Bayesian Network Expert System for Early Diagnosis of Heart Diseases

Mohamad Saad A.Sawa, Agus Naba and Harry Soekotjo Dahlan

***Abstract*—Heart disease is a deadly disease in the world. Some countries that have a high risk of death are America, Australia and England. It is difficult for a person do a medical checkup not only because of the financial matter but also because they don't care about it. This study is to help people on early diagnose to the risk of heart disease. Expert System is used as the basis of this study which is using Bayesian Network algorithms with datasets from previous studies conducted in Europe. The objective of this research are (1) to design knowledge base, inference engine with Bayesian network for early diagnose of heart diseases, (2) to develop web-based system that help people to get early diagnose of risk heart disease. This study was conducted by the help of a heart disease specialist and internist that helps validate the dataset and the results of research through application testing detection of heart disease risk. This result is consistent with the analysis that can be used by physicians and public.**

***Index Terms*—expert system, Bayesian network, heart disease, early diagnose.**

I. INTRODUCTION

ACCORDING to the research conducted by the WHO, it is noted that a heart disease is a deadly disease and the cause of the number one killer in the world. In the United States itself, each year approximately 1,400,000 people died because of it and doctors in the world are trying to find a solution to prevent it. [1]

The diseases can affect any part of the heart. However, the most common disease is a chronic disease of the coronary arteries is called Atherosclerosis. Therefore, heart disease is commonly known as coronary heart disease or coronary artery disease. This disease often becomes a trigger of a deadly heart attack. The trigger is the narrowing of the coronary arteries, where the vessel serves to supply the blood into the heart muscle. The construction caused by the accumulation of cholesterol or other proteins that come from the food they consume. This factor causes the coronary arteries become stiff. Stiffness is called Atherosclerosis.[2]

The pain or pressure in the chest, called Angina, is a warning to those who suffers from a heart disease. Some people experience short of breath or fatigue and weakness as symptoms, are indicating that their heart are not getting enough oxygen because of coronary blockage.

Mohamad Saad A.Sawa is with the Electrical Engineering Departement of Universitas Brawijaya, Malang, Indonesia (email sowamemo@gmail.com)

Dr. Eng Agus Naba, S.Si., M.T. is with the Physical Engineering Departement of University Brawijaya, Malang, Indonesia.

Dr. Ir. Harry Soekotjo Dahlan Msc. is with the Electrical Engineering Departement of Universitas Brawijaya, Malang, Indonesia

Usually before having a massive heart attack, a person will experience muscle contractions suddenly in the chest which is a small attack or a minor heart attack. Generally a minor heart attack occurs before a major attack eventually.

Heart disease can happen to anyone. When someone is experiencing pain in his left chest, he doesn't visit a doctor soon and undergo treatment is done for he thought that it is only minor physical problem. But if the verdict is known that they suffer from heart disease, the problem is worse.

Based on these problems, the researchers will examine the symptoms of the heart disease then make an expert system which provides valuable information to public and use it to detect the risk of heart disease based on the symptoms.

The result is a web-based expert system application that can be used by the public in the early diagnosis for the risk of heart disease. If they find out that they have a risk of heart disease is more than 50%, it is highly recommended to visit a doctor soon.

II. LITERATURE REVIEW

A. Heart Disease

Heart disease is an umbrella term for any type of disorder that affects the heart. Heart disease means the same as cardiac disease but not cardiovascular disease. Cardiovascular disease refers to disorders of the blood vessels and heart, while heart disease refers to just the heart. According to WHO (World Health Organization), heart disease is the main factor of death in the UK, USA, Canada and Australia. 25.4% of total deaths in the USA today are caused by heart disease.

