TELKOMNIKA Indonesian Journal of Electrical Engineering
Vol.12, No.7, July 2014, pp. 5669 ~ 5677
DOI: 10.11591/telkomnika.v12i7.5723

Hypertension Expert System with C5.0 Algorithm and Fuzzy Logic

I Gusti Made Ngurah Ardi Yasa*[1], I Ketut Gede Darma Putra[2], Ni Md Ika Marini Mandenni[3]
Department of Information Technology, Udayana University, Bali, Indonesia
Bukit Jimbaran, Badung, Bali, Indonesia, Telp. 0361-7853533
*Corresponidng author, e-mail: ngurah.ardiyasa0961@gmail.com[1], ikgdarmaputra@gmail.com[2], ika_made@yahoo.com[3]

Abstract

Expert System is a knowledge-based system which is its knowledge is coming from the experts itself. This system is expected to help users to take best decisions in solving the problems they face. This expert system can be applied in various fields of knowledges. Medical field is one area where expert system is needed in case of taking decision, such as diagnosing and treatment.Hypertension is one of medical disease that can be diagnosed from looking at the patient's physical characteristics and the patient's lifestyle. Beside using the experts knowledge in diagnosing process, we also combine it with C5.0 algorithms and fuzzy logic to get precise result. C5.0 algorithm is used to create a decision tree based on the experts, while fuzzy logic used to categorize the type of hypertensive disease that suffered by the patient and increase the level of accuracy of the diagnosing system. The accuracy of the combination between c5.0 algorithm and fuzzy logic is about 97.19%.

Keywords: *expert system, hypertension, C5.0 algorithms, fuzzy logic*

Copyright © 2014 Institute of Advanced Engineering and Science. All rights reserved.

1. Introduction

As many as 1 billion people worldwide suffer from hypertension. In the United States, nearly 1 in 3 adults (approximately 73 million people) have some degree of high blood pressure. Hypertension is a contributing factor to many other diseases including myocardial infarction (MI), stroke, heart failure, renal failure, and retinopathy, and is a leading cause of death [1].

Awareness, treatment, and control of hypertension are suboptimal. Only two-thirds of patients with hypertension are aware of their status, which means that a large segment of the population has hypertension that is unrecognized and untreated. Even in patients with known hypertension, some are not treated for various reasons; including physician and patient under-recognition of the importance of treatment.To minimize the impact of hypertension, the need for an expert system for early diagnosing patients with hypertension. Expert systems are computer programs that attempt to achieve the level of performance in a way that is comparable to solve problems with a human expert in taking a decision [2].

An Expert system is a computer program that simulates the judgment and behavior of a human being or an organization that has expert knowledge and experience in a particular field. Typically, such a system contains a knowledge base of accumulated experience and a set of rules for applying the condition to each particular situation that is described in the program. Sophisticated expert systems can be enhanced with additions to the knowledge base or to the set of rules. In other words, it is a software based system which makes or evaluates decisions based on rules established within the software [3]. Compared to the knowledge of human experts, the Expert System Knowledge has the advantages of permanent, easy to transfer, easy to edit, compatible, moderate and well suited for fault diagnosis [4].

In the domain of healthcare, it is important that the system is accurate in diagnosing because it deals with lives of persons which can never be replaced if an error occurred. There are various techniques on implementing the expert system and almost uses accurate and established algorithms. Most of these use data mining techniques [5]. In the first phase, the medical background of diseases is recorded through the creation of personal interview with doctors and patients. In the second phase, a set of rules is created where each rule contains in IF part that has the symptoms and in THEN part that has the disease that should be realized [6].

Received February 1, 2013; Revised March 21, 2014; Accepted April 14, 2014

Machine learning systems can be used to develop the knowledge bases used by expert systems. Given a set of clinical cases that act as examples, a machine learning system can produce a systematic description of those clinical features that uniquely characterise the clinical conditions. This knowledge can be expressed in the form of simple rules, or often as a decision tree [7].

