Scientific Journal of Informatics
Vol. 5, No. 1, May 2018
p-ISSN 2407-7658 http://journal.unnes.ac.id/nju/index.php/sji e-ISSN 2460-0040

Implementation of Decision Tree and Dempster Shafer on Expert System for Lung Disease Diagnosis

Abdul Muis Alfatah[1], Riza Arifudin[2], Much Aziz Muslim[3]

[1,2,3]Computer Science Department, FMIPA, Universitas Negeri Semarang
Email: [1]abdulmuis297@gmail.com, [2]riza.arifudin@gmail.com, [3]a212muslim@yahoo.com

Abstract

The expert system is a computer system that contains set of rules to solve problems like an expert. The lungs are one of the vulnerable respiratory organs. The purpose of this research is to implement decision tree and dempster shafer method on lung disease diagnosis and measure the accuracy of the system. The symptom was searched using forward chaining decision tree and the diagnosis was calculated using dempster shafer method. Dempster Shafer method calculates the possibility of a lung disease based on the density of probability value that possessed by each symptom. This research used 65 data obtained from medical record of Puskesmas Tegowanu Grobogan Regency. General symptoms and types of disease are used as a variable. Based on the results of the study, it can be concluded that the results of the diagnosis using dempster shafer method has an 83.08% accuracy.

Keywords: Expert System, Lung Disease, Decision Tree, Dempster Shafer.

1. INTRODUCTION

Expert systems are part of the high-level software or high-level programming language which attempt to duplicate the expert functionality in a particular area of expertise [1]. It can be used to overcome multiple problems by giving some advice like an expert knowledge [2]. In the application, the problem that needs to be solved is not only the algorithm but also the expert-field problem that is difficult to understand [3]. There are a knowledge base and an inference system (set of rules) [4]. In other words, the expert system is software-based systems that create or evaluate decisions based on rules defined in the software [5]. The purpose of the expert system is not to replace human role, but to represent a human knowledge into a system form, so that it can be used by many people [6]. From some previous studies, expert system provides good results for solving cases that use complex data, such as skin disease diagnosis, pregnancy illness diagnosis, asset damage analysis and digestive disease diagnosis [7-10].

The Dempster Shafer theory was introduced by Dempster (1967) and expanded by Shafer (1976). Dempster Shafer is a representation, combination, and propagation of uncertainty that has some institutive characteristic according to the way the expert thinks, but it has a strong mathematical basis [11]. This theory can be interpreted as a general form of probability theory, which is used for the dataset, not for single data [12]. In addition, this theory aims to represent and overcome the uncertainty of information. The most important about dempster shafer is the ability to combine different data sources to improve the quality of information [13].

Someone who is suffering from a disease with cough symptoms would need to consult with a doctor so that he immediately know the illness that he suffered [14]. Lung disease is a disease associated with the respiratory system in humans, can be bad if not treated immediately seriously. The lungs served as a place of exchange of oxygen needed by humans and release carbon dioxide which is the result of the rest of the respiratory process that must be removed from the body so that the body's need for oxygen remains fulfilled [15]. A doctor can diagnose the disease by analyzing the patient symptoms. But, the presence of technological advances makes it can be diagnosed more quickly by using an expert system [16].

Based on the description, the purpose of this research is to implement decision tree and dempster shafer method in Expert System Diagnosis of Lung Disease and measure the accuracy of that.

2. METHODS

2.1. Decision Tree

The decision tree is a classification method that uses a representation of a tree structure, each node represents an attribute, a branch represents the value of an attribute, and the leaf represents the class. The top node of the decision tree is referred as a root [17]. The decision tree is decision support tools that use hierarchical tree structures to classify the classes based on a set of questions [8]. The decision tree consists 3 types of nodes:

1. Decision Node: generally represented by a box.
2. Node of opportunity: generally represented by a circle.
3. End node: generally represented by a triangle.

