cary, NC, USA) was used for data pretreatment. Descriptive statistics and logistic regression analysis of the data was performed with SPSS Version 22 (IBM Corp., Armonk, NY, USA). Significance for all statistical tests was set a priori at $P < 0.05$, and all P values were two-tailed. Weka 3.8.0 (Waikato Environment for Knowledge Analysis; the University of Waikato, New Zealand) was used for structural learning of BNs and parameter estimation of BNs using the maximum likelihood estimation method. The BNs' topology and conditional probability distribution tables were drawn using Netica (Norsys Software Corp., Vancouver, BC, Canada).

Results

Characteristics of the study population. The total enrollment was 950 patients with liver cirrhosis, including 508 male (53.5%) and 442 female (46.5%) individuals ranging 17–88 years old, with a negatively skewed distribution. The overall median age was 58.0 years (QR = 20.3). Among the subjects, 156, 357, and 437 were aged < 45 , 45–59, and ≥ 60 years, respectively (Fig. 1).

Univariate analysis. There were 68 cases of HE among the 950 total patients with cirrhosis. The average prevalence of HE was 7.16%. To investigate the prevalence of concurrent HE, The patients were grouped by 23 factors such as demographic information, clinical comorbidities, clinical manifestations, and biochemical parameters. Suppl. Table S1 shows the factors and their assignment. Among the factors, 15 of them were statistically significant in patients with cirrhosis complicated by HE: infection, electrolyte disorder, hepatorenal syndrome, spontaneous peritonitis, hepatic diabetes, poor spirits, liver disease face, spider nevus, liver palm, jaundice, abdominal varicose veins, hepatomegaly, albumin, increased total bilirubin, and prothrombin time prolongation (all Ps < 0.05). Suppl. Tables S2, S3, and S4 represent differences in the prevalence of HE among patients with cirrhosis who had different characteristics.

Multivariate analysis. The multivariate logistic regression was accompanied by stepwise variable selection to analyze the 15 factors related to HE. The variable selection and rejection criteria were 0.05 and 0.1, respectively.

The final regression model contained seven factors: infection, electrolyte disorders, hepatorenal syndrome, hepatic diabetes, poor spirits, increased total bilirubin, and prothrombin time prolongation. Among them, electrolyte disorders (RR = 6.861), hepatorenal syndrome (RR = 3.467), and infection (RR = 3.021) were the major risk factors for HE in patients with cirrhosis. The relationships between hepatic diabetes, poor spirits, increased

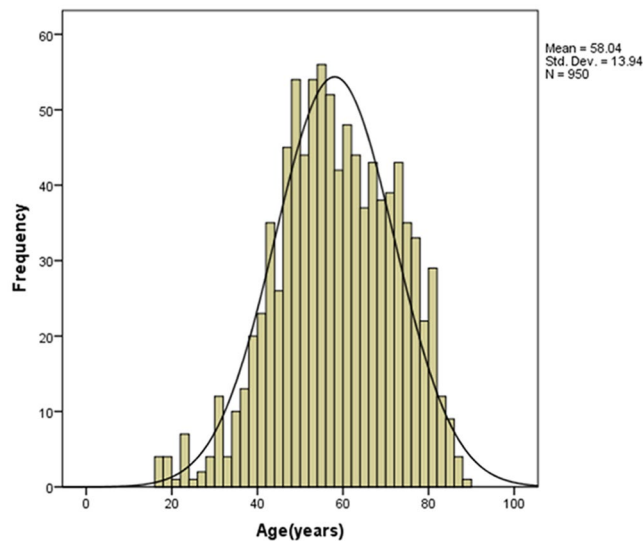


Figure 1. Age distribution characteristics of 950 patients with cirrhosis. The figure was plotted using Spss22.0 (<https://www.ibm.com>).