Hypertension is a condition in which a person experiences an increase in blood pressure above normal. Hypertension is estimated to cause 4.5% of the global disease burden and is as prevalent in many developing countries as in developed countries. Worldwide, seven million premature deaths have been attributed to hypertension. In recent decades, it has become increasingly clear that the development of stroke, ischemic heart disease, and renal failure have been attributed by hypertension [8].

Many people have hypertension for years without even knowing it. According torecent estimates, one in four adults in the United States have hypertension, but, because there are few symptoms, nearly onethird of these people don't know they have it. That is why it is called the "silent killer" [9].

2. Research Method

The type of disease that is created as the object of research is hypertension. Hypertension is classified into several types: Pre Hypertension, Stage I Hypertension and Stage II Hypertension [1]. Expert's knowledge is obtained from literature sources such as library books and internist. Basic Informations based on the results of clinical examination and laboratory tests. This system's knowledges is representated by using the C5.0 Algorithm and Fuzzy Logic. The system is developed into web-based system which giving the users ease of access. The output produced by this system is the belief of the illness and the stage of the hypertension.

2.1. C.50 Algorithm

C5.0 model done by splitting the sample based on the field that provides the maximum information gain. In C5.0 algorithms, the selection of the attributes processed using information gain. Attributes selected heuristically to find purest attributes, which is giving the most net node. If one of the branches of a decision tree derived from one class, then this branch is called pure [10]. Information gain is the main criteria that is used. So in selecting attributes for splitting objects in some classes, we should select the attributes that produce the best information gain [11].

The size of the information gain is used to select the test attributes at each node in the tree. This size is used to select the attributes or nodes on the tree. Attribute with the highest information gain value will be selected as the parent for the next node [12]. The formula to get the information gain value is:

$$I(s_1, s_2, \dots, s_m) = -\sum_{i=1}^{y} p_i \; log_2 \, (p_i) \tag{1}$$

S is a set that consists of s sample data. known as class attribute, m defines the classes in it, C_i (for i = 1, ..., m), S_1 the number of samples of S in class C_i. To classify the samples that going to be used, Use the rule (1) above to obtain the required information. According to rule (1) above, p_i is a proportion of the output class, same as the class C_i and estimated by si /s. Attribute A has specific values {a_1, a_2, ..., a_v}. Attribute A can be used for partition S into v subsets, {S_1, S_2, ..., S_v}, where S_j contains samples on S worth aj in A [11].

If A is chosen as the test attribute (for example, the best attribute for split), then this subset will be related to the branch of node set S. S_{ij} is the number of samples in class Ci in a subset S_j. To get the value of a subset of the attribute A, then use formula (2),

$$E(A) = \sum_{j=1}^{y} \frac{s_{1j} + \dots + s_{mj}}{s} \, I\left(s_{1j}, \dots, s_{mj}\right) \tag{2}$$

$\frac{s_{1j} + \dots + s_{mj}}{s}$ is the number of subsets j divided by the number of samples in S. To get the value of the gain, then used the formula [11].

$$GAIN(A) = I\,(s_1, s_2, \dots, s_m) - E(A) \tag{3}$$

Having the first branch determined, the next process is to do the second iteration. In the second iteration, the attributes of the first branch will not be used in this process. The process of determining the second branch is the same as the previous method of determining branch. The iterative process will continue until the data can not be solved anymore. From the various existing branches will form a scheme that will end in an initial hypothesis.

2.2. Fuzzy Logic

Basically, Fuzzy Logic is a multivalued logic, that allows intermediate values to be defined between conventional evaluations like true/false, yes/no, high/low, etc. Notions like rather tall or very fast can be formulated mathematically and processed by computers, in order to apply a more human-like way of thinking in the programming of computers [13].

2.2.1. Knowledge Representation

Knowledge of clinical symptoms and laboratory results of the patient is necessary for the representation of these facts along with an explanation of the question [14]. The question scheme of the system is obtained from the decision tree through C50 algorithm. The questions scheme will be ended by giving the result about which type the hypertension is. Laboratory results will represented by a fuzzy set with a membership function. Some fuzzy sets model shown below.