Decision tree as a decision support tools can provide effective decisions because they have several advantages such as:

a) Easy to understand and interpret.
b) Has a value even with small amount of data.
c) Can be combined with other decision-making techniques.
d) Exposing all problems so that all possibilities can be classified.
e) Allows to analyze in making decisions about the possibilities of some alternatives
f) Provide a framework for measuring the result of value and probability of reaching a decision.
g) Help to make the best decisions based on available information.

2.2. Dempster Shafer

Dempster shafer is a mathematical theory for looking evidence based on belief function and plausible reasoning, used to combine separate information (evidence) and calculate the probability of an event. In other words, Dempster shafer is a mathematical theory for facts [12]. There are various kinds of reasoning with complete and consistent models, but in the reality, there are many problems that can't be resolved completely and consistently. The inconsistency is due to the addition of new facts. This is called as nonmonotonic reasoning. Dempster shafer is able to overcome this inconsistency [18].

In the expert system, a disease has a number of evidences that will be used in the uncertainty factor in making a decision to diagnose a disease. To overcome a number of evidences, dempster shafer using a rule that known as Dempster's rule of combination.

$$m1 \oplus m2(Z) = \frac{\sum_{X \cap Y = Z} m1(X)m2(Y)}{1-\sum_{X \cap Y = \theta} m1(X)m2(Y)} \quad (1)$$

Where as:

m1 ⊕ m2(Z)	= *mass function* of *evidence* (Z)
m1(X)	= *mass function* of *evidence* (X)
m2(Y)	= *mass function* of *evidence* (Y)
⊕	= *direct sum operator*

2.3. Process of Planning

The steps of developing an expert system diagnosis of lung disease are shown in Figure 1.

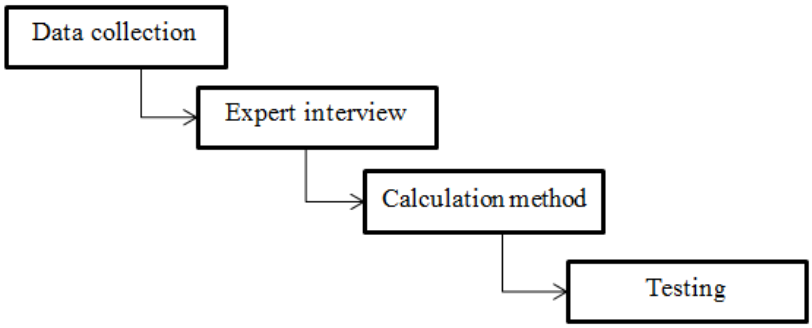

Figure 1. The Steps of Developing An Expert System

2.4. System of Design

The design of this expert system using waterfall model. A waterfall is an approach based on the assumption that major decisions must be made before encoding begins [19]. This model is often used by systems analyst [20]. There are four stages in the waterfall model [21].

1) requirement analysis stage is defining the entire software format, identifying all the needs and outlines of the system[22].
2) The design stage is designing application includes the interface design and database structure design [23].To create an interesting web-based application program (website) it must be designed beforehand, so the achieved results are suitable with the predetermined objectives [24].
3) The implementation stage is designing software which realized as a series of program or program unit [25].
4) The testing stage is to test whether the system is ready and feasible to use. The

tester can define the set of input conditions and perform testing on functional specifications of the program [26].

3. RESULT AND DISCUSSION

3.1. Data Collecting

The data used in this study is the result of medical records of 65 patients in 2016 who suffered from lung disease including Tuberculosis, Asthma, Bronchitis, and Pneumonia. This data are taken from Puskesmas Tegowanu Grobogan. This data contains the symptoms and doctor diagnosis results of patients in Puskesmas Tegowanu.

3.2. Interview

The interview was conducted with a Tegowanu Public Health Center internal medicine specialist. This interview gives the researcher a set of weight value from each symptom in lung disease shown in Table 1.