Danish scientists found that the visible physical signs of aging, such as receding hairlines and the accumulation of fatty deposits on the eyelids, are associated with a higher risk of developing heart disease and having a heart attack. They presented their findings at the American Heart Association's Scientific Sessions 2012 in Los Angeles. Lead researcher, Anne Tybjaerg-Hansen, said "The visible signs of aging reflect physiologic or biological age, not chronological age, and are independent of chronological age." They found that a receding hairline, baldness, earlobe crease and xanthelasmata (fatty deposits around the eyelids) increased ischemic heart disease risk by 39% and heart attackrisk by 57%. [3]

B. Expert System

An expert system is software that uses a knowledge base of human expertise for problem solving, or to clarify uncertainties where normally one or more human experts would need to be consulted. [4]

Expert systems are most common in a specific

problem domain, and are a traditional application and/or subfield of artificial intelligence (AI). [5]

A wide variety of methods can be used to simulate the performance of the expert [6]; however, common to most or all are: 1) the creation of a knowledge base which uses some knowledge representation structure to capture the knowledge of the Subject Matter Expert (SME); 2) a process of gathering that knowledge from the SME and codifying it according to the structure, which is called knowledge engineering; and 3) once the system is developed, it is placed in the same real world problem solving situation as the human SME, typically as an aid to human workers or as a supplement to some information system. Expert systems may or may not have learning components. [7]

Expert systems were introduced by researchers in the Stanford Heuristic Programming Project,including the "father of expert systems" Edward Feigenbaum, with the Dendral and Mycin systems. Principal contributors to the technology were Bruce Buchanan, Edward Shortliffe, Randall Davis, William vanMelle, Carli Scott, and others at Stanford. Expert systems were among the first truly successful forms of AI software. [8]

In artificial intelligence, an expert system is a computer system that emulates the decision-making ability of a human expert. Expert systems are designed to solve complex problems by reasoning about knowledge, like an expert, and not by following the procedure of a developer as is the case in conventional programming. The first expert systems were created in the 1970s and then proliferated in the 1980s. Expert systems were among the first truly successful forms of AI software. [9]

C. Tools

There are several application used on this research, from beginning until validation. These applications are: (a) PHP, (b) PhpMyAdmin, (c) Php Designer, (d) MySql

D. PHP

PHP stands for PHP: Hypertext Preprocessor, it is a server-side scripting language, like ASP, PHP scripts are executed on the server, itsupports many databases (MySQL, Informix, Oracle, Sybase, Solid, PostgreSQL, Generic ODBC, etc.) PHP is open source software. [10]

E. PhpMyAdmin

PhpMyAdmin is an open source tool written in PHP intended to handle the administration of MySQL with the use of a Browser. It can perform various tasks such as creating, modifying or deleting databases, tables, fields or rows; executing SQL statements; or managing users and permissions. [11]

F. MySql

MySQL is official software developed by Swedish company called MySQL AB, whose name was TcX Data Konsult AB. At the beginning, MySQL was using the name “mSQL” or "mini-SQL" as the interface used, apparently using mSQL was experiencing a lot of obstacles, which is very slow and inflexible. Therefore, Michael Widenius ("Monty"), hir nickname, try to develop the interface and finally MySQL were developed. At the moment, MySQL was distributed specifically for noncommercial purposes and it was free, while for commercial purpose it required to pay the license. Since version 3.23.19, MySQL has categorized as GPL licensed software, which can be used without payment. [12] [13]

G. Bayesian Statistics

Bayesian Statistics is a method used in this research to build a Bayesian Network which is used in the inference engine. Suppose that person A at a party. And Person B on way to the party, late. A friend asks person B, “Do you suppose that Karl has already had too many beers?” Based on past experience with person A at such parties, prior probability of person B having had too many beers, $p(B) = 0.2$. The probability that person A have not had too many beers, $p(\overline{B}) = 0.8$, giving prior odds, Ω' of: $\Omega' = \frac{p(B)}{p(\overline{B})} = \frac{0.2}{0.8} = 0.25$ (inverting the odds, the probability that person A have not had too many beers is 4 times the probability that have). Ω is Greek omega. [14]

