# Comparison of Belief Network Classifier and Rule-based System for Diagnosis of Liver Diseases

Mi Swe Zar Thu; Dr. Soe Hay Mar
*University of Computer Studies, (Hpa-an)*
*miswezarthu@gmail.com*

## Abstract

*Computer base methods are increasingly used to improve the quality of medical services. Expert system uses knowledge, facts and reasoning techniques to solve problems that normally require the expertise, experiences and the abilities of human experts. This paper presents the comparison of a belief network classifier and rule-based expert system for the liver diseases. Bayesian Belief Networks provide a mathematically correct and therefore more accurate method of measuring the effects of events on each other. Belief offers an approach for dealing with uncertain information in knowledge-based (expert) systems. The theory of belief networks is mathematically sound, based on techniques from probability theory. CN2 Rule is used to implement the rule-based expert system for the comparison process. The system intends to automatically create and maintain knowledge for an intelligence assistant system of various domains. Case study used in this system is, identifying liver disorder (cirrhosis and heptocellular carcinoma).Experimental Results show that Bayesian Belief Network approach outperforms over other classification algorithm.*

## 1. Introduction

Many tasks including diagnosis, pattern recognition and forecasting can be viewed as classification, as each requires identifying the class labels for instances (attributes). There are many algorithms for extracting rules and bayesian (probability-based) classifiers. Bayesian networks are powerful for knowledge representation and inference under condition of uncertainty.

In diagnostic expert systems, knowledge from a given domain is typically represented in the form of rules. To express uncertainty in the domain, each conclusion of a rule is associated with a measure of confidence in its correctness. This paper presents the comparison of a rule based expert system and belief network classifier which aims at supporting the clinician in the initial assessment of the patient with a disorder of the liver diseases. Bayesian networks are acyclic directed graphs modeling probabilistic dependencies among variables. The graphical part of a Bayesian network reflects the structure of a problem. CN2 rules are designed for the efficient induction of simple, comprehensible production rules in domains. It can be used to generate the knowledge, usually in the form of IF-THEN rules [5].

The organization of this paper is as follows: section 2 describes the related work of the proposed system. Section 3 and section 4 presents background theory used in this system, belief network and rule induction classifier. In section 5, this paper presents the proposed system design. Section 6 reflects the system implementation and experimental results. Section 7 is the conclusion and future works of the proposed system.

## 2. Related Work

This system presents the classification of liver diseases especially for cirrhosis and heptocellular carcinoma. There have been several classifier algorithms developed to solve the medical decision support system.

In Lucus and Janssens presents a rule-based expert system; it aims at supporting the clinician in the initial assessment of the patient with a disorder of the liver or biliary tract. The system incorporates the certainty-factor model to deal with uncertain medical knowledge. Although its diagnostic performance has been shown to be quite reasonable one of the problems with the HEPAR system is that its conclusions may be difficult to interpret, due to the unclear meaning of certainty factors [3,4].

In contrast with the heuristic models for reasoning with uncertainty, such as the certainty-factor model, the recent theory of belief networks is based on mathematically sound techniques, derived straight from probability theory A belief network permits the qualitative representation of dependencies and independencies among statistical variables. The theory of belief networks then permits the use of this causal knowledge for diagnostic problem-solving. Such a causal model of a medical domain may be easier to comprehend by medical students and unexperienced clinicians than a similar diagnostic rule-based system. Furthermore, a belief network designed for diagnostic problem-solving can be applied to predict findings associated with (groups of) disorders. This paper presents the comparison for diagnostic rule-based expert system and belief network classifier for liver diseases. The experimental results show that the accuracy of diagnostic system using belief network is far better than CN2 rule-based classifier algorithm.