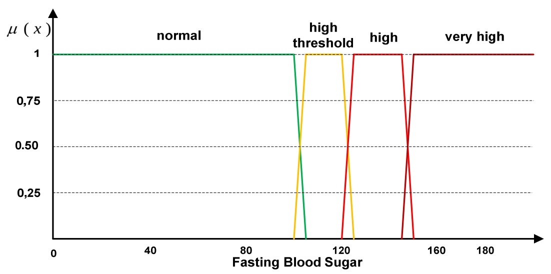


Figure 1. Membership Curves of Fasting Blood Sugar

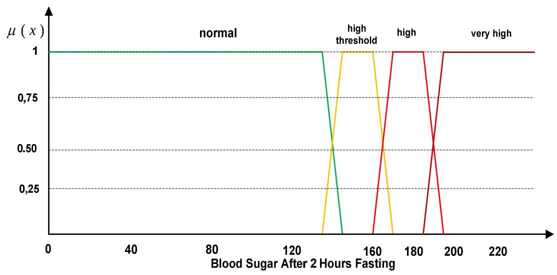


Figure 2. Membership Curves of Blood Sugar After 2 Hours Fasting

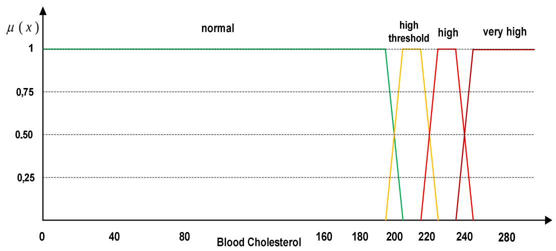


Figure 3. Membership Curves of Blood Cholesterol

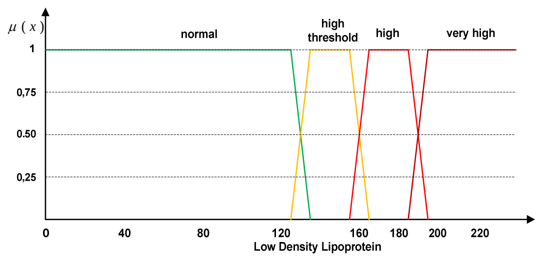


Figure 4. Membership Curves of Low Density Lipoprotein (LDL)

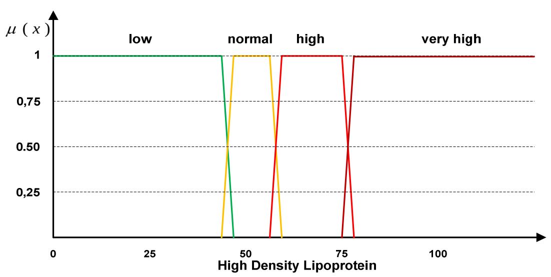


Figure 5. Membership Curves of High Density Lipoprotein (HDL)

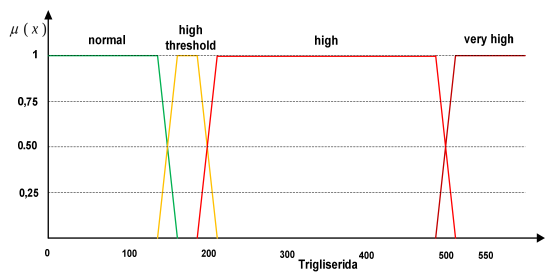


Figure 6. Membership Curves of Trigliserida

Membership functions of Fasting Blood Sugar:

$$\mu\, normal(x) = \begin{cases} 1, & x \le 100 \\ \frac{105-x}{5}, & 100 \le x \le 105 \end{cases} \qquad \mu\, very\ high\,(x) = \begin{cases} \frac{x-145}{5}, & 145 \le x \le 150 \\ 1, & x \ge 150 \end{cases}$$

$$\mu\, high\ threshold\,(x) = \begin{cases} \frac{x-100}{5}, & 100 \le x \le 105 \\ 1, & 105 \le x \le 120 \\ \frac{125-x}{5}, & 120 \le x \le 125 \end{cases} \qquad \mu\, high\,(x) = \begin{cases} \frac{x-120}{5}, & 120 \le x \le 125 \\ 1, & 125 \le x \le 145 \\ \frac{150-x}{5}, & 145 \le x \le 150 \end{cases}$$