III. METHODOLOGY

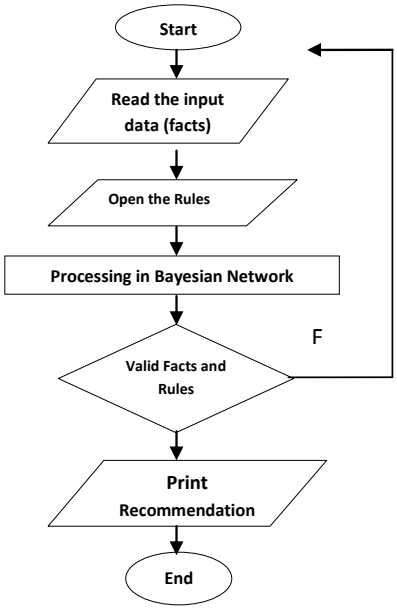

Fig. 1 Flowchart of Research

Expert systems are a piece of software [15]. Expert systems are similar to conventional software in some respects, but different in others. To build expert system software, first is getting the dataset from previous research (Lyon University) [16], analyzing the dataset with advice from the doctor, generate node for bayesian network, building application and make validation.

Thus some software engineering methods also apply to expert systems and features of the software lifecycle

can also be observed in expert systems.[17]It can be drawn an analogies between the phases in expert systems construction and corresponding phase in software development. Expert system lifecycle can be seen in Fig.1. The process of expert systems development can be compared with the traditional software lifecycle by identifying five phases in the expert systems construction process. [18]

A. Getting dataset

Knowledge base has an important role in expert system, it is a brain of the system and will help inference engine based on Bayesian network to get a better conclusion [19]. This dataset taken from University of Lyon, French and has been used by the previous research. This dataset consist of 209 patient followed by the hospital and with complete documentation to use it. Another dataset can be taken from WHO but it has limited documentation.

TABLE I
EXAMPLE OF DATASET

Age	chest_pain	rest_electro	max_heart rate	Disease
43	asympt	Normal	135	Positive
39	atyp_angina	Normal	160	Negative
39	non_anginal	Normal	160	Negative
42	non_anginal	Normal	146	Negative
49	asympt	Normal	130	Negative
50	asympt	Normal	135	Negative
59	asympt	left_vent_hyper	119	Positive

After getting dataset, the next step is to analyze to get the value of probability. The analyzing is to compare the value of current object with the whole set member.

TABLE II
EXAMPLE OF ANALYST DATASET

age	negative	positive	Grand Total	prob
31		1	1	1
32	2	1	3	0,333333
33	1		1	0
34	2	1	3	0,333333
35	2	1	3	0,333333
36	4	1	5	0,2
37	4	1	5	0,2
38	2	4	6	0,666667
39	9	1	10	0,1
40	5		5	0

Dataset analysis is done by comparing the values obtained with the total data. Once this analysis is done, the doctor will give advise and correction on the results.

B. Table Design

Based on Ulman [20], table should consist information about the variable and clear name.

C. Designing table disease_type

Disease table is used to store the information about disease and additional information related to the table. The Table III describes that Field id is the primary key that knows as the id of disease type. Fieldtype_disease is the type of heart disease.Field name_ disease is the name of the disease,and field information will store the additional information about related disease.

TABLE III
DISEASE

Field	Type	Size	Description
id	Integer	11	Primary Key
type_disease	Varchar	100	Name of the disease
name_disease	Varchar	100	Additional information for
information	text	text	current disease

D. Designing table symptoms

This table designing to store all of the symptoms of disease, related to the heart disease. The following table gives the structure of the symptoms table.

Table IV describes that field id is the primary key for unique id on symptoms table. Field symptoms is the name of the symptoms. Field information will contain additional information regarding to the symptoms.

TABLE IV
SYMPTOMS

Field	Type	Size	Description
id	Integer	11	Primary Key
symptoms	Text	100	Name of the symptoms
information	Text	11	Additional information for current symptoms

E. Designing table of mapping

Mapping table will be used for mapping between disease and symptoms. It contains the appropriate information collected from the experts.

TABLE V
MAPPING

Field	Type	Size	Description
id_mapping	Integer	11	Primary Key
id_disease	Tinyint	4	Name of the disease, symptoms
id_symptom	Tinyint	4	Additional information for
information	Text	256	current symptoms

Table V describes that field id_mapping is the primary key to identify unique elements on this table.Field id_disease is the foreign key that connected to table disease. Field id_symptoms is the foreign key that related to symptoms table, and fields information is to store additional information about mapping.