Table 1. Weight Value Symptoms

No	The Symptoms	P01	P02	P03	P04
		TB	AS	BR	PN
1	A cough with phlegm	*		*	*
2	A cough with phlegm with blood	*			
3	Asphyxiate		*	*	*
4	The body feels weak	*		*	
5		*			
6	A cough> 3 weeks	*			*
7	weight loss	*			
8	Fever in the afternoon and evening	*			
9	Often suffer from respiratory infections (Flu)			*	
10	A cough persists and recurs		*		
11	shoulder/back pain			*	
12	A cough is heavy at night		*		
13	Night Sweats	*			
14	Impaired vision			*	
15	Shortness of breath accompanied by wheezing sounds		*		
16	excessive mucus		*		
17	excessive anxiety		*		
18	Symptoms arise at night		*		
19	Shivering				*
20	chest pain	*		*	*
21	High fever				*
22	Breath and pulse beating faster		*		*
23	Fatigue and lethargy			*	*
24	pale appearance	*			*
25	Insomnia		*		*
26	decreased consciousness		*		*

Where as:
P01: Tuberculosis
P02: Asthma
P03: Bronchitis
P04: Pneumonia

3.3. Method of Calculation

the value of symptoms is a measure of disease probability, this value is determined through information from the literature and expert experience, by using probability rules. Of the four diseases, cough with phlegm has the possibility of tuberculosis (P01), Bronchitis (P03), pneumonia (P04) then the calculation of probability symptoms as follows.

Symptom 1 (Cough with phlegm)

$m1$(P01, P03, P04) = 0,16

$m1(\theta)$ = 1 – 0,16= 0,84

then calculated by the formula dempster's rule of combination. after the diagnosis, there is a modification of the final diagnosis of the formula that is 1- the highest value of the final diagnosis.

3.4. Testing

After testing 65 medical record data of lung disease patients at Puskesmas Tegowanu Grobogan, the result obtained accuracy of 83,08% as seen in Table 2.

Table 2. The Result of Data Testing

Amount of data	Matched data	Unmatched data	Accuracy
65	54	11	83,08%

From the data shown in table 2, there are 54 matched data according to doctor diagnosis from 65 cases. The accuracy of dempster safer method compared with doctor diagnosis as follows.

Amount of data = 65
Matched data = 54
Unmatched data = 11

$$\text{Accuracy} = \frac{\text{Amount of data} - \text{total unmathed data}}{\text{Amount of data}} x100\%$$

$$= \frac{65-11}{65} x\ 100\%$$

= 83,08%

The testing of medical record data using a system provides 54 accurate data out of 65 data. The rest 11 data is less accurate due to the system with the dempster shafer method uses a strong calculation, so it doesn't easily influenced by other factors. While the specialist in internal medicine as an expert has other considerations in determining the disease suffered by the patient.

3.5. Implementation

A consultation page is a page that users can use to consult. The user can consult with the system without login. In conducting the consultation, the user fills in the name and any symptoms experienced. Figure 2 shows the consultation page.

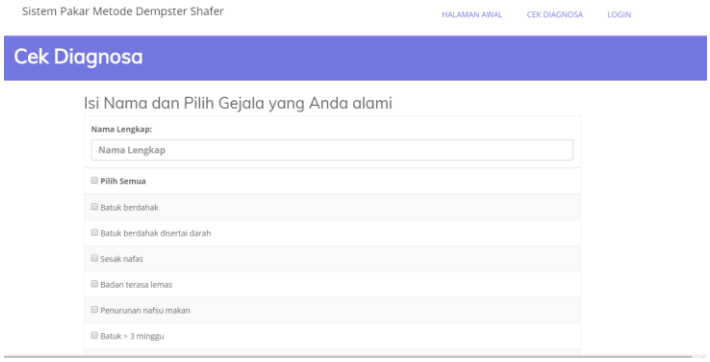

Figure 2. Consultation Page

The results page of the consultation is a page that provides diagnostic information from the input symptoms. This page shows the patient symptoms experienced, the treatment advice and his diagnosis. Figure 3 shows the results page of the consultation.