Factors	β	SE	Wald χ^2	P	RR(95% CI)
Electrolyte disorder	1.926	0.369	27.173	<0.001	6.861(3.326~14.154)
Hepatorenal syndrome	1.243	0.638	3.799	0.051	3.467(0.993~12.101)
Infection	1.106	0.331	11.160	0.001	3.021(1.579~5.778)
Poor spirit	1.001	0.292	11.755	0.001	2.721(1.535~4.822)
Hepatic diabetes	0.871	0.340	6.558	0.010	2.390(1.227~4.656)
Prothrombin time (s)	0.830	0.267	9.691	0.002	2.293(1.360~3.866)
Total bilirubin (mmol/L)	0.703	0.172	16.720	<0.001	2.020(1.442~2.829)

Table 1. Multivariate logistic regression analyses of related factors of hepatic encephalopathy. $\alpha_{in} = 0.05$, $\alpha_{out} = 0.10$.

total bilirubin, prothrombin time prolongation, and HE were relatively weak, with RR values ranging 2.1–2.7 (Table 1).

Logistic regression was used to screen possible factors related to HE and evaluate their associated risk intensity. However, logistic regression analysis requires independence among variables and cannot account for the correlation between independent variables. The model can only explain the direct relationship between these factors and HE. In terms of information about current status, the results showed only factors associated with HE and did not reflect the complex relationship between the factors, diseases, and outcomes.

Bayesian network model. BNs are DAGs, which can help to describe the relationship between HE and these possible factors.

Construct Bayesian network structure. The algorithm for constructing a BN structure is the tabu search algorithm, which is achieved by the Weka tabu search module. The path is Weka.classifiers.bayes.net.search.local.TabuSearch. The BN model of HE-related factors consists of 8 nodes and 10 directed edges. Each node represents HE, electrolyte disorders, infections, poor spirits, hepatorenal syndrome, diabetes, prothrombin time, and total bilirubin (Fig. 2).

Figure 2 shows that various factors establish links with HE through complex relationships: hepatorenal syndrome, electrolyte disorders, infections, poor spirits, total bilirubin, and prothrombin time are directly related to HE, and hepatic diabetes is indirectly associated with total bilirubin. In addition, hepatorenal syndrome is indirectly linked with HE via its association with electrolyte disorders and infections.

Bayesian network parameter learning. Based on the BN structure, we used the maximum likelihood estimation method to estimate the probability of each node in the network conditions. Figure 3 shows the a priori probability of each node in the network. The probability of HE in patients with liver cirrhosis was 0.0736. As depicted in Table 2, if patients had hepatorenal syndrome but not electrolyte disorder, the incidence of HE was reduced to 0.048; if only hepatorenal syndrome occurred, the incidence of HE was 0.300; if only electrolyte disorder, the probability of concurrent HE was 0.383; and when hepatorenal syndrome and electrolyte disorder occurred



Figure 2. The Bayesian Network model for related factors of HE. The figure was plotted using Weka3.8.0 (<http://weka.wikispaces.com/>).

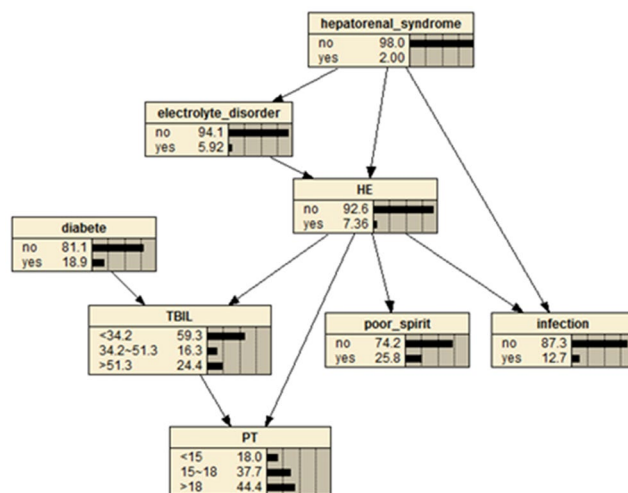


Figure 3. Prior probability of each node in the Bayesian Network. The figure was plotted using Netica (www.norsys.com).