F. Designing table of question

This table is used to store question related to the symptoms and the rule of expert system. This data connected each other and linking the question.

TABLE VI
QUESTION TABLE STRUCTURE

Field	Type	Size	Description
id_question	Integer	11	Primary Key
question	Tinyint	4	Name of the disease,
id_mapping	Tinyint	4	symptoms
prev_tree	Text	256	Additional information for
information	Text	256	current symptoms

Table VI describes thatfield id_question is the primary key on the table, it's guarantee that each row

has unique id.Field question contain the information about the question that will be displayed, user type sentences related to the symptoms and disease. Field id_mapping store the id of mapping table, it's foreign key. Field prev_tree is the id of question table and expressing the previous question, and fields information is to store additional information about question.

G. Designing User_Answering table

This table is used to store answer related to the question and the rule of expert system.

TABLE VII
USER ANSWERING TABLE STRUCTURE

Field	Type	Size	Description
id_answer	Integer	11	Primary Key
question	Text	4	Question
yes_answer	Text	4	
no_answer	Text	256	
text_answer	Text	256	

Table VII describes that field id_answer is the primary key on the table, it's guarantee that each row has unique id. Field question contain the information about the question that will be displayed, user type sentences related to the symptoms and disease. Field yes_answer store the '1' value on table. Field no_answer store the '1' value on table, and field text_answer store the text answer value on table.

H. Designing Bayesian table

This table is used to store value of probability each symptom to diagnose heart disease.

TABLE VIII
BAYESIAN TABLE STRUCTURE

Field	Type	Size	Description
id	Integer	11	Primary Key
group_sym	Integer	11	Question
id_symptom	Integer	11	
min_val	float		
max_val	float		
prob_value	float		

Table VIII describes that Field id store the id value on table. Field group_sym store the group of symptom on table. Field id_symptom store the id value of symptom on table. Field min_val store the minimum value on table. Field max_val store the maximum value on table, and field prob_value store the value of probability on table.

I. User Interface Design

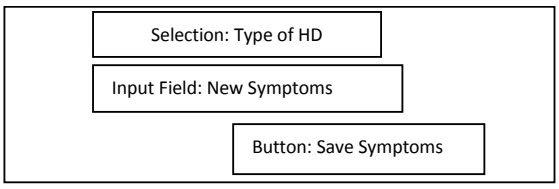

Fig.2 Form of Input New Symptoms

User interface of the system required three important points in the conceptual level. User will enter information about new symptoms of the disease that has been stored in the database. This information is entered in the form presented below:

User should choose the option that contains the suitable disease and then fill heart symptoms found in the input box, then afterwards will be able to continue to save by pressing the save button.

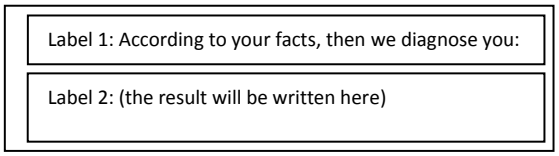

Fig.3 Form of Diagnose/Recommendation

After going through several stages of discussions between the system with the user, then the expert system will provide diagnose result presented in the following designs:

Label1 provide information about the title of this form, then Label2 provide information on the result as recommendation of an expert system based on existing facts and data have been previously entered.

To manage the rules that can be changed according to our wishes, it provided a form that helps in the rules management.

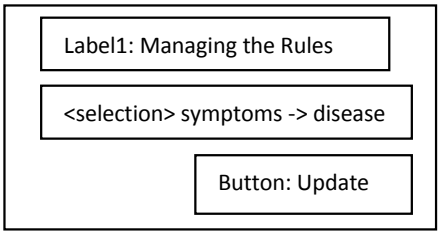

Fig.4 Form of Manage the rule

Form above shows that at the top of the label containing the title information, by the selection it can be configuration of disease and symptoms, where on the left is a symptoms and on the right is the disease. To manage the rules, users simply click and select the left content on the right content, then press button update to saving.