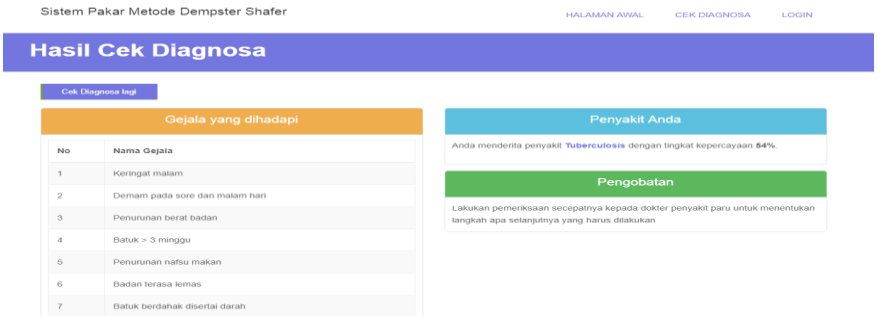

Figure 3. Admin and Expert Diagnosis Result Page

4. CONCLUSION

Dempster shafer method and decision tree is a method to calculate the uncertainty of a problem, this uncertainty is due to the addition of new facts. Dempster shafer can optimize the diagnosis that is produced because the system is not only based on the rule but has a value. The level of trust is more accurately supported by the decision tree as a supporter in disease prediction. Dempster shafer works by using the density value or weight value of each known fact, derived from an expert. The weights of these factare combined to produce a combination of known factual density values. The end result is a fact or a combination of facts with the greatest weight value. After

implementing 65 medical record data taken from Puskesmas Tegowanu using dempster shafer method and decision tree, it can be seen that the accuracy of dempster shafer 83,08%, with 54 data produce the same diagnosis with the doctor.

5. REFFERENCES

[1] Hustinawaty, R. A. (2014). The Development of Web Based Expert System for Diagnosing Children Disease Using Php And Mysql. International Journal of Computer Trends and Technology (IJCTT), 10(4), 197-202.

[2] Naik, V. M., & Lokhanday, S. (2012). Building a Legal Expert System for Legal Reasoning in Specific Domain-A Survey. International Journal of Computer Science & Information Technology, 4(5), 175.

[3] Setyabudi, W.U., Sugiharti E., & Arini F.Y. 2017. Expert System Diagnosis Dental Disease Using Certainty Factor Method. *Scientific Journal of Informatics*, 4(1): 44.

[4] Pramesti, A.A., Arifudin R., & Sugiharti E. 2016. Expert System for Determination of Type Lenses Glasses using Forward Chaining Method. *Scientific Journal of Informatics*, 3(2): 177.

[5] Josephine, M.S. & Jeyabalaraja V. 2012. Expert System and Knowledge Management for Software Developer in Software Companies. *International Journal of Information and Communication Technology Research*, 3(2): 243.

[6] Muslim, M.A., Kurniawati I., & Sugiharti E. 2015. Expert System Diagnosis Chronic Kidney Disease Based on Mamdani Fuzzy Inference System. *Journal of Theoretical and Applied Information Technology*, 78(1): 70.

[7] Maseleno, A. & Md. M. Hasan. 2012. Skin Diseases Expert System using Dempster Shafer Theory. *Journal Intelligent System and Applications*, 5: 38-44.

[8] Minardi, J. & Suyatno. 2016. Sistem Pakar Untuk Diagnosa Penyakit Kehamilan Menggunakan Metode Dempster Shafer dan Decisiona Tree. *Jurnal SIMETRIS*, 7(1).

[9] Purwanto, A. 2015. Analisa dan Perancangan Sistem Pakar Kerusakan pada Aset UKM STIKOM Bali Menggunakan Metode Dempster Shafer. *Konferensi Nasional Sistem & Informatika STMIK STIKOM 2015.*

[10] Istiqomah, N.Y & Fadlil, A. 2013. Sistem Pakar Untuk Mendiagnosa Penyakit Saluran Pencernaan Menggunakan Metode Dempster Shafer. *Jurnal Sarjana Teknik Informatika*, 1(1): 32-41.