Membership functions of the Blood Sugar After Fasting 2 Hours:

$$\mu\, normal(x) = \begin{cases} 1, & x \le 135 \\ \frac{145-x}{10}, & 135 \le x \le 145 \end{cases} \qquad \mu\, high\,(x) = \begin{cases} \frac{x-160}{10}, & 160 \le x \le 170 \\ 1, & 170 \le x \le 185 \\ \frac{195-x}{10}, & 185 \le x \le 195 \end{cases}$$

$$\mu\, high\ threshold\,(x) = \begin{cases} \frac{x-135}{10}, & 135 \le x \le 145 \\ 1, & 145 \le x \le 160 \\ \frac{170-x}{10}, & 160 \le x \le 170 \end{cases} \qquad \mu\, very\ high\,(x) = \begin{cases} \frac{x-185}{10}, & 185 \le x \le 195 \\ 1, & x \ge 195 \end{cases}$$

Membership functions of Blood Cholesterol:

$$\mu\, normal(x) = \begin{cases} 1, & x \le 195 \\ \frac{205-x}{10}, & 195 \le x \le 205 \end{cases} \qquad \mu\, high\,(x) = \begin{cases} \frac{x-215}{10}, & 215 \le x \le 225 \\ 1, & 225 \le x \le 235 \\ \frac{245-x}{10}, & 235 \le x \le 245 \end{cases}$$

$$\mu high\ threshold(x) = \begin{cases} \frac{x-195}{10}, & 195 \le x \le 205 \\ 1, & 205 \le x \le 215 \\ \frac{225-x}{10}, & 215 \le x \le 225 \end{cases} \qquad \mu\, very\ high\,(x) = \begin{cases} \frac{x-235}{10}, & 235 \le x \le 245 \\ 1, & x \ge 245 \end{cases}$$

Membership functions of Low Density Lipoprotein (LDL)

$$\mu\, normal(x) = \begin{cases} 1, & x \le 125 \\ \frac{135-x}{10}, & 125 \le x \le 135 \end{cases} \qquad \mu\, high\,(x) = \begin{cases} \frac{x-155}{10}, & 155 \le x \le 165 \\ 1, & 165 \le x \le 185 \\ \frac{195-x}{10}, & 185 \le x \le 195 \end{cases}$$

$$high\ threshold(x) = \begin{cases} \frac{x-125}{10}, & 125 \le x \le 135 \\ 1, & 135 \le x \le 155 \\ \frac{165-x}{10}, & 155 \le x \le 165 \end{cases} \qquad \mu\, very\ high\,(x) = \begin{cases} \frac{x-185}{10}, & 185 \le x \le 195 \\ 1, & x \ge 195 \end{cases}$$

Membership functions of High Density Lipoprotein (HDL):

$$\mu\, low\,(x) = \begin{cases} 1, & x \le 40 \\ \frac{45-x}{5}, & 40 \le x \le 45 \end{cases} \qquad \mu\, high\,(x) = \begin{cases} \frac{x-60}{5}, & 60 \le x \le 65 \\ 1, & 65 \le x \le 75 \\ \frac{80-x}{5}, & 75 \le x \le 80 \end{cases}$$

$$\mu\, normal\,(x) = \begin{cases} \frac{x-35}{5}, & 35 \le x \le 45 \\ 1, & 45 \le x \le 60 \\ \frac{65-x}{5}, & 60 \le x \le 65 \end{cases} \qquad \mu\, very\ high\,(x) = \begin{cases} 0, & x \le 75 \\ \frac{x-75}{5}, & 75 \le x \le 80 \\ 1, & x \ge 80 \end{cases}$$