Parent nodes		HE	
Hepatorenal syndrome	Electrolyte disorder	YES	NO
YES	YES	0.667	0.333
YES	NO	0.300	0.700
NO	YES	0.383	0.617
NO	NO	0.048	0.952

Table 2. The conditional probability distribution table with hepatorenal syndrome and electrolyte disorder as parent nodes.

together, The incidence of HE was 0.667. When the above two phenomena occur together, we should pay great attention to prevent the occurrence of HE.

Bayesian network evaluation. We used the area under the receiver operating characteristic (ROC) curve as the evaluation index. Figure 4 shows that the network's area under the ROC curve is 0.843. Thus, the network has high sensitivity and specificity and reflects the dependence among nodes.

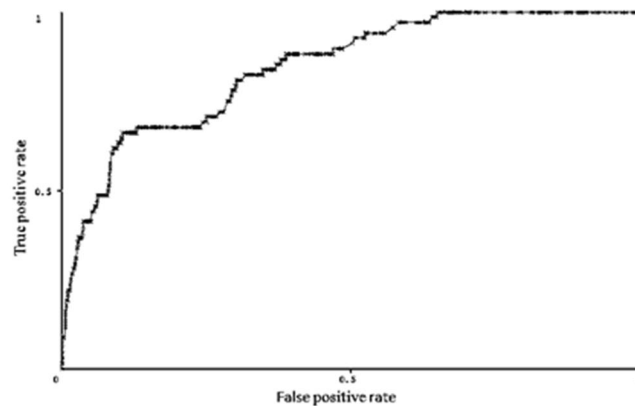


Figure 4. ROC curve of the Bayesian network of hepatic encephalopathy. The figure was plotted using Weka3.8.0 (<http://weka.wikispaces.com/>).

Reasoning model. BN reasoning refers to calculating the probability of an event in a given network structure, given the available evidence. Its biggest advantage is automatic updating of network probabilities according to different degrees of information mastery.

According to the patients' visiting process and the order in which information was obtained by the doctors, we used the BN to predict which patients would have concurrent HE. When a patient with cirrhosis does not attend a clinical examination, the likelihood of HE is 0.0736 by a priori information about the BN. If a clinical examination reveals that the patient is in poor spirits and has hepatic diabetes, then the BN shows that the likelihood that the patient has HE is 0.155 (Suppl. Fig. S1). If the patient's total bilirubin is greater than 51.3 mmol/L, or prothrombin time is longer than 18s, then the probability of concurrent HE increases to 0.341 (Suppl. Fig. S2). If the patient has any of the factors of infection, electrolyte disturbance, or hepatorenal syndrome, the probability of concurrent HE is more than 0.60: 0.615, 0.834, and 0.867, respectively (Suppl. Fig. S3, S4, and S5). Thus, infections, electrolyte disorders, and hepatorenal syndrome are closely associated with HE: if those three conditions are all present, the likelihood of concurrent HE is 0.968 (Suppl. Fig. S6).

Discussion

In the medical field, prediction of disease risk requires the establishment of a statistical model of risk factors and diseases and prediction of the probability of disease according to levels of multiple risk factors¹⁵. Our study's data are cross-sectional survey data that contain both causal and outcome variables of HE. Thus, it is not reasonable to use them to predict the risk of HE with a logistic regression model. Moreover, the use of a logistic regression model requires knowing all variables that predict disease risk⁹. From the perspective of economic rationality, doctors cannot require patients to undergo all tests that can exclude the disease at once; thus, logistic regression does not conform to clinical practice. BN reasoning can handle incomplete datasets and probabilistic reasoning based on the amount of information available to the patient. With increased amounts of patients' information collected, the network probabilities are updated constantly, and the likelihood of HE among patients with cirrhosis is evaluated according to different symptoms and signs. Therefore, the BN method is more suited to the clinical process of sequential diagnosis and treatment^{11,12,15}.

All clinical factors contribute to the occurrence and development of HE through complex network relationships^{5,30}. In this study, we used logistic regression with univariate and multivariate models to screen the main correlates of cirrhosis complicated with HE, and the BN structure was constructed to estimate the conditional probability of each node^{31–33}. The area under the ROC curve was 0.843, which indicated high sensitivity and specificity of the network construction of the tabu search algorithm. The BN can reflect the relationship between the various nodes so as to provide a basis for evaluation of influencing factors on HE.