## 3. Bayesian Belief Network

Bayesian Belief Networks are an emerging modeling approach of artificial intelligence (AI) research that aim to provide a decision-support framework for problems involving uncertainty, complexity and probabilistic reasoning [6]. It is based on conceptualising a model domain of interest as a graph (i.e. network) of connected nodes and linkages. In the graph, nodes represent important domain variables, and a link from one node to another represents a dependency relationship between the corresponding variables. To provide quantitative description of the dependency links, Bayesian Belief Networks (BBNs) utilise probabilistic relations, rather than deterministic expressions. For example, there is a network consisting of five variables (nodes) A,B,C,D,E. All the variables are dependent on (i.e. influence) each other for the BBN as in Figure 1.

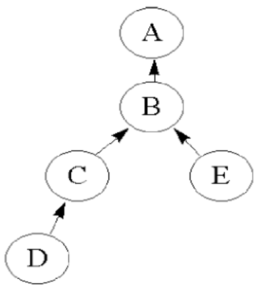


**Figure 1 Example Node-link structure for BBN**

The chain rule from probability theory to calculate the joint distribution p(A,B,C,D,E) is as follow:

p(A,B,C,D,E) = p(A|B,C,D,E) * p(B|C,D,E) * p(C|D,E) * p(D|E) * p(E)

Suppose the set of variables in a BBN is {A1 ,A2 ,…,An} and that parents(Ai ) denotes the set of parents of the node Ai in the BBN. Then the joint probability distribution for {A1 ,A2 ,…,An} is:

$$p(A_1, \ldots\ldots A_n) = \prod_{i=1}^{n} p(A_i \mid parents(A_i))$$

A belief network consists of two components –

- A qualitative representation of the variables and relationships between the variables discerned in the domain, expressed by means of a directed acyclic graph G = (V (G), A(G)), where V (G) = (V1, V2, …, Vn) is a set of vertices, taken as the variables, and A(G) a set of arcs (Vi, Vj ), where Vi, Vj $\in$ V (G), taken as the relationships between the variables.

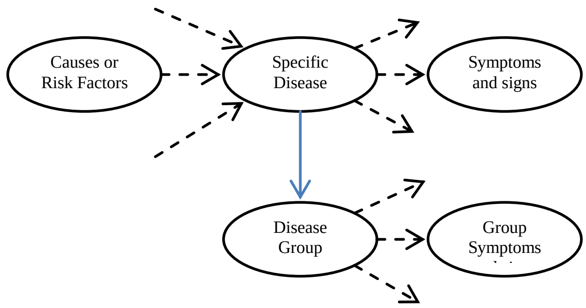


**Figure 2 Design of Belief Network for Diagnosis**

- A quantitative representation of the `strengths' of the relationships between the variables, expressed by means of assessment functions.

After a belief network has been specified, it can be used for reasoning about uncertain information in the domain concerned. The process of computing the updated probability distribution after entering specific evidence into the belief network is called evidence propagation [1]. Figure 2 presents the design of Belief Network for diagnosis.

## 4. CN2 Rule Induction

The CN2 algorithm inductively learns a set of preposition if/then rules from a set of training examples [2]. To do so, it performs a general to specific beam search through rule-space for the "best" rule; then it removes the training example covered by that rule and repeats the previous two steps until no more good rules can be found. This algorithm can be summarized as follows:

**Repeat**

- start with the general rule:"everything à <class>"
- specialize the rule;
- retain the more significant disjunctive term;

**Until** no more rules to find.

The CN2 algorithm consists of two main procedures: a control algorithm for repeatedly executing the search and a search algorithm performing a beam search for a good rule.

## 5. System Architecture

This paper presents the diagnosis of patients with disorders of the liver. This system mainly focused in the diagnosis of following liver diseases

- Hepatocellular Carcinoma
- Cirrhosis

Figure 3 presents the architecture of the system.