Membership functions of Trigliserida:

$$\mu\, normal(x) = \begin{cases} 1, & x \leq 140 \\ \frac{165-x}{25}, & 140 \leq x \leq 165 \end{cases}$$

$$\mu\, high\ (x) = \begin{cases} \frac{x-190}{25}, & 190 \leq x \leq 215 \\ 1, & 215 \leq x \leq 490 \\ \frac{515-x}{25}, & 490 \leq x \leq 515 \end{cases}$$

$$\mu\, high\ threshold\ (x) = \begin{cases} \frac{x-140}{25}, & 140 \leq x \leq 165 \\ 1, & 165 \leq x \leq 190 \\ \frac{215-x}{25}, & 190 \leq x \leq 215 \end{cases}$$

$$\mu\, very\ \ high\ (x) = \begin{cases} \frac{x-490}{25}, & 490 \leq x \leq 515 \\ 1, & x \geq 515 \end{cases}$$

2.2.2. Implication and Composition

Calculation of the degree of fuzzy membership for each symptom is determined by the value assigned by the user [15].

Table 1. Fuzzy Rule

Rule Number	Rule	Value
R000001	IF Parental history of suffering from hypertension is YES THEN	CF = 0.90
R000002	IF Eat Food With High Salt Levels is often AND Stress is often THEN	CF = 0.70
R000003	IF Eat Food With High Salt Levels is often AND Sleep Apnea is often THEN	CF = 0.75
R000014	IF Stress is often AND vertigo is often THEN	CF = 0.65
R000035	IF Smoking is very often AND Vertigo is often THEN	CF = 0.72
R000045	IF Sistolic Blood Pressure Normal AND Diastolic Blood Pressure Normal THEN	CF = 0.50
R000057	IF Sistolic Blood Pressure High AND Diastolic Blood Pressure High THEN	CF = 0.90
R000068	IF Sistolic Blood Pressure High Threshold AND Diastolic Blood Pressure Very High THEN	CF = 0.85

Based on the degree of membership, calculate the implication function with MIN function [16]-[18].μ(x) is the degree of membership for x and w_i is the result of implication.

$$wi = MIN\big(\mu(x), \mu(y)\big) \tag{4}$$

The process of composition is made to obtain the value z_i of each rule. The certainty value from expert of each rule is value of z_i [14].

2.2.3. Defuzzification

Defuzzification process is done using weighted average method defuzzification by calculating the average value of z_i [16-18].

$$\boldsymbol{z} = \frac{\sum_{j=1}^{n} w_i z_i}{\sum_{j=1}^{n} w_i} \tag{5}$$

wi is the result of implication and z_i is the result of composition [14]. The results of defuzzification demonstrate the value of the belief for the syndrome experienced by patients.

2.2.4. Certainty Factor Calculation

The result of defuzzification process will be used to calculate the value of belief for the diagnosis. Firstly, will be calculated certainty factor (CF) sequential as follows [16].

$$CF(x,y) = CF(x) * CF(y) \tag{6}$$

CF(x,y) is result of certainty factor sequential, CF(x) is result of defuzzification and CF(y) is the expert certainty value of each rule. In this study, CF sequential multiplies with the weight value of each phase of disease. CF sequential from several rules generated combined using the following calculation of the combined CF as follows [17].

$$CF(x,y) = CF(x) + CF(y) - \big(CF(x) * CF(y)\big) \tag{7}$$

The results of combined CF suggest the diagnosis of the disease to the symptoms experienced by patients [14].

3. Results and Discussion

The main objective of this research is to create an expert system which can be used for diagnosing hypertension. The expectation of this system is to help users to overcome the effects of hypertension by telling them early about their condition.

3.1. Knowledge Acquisition

The Knowledge Acquisition role is transferring knowledge that gained from completing a variety of literature and experts knowing into a knowledge base. This knowledge base will be the most important part of an expert system for storing all the knowledge to be used as a standard for decision making.