This system can view on database what is user answering according to the question provided by the system. In future, this user answer can be a FAQ (Frequently Asked Question) to provide the best question for the user.

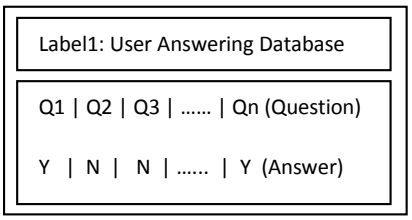

Fig.5Form of User Answering Database

Form above shows that at the top of the label containing the title information about user answering database, the box below show the question number and the answer of the user.

IV. RESULT AND DISCUSSION

System implementation generally benefits from high levels of user involvement. User participation in the design and operation of information systems have several positive results. First, if users are heavily involved in systems design, they move opportunities to mold the system according to their priorities and more opportunities to control the outcome. Second, they are more likely to react positively to the change process. Incorporating user knowledge and expertise leads to better solutions.

A. *Result*

This menu is the center of this research, namely to make the system running and try the accuracy of the application.

B. *Home menu*

The main menu contains some information, including title of research, researcher name, and displays hyperlinks to other menus like the menu for editing the knowledge base, running applications and the menu contains information about the application (see Fig. 6).

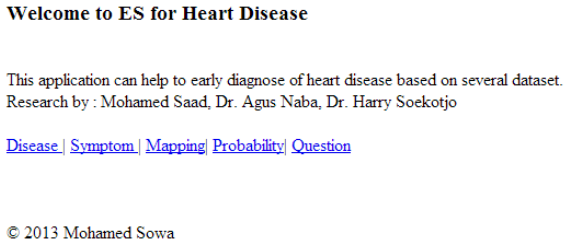

Fig. 6. Top menu header

C. *Maintaining the Bayesian Network*

This menu provides an easy way to manage probability according to likelihood on bayesian network to the application. Admin can provide additional information or about diseases, symptoms that are used (See Fig 7).

Fig. 7 Mapping Configurations

D. *Disease Management*

In this section, application provides an easy way to the admin to make changes to the disease property, or information about the disease (See Fig. 8)

E. *Symptoms Management*

In this section, application provides an easy way to the admin to make changes to the symptoms property, or information about the symptoms (See Fig. 9).

Fig. 8 Disease Management

Fig. 9 Symptoms Management Menu

F. *Running the expert system (ages)*

This menu is the center of this research, namely to make the system running and try the accuracy of the application (See Fig. 10).

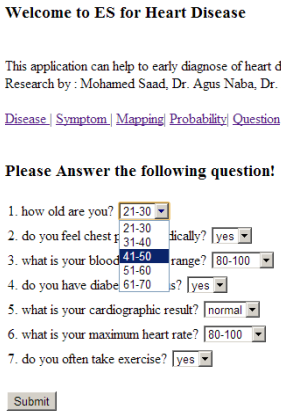

Fig. 10 Running the ES by Filtering ages

Fig.10 explain that the question of the age contains options with 5 categories, which range between the ages of 21-30, 31-40, 41-50, 51-60 and 61-70 years. Users simply choose among the range to answer questions.

G. *Running the expert system (chest pain)*

Fig.11 explain that the question of the chest pain contains options with 2 categories, which is 'yes' and 'no' answer of question. Users simply choose among the range to answer questions

H. *Running the expert system (blood pressure)*

Fig.12 explain that the question of the blood pressure

contains options with 6 categories, which range between the 80-100, 100-120, 120-140, 140-160, 160-180, and 180-200. Users simply choose among the range to answer questions