Our results show that hepatorenal syndrome, electrolyte disorders, poor spirits, infections, prothrombin time, and total bilirubin are directly associated with HE. Hepatogenic diabetes and HE are linked through total bilirubin³⁴ (Fig. 2). Studies have suggested that impaired glucose tolerance is a key determinant of the prognosis of patients with cirrhosis. Abnormal blood glucose in patients with cirrhosis increases the risk of various complications and affects patients' prognosis. Therefore, hepatogenic diabetes is likely to cause HE indirectly by aggravating hepatic impairment⁵. Hepatorenal syndrome is also related to electrolyte disorder and infection. Mathurin et al. showed that the risk coefficient of HE increased with decreased serum sodium and serum cholinesterase levels and increased blood urea nitrogen levels in patients with liver cirrhosis and renal insufficiency³⁵. A large number of patients suffer from anorexia and ascites among those with liver cirrhosis. Salt intake should be strictly controlled as part of treatment. However, long-term administration of high-dose diuretics can lead to long-term low sodium, chloride, and potassium status. Severe imbalances disrupt electrolyte and acid-base balance in cells' internal and external fluids^{36,37}, which seriously affects their metabolism and physiological functions. This eventually leads to brain edema and induces HE³⁸. Some studies have shown that infection is related to HE^{39,40}; however, whether or not there is a causal relationship is still unclear. It is generally believed that when an organism becomes infected, catabolism increases; ammonia production increases; and liver, brain, kidney, and other solid organs are damaged. Hypoxia or hyperthermia increase the brain's sensitivity to ammonia toxicity and induce or aggravate HE^{38,41}. To a certain extent,

increased total bilirubin and prolonged prothrombin time may reflect the deterioration of liver function, and severe cases of liver cirrhosis may be complicated by HE³⁴.

The BN was used to induce the probability of HE in patients with cirrhosis. When a patient with cirrhosis has not been clinically examined, the relevant symptoms and indicators are unknown. The BN showed that the probability of concurrent HE in such cases was 0.0736. If the clinical examination found that the patient was apathetic and had hepatic diabetes, the likelihood of concurrent HE increased to 0.155; if the patient's total bilirubin was greater than 51.3 mmol/L and prothrombin time was longer than 18 s, the probability of concurrent HE increased to 0.341; if the patient showed any of the factors of infection, electrolyte disorders, or hepatorenal syndrome, the probability of concurrent HE exceeded 0.60 (i.e., 0.615, 0.834, and 0.867, respectively). Infection, electrolyte disorders, and hepatorenal syndrome are closely related to HE: if the patient had all three conditions, the probability of concurrent HE was 0.968. The results show that the BN constructed by the tabu search algorithm has a flexible reasoning mechanism, which is helpful for disease identification and diagnosis. It has important clinical diagnostic value and could serve as an effective method to reduce harm from HE.

The complex network of relationships between HE-related factors revealed by the BN is in line with clinical practice. By constructing the BN model, the clinical manifestations of patients with cirrhosis and the factors affecting their probability of occurrence of HE are used to identify the cirrhosis with HE, and the network probabilities can be updated as the doctor receives increasing amounts of patient information. Although this study lists only some of the key issues, it shows that compared with a logistic probability prediction model, the BN is more in line with the sequential process of clinical diagnosis and treatment. BNs have other advantages, including their ability to combine prior information and sample information. Subjective bias caused by the application of prior information and bias caused by using only sample information can be avoided⁴². BNs are also suitable for datasets with a variety of uncertain information and excessive numbers of variables¹¹: they facilitate easy understanding and explanation through visual images of the results. However, BNs also have shortcomings: for example, the edges between variables only represent probability dependencies and do not represent causal relationships. Therefore, we cannot perform causal analysis from only the BN's directed edges. It is necessary to combine various aspects of professional knowledge to determine causality⁴³. The data in this study are unbalanced, which may cause incorrect classification. In future research, we intend to deal with the impact of data imbalance on the classification model at the data and algorithm levels and hope to achieve better results. For example, we can expand the scope of data collection to increase the number of HE-positive cases and continuously verify and improve the accuracy and applicability of model classification identification. In addition, this study is based on a cross-sectional survey; thus, the BN can only reveal the relevant factors related to the occurrence of HE, and the causal relationship needs further verification. Only in a BN built making use of causal relationships do the directed edges between the variables represent causal relationships⁴⁴.