3.2. Knowledge Representation

The first thing that is done in representing the knowledge is started from generating a decision tree using C5.0 algorithm. The process of forming a decision tree start from selecting the initial tree branches. This branch is obtained from the value of the attribute that has the highest gain of all the existing attributes. In this case we use 15 attributes that will be used in generating a decision tree. The decision tree made up three-pronged branches, that in every branch later is the result of the question in the previous branch.

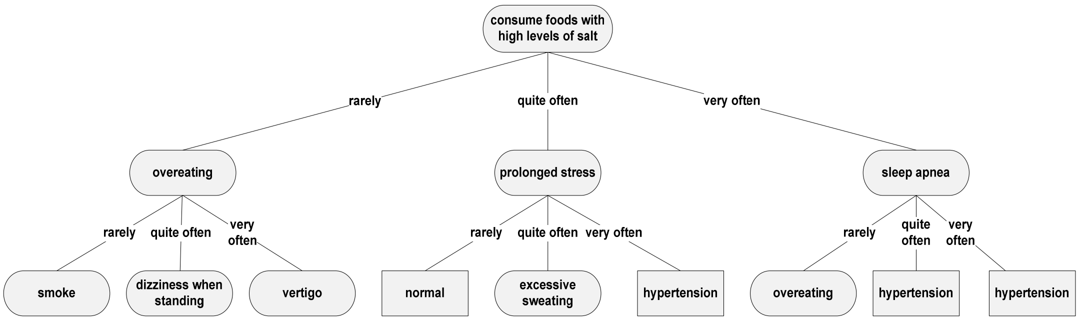


Figure 7. Decision Tree

After determining the first branch, the process continues to the next iteration. In the second iteration, the first branch attributes will not be used in this process. The process of determining the second branch is the same as the previous method to determine the first branch. Iterative process will continue until the data can not be solved again. From the various existing branches will form a scheme that will ended up by giving an initial hypothesis.

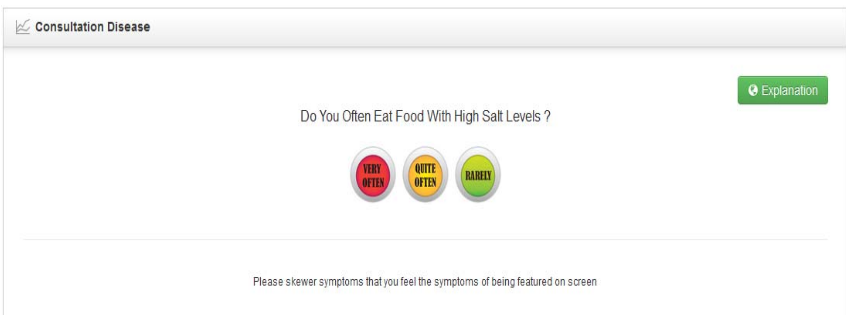


Figure 8. Consultation User Interface

The fuzzification process in the diagnosing hypertension is divided into 2 phases. The first phase begins with processing the data proximity of the decision tree with the data from the knowledge base that has undergone the fuzzification process. The data were fed by the user will be matched with hypotheses that have been made by fuzzy logic. The next process is processing the result of a laboratory data that may be owned by the user. Data from a laboratory results in numbers will be grouped into several classes.