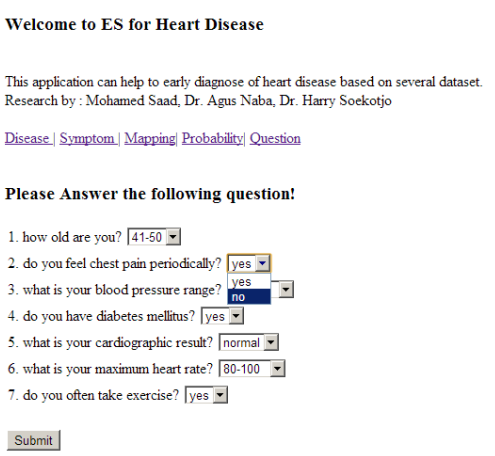

Fig. 11 Running the ES by Filtering chest pain

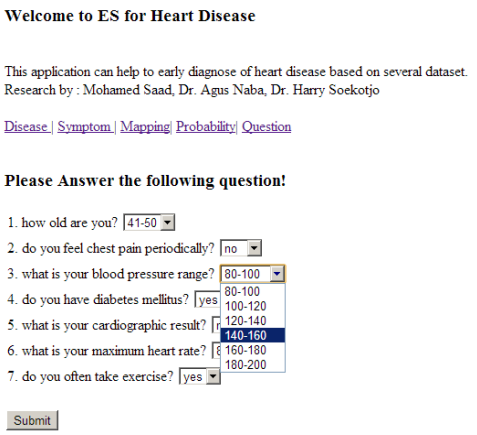

Fig. 12 Running the ES by Filtering blood pressure range

I. Running the expert system (diabetes mellitus)

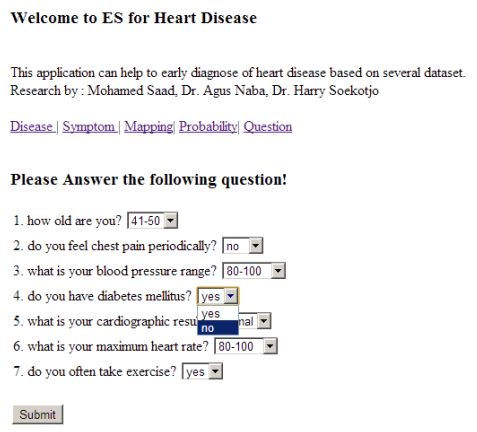

Fig. 13 Running the ES by Filtering diabetes mellitus

Fig.13 explain that the question of the diabetes mellitus contains options with 2 categories, which is 'yes' and 'no' answer of question. Users simply choose among the range to answer questions.

J. Running the expert system (cardiographic)

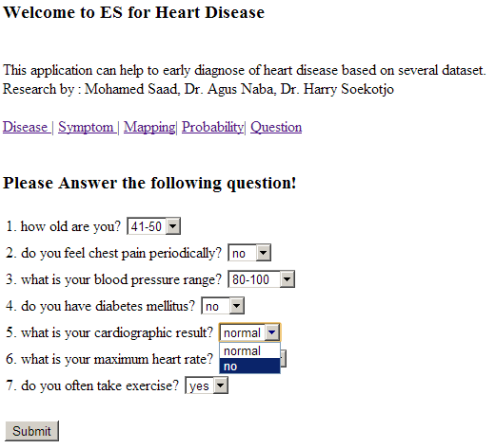

Fig. 14 Running the ES by Filtering cardiographic result

Fig.14 explain that the question of the cardiographic result contains options with 2 categories, which is 'normal' and 'no' answer of question. Users simply choose among the range to answer questions

K. Running the expert system (heart rate)

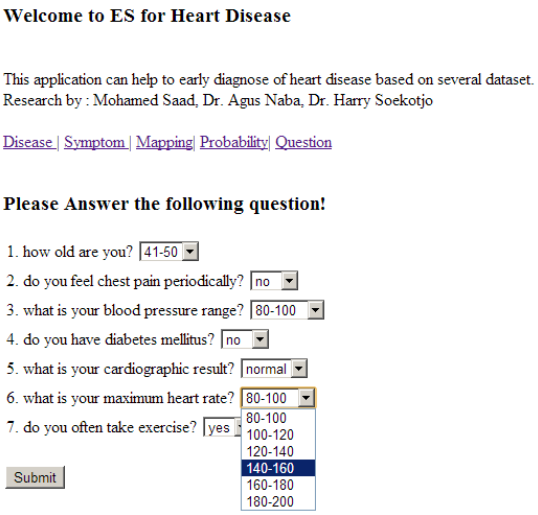

Fig. 15 Running the ES by Filtering maximum heart rate

Fig.15 explain that the question of the heart rate contains options with 6 categories, which range between the 80-100, 100-120, 120-140, 140-160, 160-180, and 180-200. Users simply choose among the range to answer questions.