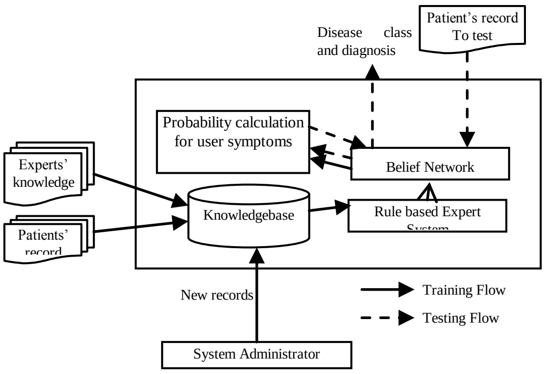


**Figure 3 System Architecture**

Comparison of rule-based expert system and belief network classifier is implemented. Since a relationship between two variables implies their statistical dependence, a belief network can be used to represent medical knowledge. Such a model of a medical domain may be easier to comprehend by medical students and unexperienced clinicians than a similar diagnostic rule-based system. Rule induction allows for the implementation of reasoning models, and conclusion is drawn from its facts. First this system is trained using training data from patients' records using Belief Network. Then disease class and diagnosis can be obtained by applying patient's symptoms into the system.

## 5.1. Process Flow of the System

Process flow of the System is as follows:

- There must be two kinds of user, Admin user and normal user (clinicians or patients)
- When normal user sets patient's attribute into the system, there is two kinds of process.
- The first one is processing by rule based algorithm; rules are extracted from knowledge base by CN2 rule induction method.
- Then user attributes are checked with IF THEN rules and result is generated.
- The other one is processing by Belief Network; BBN consists of two process, preparing BBM Graph and conditional table.
- Then probability for each class is computed and class with maximum probability is generated as the final results
- Finally, accuracy of both processes has been computed and there is performance analysis over those two algorithms.

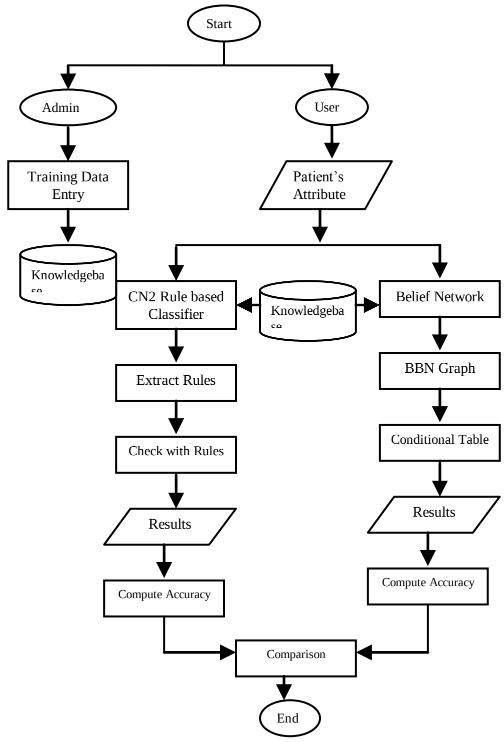


**Figure 4 Process flow of the System**

## 6. System Implementation

This system is implemented as Windows based diagnosis system, developed using Java programming language. Both rule based and BBN algorithms are implemented in this system. Each system has two main process; Training and Testing. Then Diagnosis tab for classifying user symptoms into specific kind of disease. There are 21 attributes used in this system as in Table 1.

**Table 1 Attributes for Classifier**

| Attribute | Values |
|---|---|
| Class | Carcinoma, Cirrhosis, None |
| Age | 10-19, 20-29, 30-39, 40-49, 50-59, 60-69, 70-79, 80-89, 90-99. |
| AlkoholAbuse | Always, No, Often, Sometimes |
| AlkalinePhospate | < 3 x normal, > 3 x normal, normal |
| Antivirals | No, Yes |
| Ascities | No, Yes |
| BleedingEasily | No, Yes |
| BloodInStool | No, Yes |
| DarkUrine | No, Yes |
| Diaherra | No, Yes |
| DistensionAbdomen | No, Yes |
| Edema | No, Yes |

| | |
|---|---|
| Fatigue | No, Yes |
| Headache | No, Yes |
| IntestinalBleeding | No, Yes |
| Itching | No, Yes |
| Jaundice | No, Yes |
| LossOfApetite | No, Yes |
| Malaise | No, Yes |
| MuscleJoints | No, Yes |
| NeedSleep | No, Yes |
| NFLD | No, Yes |
| PortalHypertension | No, Yes |
| SerumAlbumin | No, Yes |
| Sex | Female, Male |
| Vomit | No, Yes |
| WeightLoss | No, Yes |