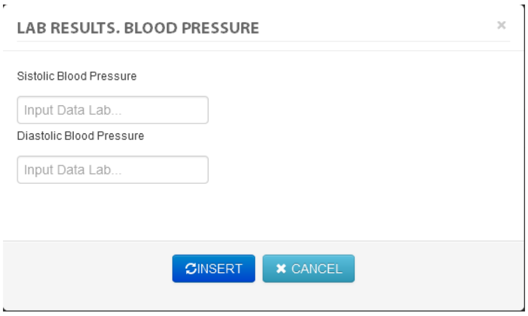
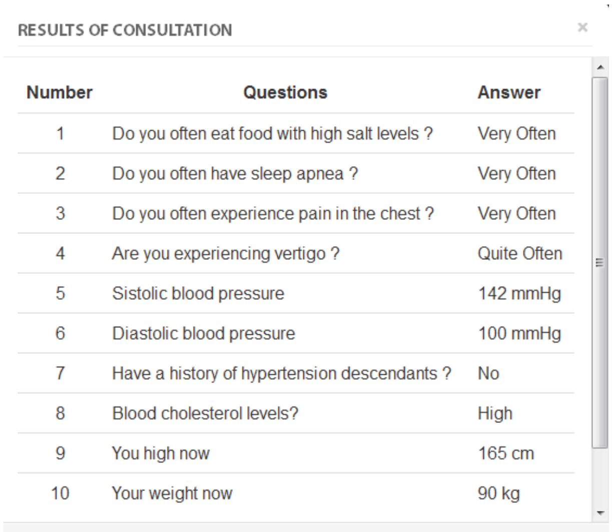


Figure 9. Consultation User Interface With Fuzzy Logic

Results from each of these classes will be processed by implicating classes from every existing laboratory results.

3.3. System Performance

The performance of the expert system is obtained from the comparison of the results of the diagnosis made by a real expert with the diagnoses given by the expert system. Based on the structure of the system above, this expert system was tested in many different cases. One example of consultation and diagnosis of the system is given in Figure 10.



RESULTS OF CONSULTATION

Number	Questions	Answer
1	Do you often eat food with high salt levels ?	Very Often
2	Do you often have sleep apnea ?	Very Often
3	Do you often experience pain in the chest ?	Very Often
4	Are you experiencing vertigo ?	Quite Often
5	Sistolic blood pressure	142 mmHg
6	Diastolic blood pressure	100 mmHg
7	Have a history of hypertension descendants ?	No
8	Blood cholesterol levels?	High
9	You high now	165 cm
10	Your weight now	90 kg

Figure 10. Results of Consultation

Total sample of cases used to determine the accuracy of the data is measuring the accuration percentage by comparing 900 data samples. The data will be processed using algorithm c5.0 which will generating a decision tree. Decision tree will be used as a knowledge base system. Grievances that perceived by users will be matched with a decision tree, that will give a conclusion about the illness suffered by users.The performance of the expert system with C5.0 algorithm can be seen in Table 2.

Table 2. Expert Systems Performance with C5.0 algorithm

Scheme Case	Total Cases	Same Case With Scheme	Correct	Wrong	Accuracy (%)
1	900	50	44	6	88
2	900	39	37	2	95
3	900	48	45	3	93.75
4	900	49	48	1	98
5	900	64	60	4	93.75
6	900	60	57	3	95
7	900	57	57	0	100
8	900	38	37	1	97.22
9	900	65	65	0	100
10	900	40	38	2	95
Average of the difference result					95.57

Expert System Performance with a combination of C5.0 and fuzzy logic algorithm has a better accuracy rate than the use of two methods separately. The combination can improve the accuracy of diagnoses predicted by the expert system. The expert system performance with a combination of C5.0 algorithms and fuzzy logic can be seen in Table 3.

Table 3. Performance with a Combination of C5.0 Algorithms and Fuzzy Logic

Case	System Diagnosis Fuzzy (%)	System Diagnosis Fuzzy & C50 (%)	Expert Diagnosis (%)	The difference result (%) Fuzzy	The difference result Fuzzy & C5.0 (%)
1	59	58	55	6.78	5.17
2	69	71	72	4.17	1.39
3	38	37	35	7.89	5.41
4	77	74	75	2.6	1.33
5	61	61	64	6.25	4.69
6	66	62	64	4.47	3.12
7	69	71	70	2.78	1.41
8	67	64	65	2.99	1.54
9	74	71	72	4	1.39
10	71	76	74	4.05	2.63
Average of the difference result				4.60	2.81

4. Conclusion

Expert system for diagnosing Hypertension has been developed into a web-based platform to receive feed in a form of physical symptoms and laboratory test results. the questions systematic that arises is a decision tree derived from processing the required data by using the C5.0 algorithm. Fuzzy rules relating each disease symptoms using expert certainty. This system provides an output of hypertension illness possibility that suffered by users. The combination of C50 with fuzzy logic algorithm increase the expert system hypotheses accuracy. From statistical Performance accuracy of c5.0 is 95.7% and Fuzzy logic 95.4%. Test results with a combination of c5.0 algorithm with fuzzy logic system shows that the developed expert system has 97.19% accuracy of the similarity to a real expert.