L. Running the expert system (exercise)

Fig.16 explain that the question of the exercise result contains options with 2 categories, which is 'yes' and 'no' answer of question. Users simply choose among the range to answer questions

M. Result of Recommendation

After user answering all of the question, user will see the result of application expert system and the result can be seen in Fig. 17.

Application resultsprovide recommendations

ondecisions takenby the computerbased on thefactthatthe userhasentered.

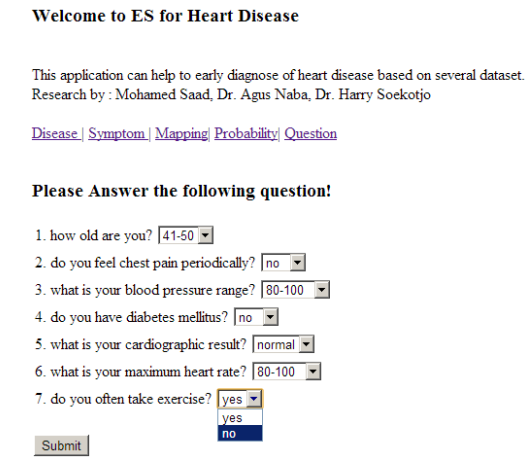

Fig. 16 Running the ES by Filtering exercise

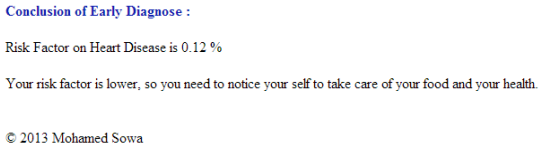

Fig. 17 Result of Recommendation

This result give to the user the risk of getting heart disease in the percentages.Preliminary results show the diagnosis on a person's risk of heart disease, based on advice from a doctor if the value of a percentage above 50%, the user is encouraged to immediately go to the doctor to check their health. This is done for the early detection of heart disease before a person having a heart attack.

N. Testing the Application

To get the best result, the application must be tested by developer. Testing task differentiate into three steps, verification, validation and Prototype Testing.

O. Verification

The goal of this test is to check whether the results meets design requirement. This test conducted by developer that produce verification document about the application compared to its design. In this application, use the PHP programming language and MySQL database.Although this database free but also reliable.

In the developing of software, a programmer must know how the software compilation steps, because the largest part of the programmer's attention is the problem of modularity, especially in the treatment of such software. This is expected not to the software that was made was not able or are too difficult to develop, it needs to be made clear, which helps not only from the programmer vision but on the developer as well.

The following forms are the example of the test.

P. Validation

The objectives of validation test are to check whether the recommendations of the software are the same as the recommendation of the expert [9]. The following table shows the test results.

TABLE IX
TEST OF APPLICATION

Facts	Recommendation
No information about the number of symptoms on the header menu	Add text that contain information total of symptoms on the left of symptoms menu
Question still in raw mode.	Add another color and change name of the field.
Symptoms table doesn't contain column number	Add one column which show the ordered number

TABLE X
VALIDATION RESULT

No	The patient sample	Symptoms	The doctor's recommendation	The ES solution
1	Patient 1	age, chest pain, blood pressure, diabetes mellitus, cardiographics, heart rate max, exercise	V	HD (34.7%)
2	Patient 2	(same above)	V	HD (2.9%)
3	Patient 3	(same above)	V	HD (13.6%)
4	Patient 4	(same above)	V	HD (0.8%)
5	Patient 5	(same above)	V	HD (23.4%)

It is important for users to know the outcome of the application and then consider all the recommendations made by the system. However, a clear validation of experts and doctors will also increase the trust to users using the application. Users will not doubt and they can make decisions on applications, because the purpose of this application is to assist users in diagnosing heart disease.