[11] Sinaga, D.M., & Sari, N. 2016. Penerapan Metode Dempster Shafer Untuk Mendiagnosa Penyakit Dari Akibat Bakteri Salmonella. *Cogito Smart Journal*, 2(2): 96-98.

[12] Chowdury &Shakhawat. 2013. Fusing Probability Density Function into Dempster-Shafer Theory of Evidence for The Evaluation of Water Treatment Plant. *Environt Monit Assess* (2013) 185:3917-3929.

[13] Chaabane & Ben, S. 2011. A New Method for the Estimation of Mass Function in the Dempster-Shafer's Evidence Theory: Application to Colour Image Segmentation. *Circuits Syst Signal Process* (2011) 30: 55-71.

[14] Rahmawati, E & Wibawanto, H. (2016). Sistem Pakar Diagnosis Penyakit Paru-Paru Menggunakan Metode Forward Chaining. *Jurnal Teknik Elektro*, 8(2): 64.

[15] Minarni & Novriani, V. 2014. Rekayasa Perangkat Lunak Diagnosa Penyakit Paru-Paru Menggunakan Metode Backward Chaining Berbasis Web. *Jurnal*

Teknologi Informasi & Pendidikan, 7(2) : 84-85.
[16] Handarko, J.L., & Alamsyah A. (2015). Implementasi Fuzzy Decision Tree untuk Mendiagnosa Penyakit Hepatitis. *UNNES Journal of Mathematics*, 4(2): 157-164.
[17] Gorunescu, F. (2011). Data Mining Concepts, Models and Tehniques. *Intelligent Systems Reference Library*, Volume 12. Springer-Verlag Berlin Heidelberg.
[18] Rikhiana, D.E. & Fadlil, A. (2013). Implementasi Sistem Pakar Untuk Mendiagnosa Penyakit Dalam Pada Manusia Menggunakan Metode Demspter Shafer. *Jurnal Sarjana Teknik Informatika*, 1(1): 1-10.
[19] Patel, U.A., & Jain N.K. (2013). New Idea in Waterfall Model for Real Time Software Development. *International Journal of Engineering Research & Technology (IJERT)*, 2(4): 115.
[20] Roviaji, R., & Muslim, M. A. (2017, March). Pembuatan Sistem Informasi Gardu Induk PT. PLN (Persero) App Semarang Se-Kota Semarang dengan Java Android. In *Prosiding SAKTI (Seminar Ilmu Komputer dan Teknologi Informasi)* 2(1): 182-185.
[21] Pressman, R.S. (2001). *Software Engineering*. Online. Tersedia di http://www.resource.mitfiles.com/ [diakses 18-2-2017].
[22] Nugroho, Z. A., & Arifudin, R. (2015). Sistem Informasi Tracer Study Alumni Universitas Negeri Semarang Dengan Aplikasi Digital Maps. *Scientific Journal of Informatics*, 1(2), 153-160.
[23] Putra, A. T. (2015). Pengembangan E-Lecture menggunakan Web Service Sikadu untuk Mendukung Perkuliahan di Universitas Negeri Semarang. Scientific Journal of Informatics, 1(2), 168-176.
[24] Muslim, M. A. (2012). Pengembangan Sistem Informasi Jurusan Berbasis Web Untuk Meningkatkan Pelayanan Dan Akses Informasi. *Jurnal MIPA*, 35(1).
[25] Purwinarko, A. 2014. Model Expertise Management System di Universitas Negeri Semarang. *Scientific Journal of Informatics*, 1(2): 177-184
[26] Mustaqbal, M. S., Firdaus, R. F., & Rahmadi, H. (2016). Pengujian Aplikasi Menggunakan Black Box Testing Boundary Value Analysis (Studi Kasus: Aplikasi Prediksi Kelulusan SMNPTN). *Jurnal Ilmiah Teknologi Informasi Terapan*, 1(3): 31 -36.