Data Availability

All data generated or analyzed during this study are included in its Supplementary Information files.

References

- Wei, L. & Liu, Y. Consensus on the diagnosis and treatment of hepatic encephalopathy in China (2013, Chongqing). *Chinese Journal of the Frontiers of Medical Science (Electronic Version)* **6**, 81–93 (2014).
- Bismuth, M., Funakoshi, N., Cadranet, J. F. & Blanc, P. Hepatic encephalopathy: from pathophysiology to therapeutic management. *European Journal of Gastroenterology & Hepatology* **23**, 8 (2011).
- Khungar, V. & Poordad, F. Management of overt hepatic encephalopathy. *Clinics in Liver Disease* **16**, 73–89 (2012).
- Jiang, H., Yao, D. & Yao, X. Cirrhosis with hepatic encephalopathy. *Chinese General Practice* **6**, 452–454 (2003).
- Kalaitzakis, E., Josefsson, A. & Bjornsson, E. Type and etiology of liver cirrhosis are not related to the presence of hepatic encephalopathy or health-related quality of life: a cross-sectional study. *BMC gastroenterology* **8**, 46, <https://doi.org/10.1186/1471-230X-8-46> (2008).
- Yang, Y., Zhang, Z., Shi, F., Li, M. & Zhao, L. Advances in research on treatment of mild hepatic encephalopathy with traditional Chinese and western medicine. *Clinical journal of Chinese medicine*, 142–144 (2014).
- Sesen, M. B., Nicholson, A. E., Banares-Alcantara, R., Kadir, T. & Brady, M. Bayesian networks for clinical decision support in lung cancer care. *PloS one* **8**, e82349, <https://doi.org/10.1371/journal.pone.0082349> (2013).
- Lee, S.-M. & Abbott, P. A. Bayesian networks for knowledge discovery in large datasets: basics for nurse researchers. *Journal of Biomedical Informatics* **36**, 389–399, <https://doi.org/10.1016/j.jbi.2003.09.022> (2003).
- Lee, S. M., Abbott, P. & Johantgen, M. Logistic regression and Bayesian networks to study outcomes using large data sets. *Nursing Research* **54**, 133 (2005).
- Garcia-Herrero, S., Mariscal, M. A., Gutierrez, J. M. & Ritzel, D. O. Using Bayesian networks to analyze occupational stress caused by work demands: preventing stress through social support. *Accident; analysis and prevention* **57**, 114–123, <https://doi.org/10.1016/j.aap.2013.04.009> (2013).
- Sperotto, A., Molina, J. L., Torresan, S., Critto, A. & Marcomini, A. Reviewing Bayesian Networks potentials for climate change impacts assessment and management: A multi-risk perspective. *Journal of environmental management* **202**, 320–331, <https://doi.org/10.1016/j.jenvman.2017.07.044> (2017).
- Wang, K. J., Makond, B. & Wang, K. M. Modeling and predicting the occurrence of brain metastasis from lung cancer by Bayesian network: a case study of Taiwan. *Computers in biology and medicine* **47**, 147–160, <https://doi.org/10.1016/j.combiomed.2014.02.002> (2014).
- Kumari, M., Vohra, R. & Arora, A. Prediction of Diabetes Using Bayesian Network. *International Journal of Computer Science & Information Technolo* (2014).
- Deeter, A., Dalman, M., Haddad, J. & Duan, Z. H. Inferring gene and protein interactions using PubMed citations and consensus Bayesian networks. *PloS one* **12**, e0186004, <https://doi.org/10.1371/journal.pone.0186004> (2017).
- Zhang, X., Yuan, Z., Ji, J., Li, H. & Xue, F. Network or regression-based methods for disease discrimination: a comparison study. *BMC Medical Research Methodology* **16**, <https://doi.org/10.1186/s12874-016-0207-2> (2016).
- Yazbek, M. A. *et al.* Association analysis of anti-Epstein-Barr nuclear antigen-1 antibodies, anti-cyclic citrullinated peptide antibodies, the shared epitope and smoking status in Brazilian patients with rheumatoid arthritis. *Clinics* **66**, 1401–1406, <https://doi.org/10.1590/s1807-59322011000800016> (2011).