### 6.1. Accuracy

Estimating classifier accuracy is important since it determines to evaluate how accurately a given classifier will label future data, data on which the classifier has not been trained. Accuracy estimates also help in the comparison of different classifiers. A test set is used to determine the accuracy of the model. Usually, the given data set is divided into training and test sets, with training set used to build the model and test set used to validate it.

**Accuracy ratio**: the percentage of test set samples that are correctly classified by the model

The Sensitivity and specificity measures can be used to determine the accuracy measures. Precision may also be used to access the percentage of samples labeled as for example, “cancer” that actually are “cancer” samples. These measures are defined as:

**Sensitivity = tpos / pos**

**Specificity = tneg / neg**

**Precision = tpos / (tpos + fpos)**

Where,

tpos = the number of true positives (“cancer” samples that were correctly classified as such),

pos = the number of positive (“cancer”) samples

tneg = the number of true negative (“not cancer” samples that were correctly classified as such)

neg = the number of negative samples

fpos = number of false positive (“not_cancer” samples that were incorrectly labeled as cancer)

$$\textbf{accuracy} = \textbf{sensitivity}\,\frac{\textbf{pos}}{(\textbf{pos} + \textbf{neg})} + \textbf{specificity}\,\frac{\textbf{neg}}{(\textbf{pos} + \textbf{neg})}$$

### 6.2. Experimental Results

It is tested with various training data size. In this system, two third of samples is used as training data and the rest is used as testing data. Table 2 shows the results of this system.

**Table 2 Classifier accuracy for BBN and CN2**

| | Samples | BBN | CN2 |
|---|---|---|---|
| **Training Accuracy** | 400 | 93.56 % | 56.20 % |
| **Testing Accuracy** | 200 | 93.27 % | 56.35 % |
| **Training Accuracy** | 713 | 92.87 % | 60.04 % |
| **Testing Accuracy** | 350 | 90.40 % | 60.33 % |

BBN can solve the problems of uncertainty, complexity and probabilistic reasoning. In the domain of liver diseases, one of the problems with rule based system is that its conclusions may be difficult to interpret, due to the unclear meaning of certainty factors.

## 7. Conclusion

This system presents the comparison for Bayesian Belief Network and CN2 rule induction algorithms. Belief networks are efficient system for uncertainty data based on probability theory of mathematics. Furthermore, a belief network designed for diagnostic problem-solving can be applied to predict findings associated with (groups of) disorders. This system can support medical treatment about liver diseases. According to the experimental results BBN performs significantly better than CN2 rule induction algorithm.

## 9. References


[1] J. Cheng and R. Greiner, "Learning Bayesian Belief Network Classifiers: Algorithms and System", IBM Toronto Lab, Canada, 1998.

[2] P. Clark and T. Niblett, "The CN2 Induction Algorithm", In: Machine Learning Journal, 3 (4), pp261-283, Netherlands: Kluwer (1989)

[3] P.J.F. Lucas and A.R. Janssens, Development and validation of HEPAR, an expert system for the diagnosis of disorders of the liver and biliary tract, Journal of Medical Informatics 16 (1991) 259-270.

[4] P.J.F. Lucas, R.W. Segaar and A.R. Janssens, HEPAR, an expert system for the diagnosis of disorders of the liver and biliary tract, Liver 9 (1989) 266-275.

[5] G. L. Pappa and A. Freitas, "Automatically Evolving Rule Induction Algorithms", Computing Laboratory, University of Kent, 2007.

[6] B. Zheng, Y. Chang, X. Wang and W. F. Good, "Comparison of Artificial Neural Network and Bayesian Belief Network in a Computer-Assisted Diagnosis Scheme for Mammography", in the proceeding of IEEE, 1999.