Q. Prototype Testing

TABLE XI
TESTING OF PROTOTYPE

Items	The rule base method	Experts System Software
Managing likelihood on Bayesian Network	Difficult	Easy
Running the Expert System & Consultation	Difficult	Easy
Speed/Quick Ability	Slow	Much Faster
Ability to exploit a considerable amount of knowledge	Need a lot of time	The same ability because using database
Reliability	Lower	Higher
Scalability	Difficult	Easy

Application testing software testing tested by the user and software analyst. If the functional and result of the

application are reliable, then this testing will get better result.

This test is conducted to compare between the experts system software and the rule base method in dealing with the disease. The Table XI shows the summary of the test.

V. CONCLUSION

After developing application for expert system to diagnose heart disease,the conclusionsare : (1) From the designing of expert system, dataset of heart disease research obtained from a previous study, consisting of 209 patients with different types of symptoms taken from Lyon University. This data has been analyzed to obtain the probability value used by the Bayesian network as node and applying in this research. (2) The development of the expert system application using php programming as web base language consists of several development steps: feasibility analysis, conceptual design, knowledge acquisition, knowledge representation, and validation. The result is a prototype of an expert system in early diagnose of heart disease which has been tested using verification, validation, and prototype, (3) Implementation of this research is an application that gives the user a few questions about user health information who subsequently processed by the system and processed using an expert system. The result is a percentage of a risk heart disease. According to the doctor's advice, if the percentage is more than 50%, the users are expected to check their health to the doctor. (4) In this research, expert has already been testing and analyzing the application and the application to deploy to the user on the website saadsowa.com. From the validation and deployment to the people, people can get a lot of benefit by using this application and the doctor said that this research can be improved again in the next research.

ACKNOWLEDGMENT

I would like to thanks to my supervisor, Agus Naba and Harry Soekotjo Dahlan. Thank you for everything.

REFERENCES

[1] WHO.(2008). *Heart disease facts*. Available : http://www.who.int/mediacentre
[2] Nordqvist, Christian.(2013). What is heart disease?, National Institute of Health. Available : http://www.medicalnewstoday.com/articles/237191.php
[3] Giarratano, R., Expert Systems Principles and Programming, United States : PWS Publishing, 1998.
[4] Rolston, D., Principles Artificial Intelligence And Expert Systems Development, Singapore : McGraw Hill, 1988.
[5] Klein, R., The Instructional Design Knowledge Base: Theory, Research, and Practice, British : Routledge, 2010.
[6] Parsaye, K. and Chignell, M., Expert Systems for Experts, United States : John Wiley, 1986.
[7] Jackson, P., Introduction to Expert System, United States : Addison Wesley, 1998.
[8] Negnevitsky, M,Artificial Intelligence, A Guide to Intelligent Systems, England : Addison-Wesley, 2005.
[9] Fuecks, H., The PHP Anthology, Australia : Sitepoint, 2003.
[10] Powers, D., PHP Solutions: Dynamic Web Design made easy, United States : friends of LED, 2010.
[11] Dubois, P., MySQL (4th Edition), United States : Addison-Wesley Professional, 2008.
[12] Williams, T., Learning MySql, United States : O'Reilly Media, 2006.
[13] Brachman, R., Levesque, H., Knowledge Representation and Reasoning, United States : Elsevier, 2004.
[14] Martin, J., Building Expert System: A Tutorial, United States : Prentice Hall, 1988.
[15] Rakotomalala, R., (2013). Pratique de la Regression, Universete Lumiere Lyon 2. Available : http://www.eric.univ-lyon2.fr
[16] Levesque, H., The Logic of Knowledge Bases, United States : the MIT Press, 2001.
[17] Mira, A. (1998).Methodology and Tools in Knowledge-Based Systems, 11th International Conference IEA, Spain.
[18] Gruber, T., The Acquisition of Strategic Knowledge, United States : Academic Press, 1989.
[19] Ullman, J., Principles of Database and Knowledge-Base System, Vol 1: Classical Database System,United States : Computer Science Press, 1990.
[20] Milton, N., Knowledge Acquisition in Practice: A Step-by-Step Guide (Decision Engineering), United States : Springer, 2010.