17. Pan, J. *et al.* Prevalence of hyperlipidemia in Shanxi Province, China and application of Bayesian networks to analyse its related factors. *Scientific Reports* **8**, <https://doi.org/10.1038/s41598-018-22167-2> (2018).
18. Chen, R. & Herskovits, E. H. Clinical Diagnosis Based on Bayesian Classification of Functional Magnetic-Resonance Data. *Neuroinformatics* **5**, 178–188, <https://doi.org/10.1007/s12021-007-0007-2> (2007).
19. Alonso-Montesinos, J., Martínez-Durbán, M., del Sagrado, J., del Águila, I. M. & Batlles, F. J. The application of Bayesian network classifiers to cloud classification in satellite images. *Renewable Energy* **97**, 155–161, <https://doi.org/10.1016/j.renene.2016.05.066> (2016).
20. Alireza, S. H. A probabilistic bayesian classifier approach for breast cancer diagnosis and prognosis. (2012).
21. Agrahari, R. *et al.* Applications of Bayesian network models in predicting types of hematological malignancies. *Sci Rep* **8**, 6951, <https://doi.org/10.1038/s41598-018-24758-5> (2018).
22. Liao, Y., Xu, B., Wang, J. & Liu, X. A new method for assessing the risk of infectious disease outbreak. *Sci Rep* **7**, 40084 (2017).
23. Myte, R. *et al.* Untangling the role of one-carbon metabolism in colorectal cancer risk: a comprehensive Bayesian network analysis. *Scientific Reports* **7**, 43434 (2017).
24. Xu, S. *et al.* Cognition, quality-of-life and symptom clusters in breast cancer: using Bayesian networks to elucidate complex relationships. *Psycho-Oncology* **27** (2017).
25. Kumar, B. S. & Ch, V. Inverse modeling approach for evaluation of kinetic parameters of a biofilm reactor using tabu search. *Water Environment Research A Research Publication of the Water Environment Federation* **86**, 675–686 (2014).
26. Ferenci, P. *et al.* Hepatic encephalopathy—definition, nomenclature, diagnosis, and quantification: final report of the working party at the 11th World Congresses of Gastroenterology, Vienna, 1998. *Hepatology* **35**, 716–721, <https://doi.org/10.1053/jhep.2002.31250> (2002).
27. Hu, M. *et al.* Prevalence, awareness, treatment, and control of hypertension and associated risk factors among adults in Xi'an, China: A cross-sectional study. *Medicine* **95**, e4709 (2016).
28. Rao, H. *et al.* The higher prevalence of truncal obesity and diabetes in American than Chinese patients with chronic hepatitis C might contribute to more rapid progression to advanced liver disease. *Alimentary Pharmacology & Therapeutics* **46**, 731 (2017).
29. Battaglioli, T. *et al.* Cigarette smoking and alcohol consumption as determinants of survival in non-Hodgkin's lymphoma: a population-based study. *Annals of oncology: official journal of the European Society for Medical Oncology* **17**, 1283–1289, <https://doi.org/10.1093/annonc/mdl096> (2006).
30. Dhiman, R. K. & Chawla, Y. K. Minimal hepatic encephalopathy. *Indian Journal of Gastroenterology* **28**, 5–16 (2009).
31. Larrañaga, P., Sierra, B., Gallego, M. J., Michelena, M. J. & Picaza, J. M. Learning Bayesian Networks by Genetic Algorithms: A case study in the prediction of survival in malignant skin melanoma. **1211**, 261–272 <https://doi.org/10.1007/BFb0029459> (1997).