References

[1] Jeffery Martin. Hypertension Guidelines: Revisiting the JNC 7 Recommendations. *The Journal of Lancaster General Hospital*. 2008; 3(3): 91-97.

[2] Joseph, Hans-D. Web-Based Expert System for Classification of Industrial and Commercial Waste Products. *Journal of Emerging Trends in Computing and Information Sciences*. 2011; 2(6): 357-262.

[3] Josephine, Jeyabalaraja. Expert System and Knowledge Management for Software Developer in Software Companies. International *Journal of Information and Communication Technology Research*. 2012; 2(3): 243-247.

[4] Feng Yongxin, Yang Tao.Study of Fault Diagnosis Method for Wind Turbine with Decision Classification Algorithms and Expert System. *TELKOMNIKA Indonesian Journal of Electrical Engineering*. 2012; 10(5): 905-910.

[5] Romeo Mark, Jaewan Lee. Healthcare Expert System based on Group Cooperation Model. *International Journal of Software Engineering and Its Application.* 2008; 2(1): 105-116.

[6] Santosh Kumar, Dipti Prava. An Expert System for Diagnosis of Human Diseases. International *Journal of Computer Applications*. 2010; 1(13): 71-73.
[7] Prasadl, Krishna. An Approach to Develop Expert Systems in Medical Diagnosis Using Machine Learning Algorithms (Asthma) and A Performance Study. *International Journal on Soft Computing (IJSC)*. 2011; 2(1): 26-33.
[8] Syer Ree Tee, Xin Yun Teoh. The Prevalence of Hypertension and Its Associated Risk Factors In Two Rural Communities In Penang, Malaysia. *IeJSME.* 2010: 4(2): 27-40.
[9] Christine Laine ed. Living With Hypertension. American College of Physicians and Microlife Corporation. Report number: 4. 2004.
[10] Prof. Nilima Patil, Prof Rekha. Comparison of C5.0 & CART Classification algorithms using pruning technique*. International Journal of Engineering Research & Technology (IJERT)*. 2012; 1(4): 1-5.
[11] Ernawati. Prediksi Status Keaktifan Studi Mahasiswa Dengan Algoritma C5.0 Dan K-Nearest Neighbor Thesis. Bogor: Postgraduate IPB. 2008.
[12] Mehmed Kantardzic. Data Mining: Concepts, Models, Methods, and Algorithms. *IEEE Press, A john Wiley & Sons, Inc, Publication*. 2003: 177-178.
[13] LA Zadeh. Making computers think like people. *IEEE. Spectrum*.1998; 21(8) 26-32.
[14] Darma Putra. Fuzzy Expert System for Tropical Infectious Disease by Certainty Factor. *TELKOMNIKA Indonesian Journal of Electrical Engineering*. 2012; 10(4): 825-836.
[15] Putu Manik Prihatini, Katut Gede Darma Putra. Expert knowledge based system with uncertainity for Tropical Infections Dieseased Diagnosis*. IJSCSI*. 2012; 3(4): 157-163.
[16] Sivanandam SN, Sumathi S, Deepa SN. Introduction to Fuzzy Logic using MATLAB. New York: Springer. 2007.
[17] Negnevitsky Michael. Artificial Intelligence A Guide to Intelligent Systems Second Edition. England: Pearson Education. 2002
[18] Konar Amit. Artificial Intelligence and Soft Computing Behavioral and Cognitive Modelling of the Human Brain. Washington DC: CRC Press. 2000.