32. Ghosh, J. K. & Valtorta, M. In *Southeast Regional Conference, 2000, Clemson, South Carolina, Usa, April*. 239–240.
33. Scutari, M. Learning Bayesian Networks with the bnlearn R Package. *Journal of Statistical Software* **35**, 2010 (2010).
34. Takikawa, Y., Endo, R., Suzuki, K., Tsubouchi, H. & Fulminant Hepatitis Study Group of, J. Early prediction of short-term development of hepatic encephalopathy in patients with acute liver disease unrelated to paracetamol. A prospective study in Japan. *Journal of hepatology* **51**, 1021–1029, <https://doi.org/10.1016/j.jhep.2009.09.011> (2009).
35. Mathurin, S. *et al.* [Renal failure in patients with cirrhosis and ascites: incidence, etiology and predictive factors]. *Acta Gastroenterologica Latinoamericana* **38**, 116 (2008).
36. Cordoba, J. & Minguez, B. Hepatic encephalopathy. *Seminars in liver disease* **28**, 70–80, <https://doi.org/10.1055/s-2008-1040322> (2008).
37. Shaikh, S., Mal, G., Khalid, S., Baloch, G. H. & Akbar, Y. Frequency of hyponatraemia and its influence on liver cirrhosis-related complications. *Journal of the Pakistan Medical Association* **60**, 116–120 (2010).
38. Rai, R. *et al.* Reversal of Low-Grade Cerebral Edema After Lactulose/Rifaximin Therapy in Patients with Cirrhosis and Minimal Hepatic Encephalopathy. *Clinical and translational gastroenterology* **6**, e111, <https://doi.org/10.1038/ctg.2015.38> (2015).
39. Vaquero, J. *et al.* Infection and the progression of hepatic encephalopathy in acute liver failure. *Gastroenterology* **125**, 755–764, [https://doi.org/10.1016/s0016-5085\(03\)01051-5](https://doi.org/10.1016/s0016-5085(03)01051-5) (2003).
40. Xie, Y. *et al.* Bacterial distributions and prognosis of bloodstream infections in patients with liver cirrhosis. *Sci Rep* **7**, 11482, <https://doi.org/10.1038/s41598-017-11587-1> (2017).
41. Pozdeev, V. I. *et al.* TNF α induced up-regulation of Na(+), K(+), 2Cl(-) cotransporter NKCC1 in hepatic ammonia clearance and cerebral ammonia toxicity. *Sci Rep* **7**, 7938, <https://doi.org/10.1038/s41598-017-07640-8> (2017).
42. Rubin, D. L., Burnside, E. S. & Shachter, R. *A Bayesian Network to Assist Mammography Interpretation*. (Springer US, 2005).
43. Scutari, M., Auconi, P., Caldarelli, G. & Franchi, L. Bayesian Networks Analysis of Malocclusion Data. *Sci Rep* **7**, 15236, <https://doi.org/10.1038/s41598-017-15293-w> (2017).
44. S, G.-H., MA, M., JM, G. & DO, R. Using Bayesian networks to analyze occupational stress caused by work demands: preventing stress through social support. *Accident Analysis and Prevention* **57**, 114–123 (2013).

Author Contributions

Z.Z., J.Z. and Z.W. conceived the study; Z.Z., J.Z., H.R., W.M.S. and J.C.L. conducted the survey and collected data; J.Z. and Z.W. processed the data; Z.Z., J.H.P. and Y.B.Z. analyzed the results; Z.Z. and Z.W. wrote the main manuscript text; L.X.Q. gave constructive suggestions for the manuscript. All authors revised the manuscript for important intellectual content and approved the final version.

Additional Information

Supplementary information accompanies this paper at <https://doi.org/10.1038/s41598-019-42791-w>.

Competing Interests: The authors declare no competing interests.

Publisher's note: Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2019