Rule Extraction for Fuzzy Expert System to Diagnose Coronary Artery Disease

Wiga Maulana Baihaqi
Department of Electrical Engineering and Information Technology
Universitas Gadjah Mada
Yogyakarta, Indonesia
wiga.mti15@mail.ugm.ac.id

Noor Akhmad Setiawan
Department of Electrical Engineering and Information Technology
Universitas Gadjah Mada
Yogyakarta, Indonesia
noorwewe@ugm.ac.id

Igi Ardiyanto
Department of Electrical Engineering and Information Technology
Universitas Gadjah Mada
Yogyakarta, Indonesia
igi@ugm.ac.id

Abstract*— In many cases, Coronary artery disease (CAD) results in death. CAD occurs when atherosclerosis blocks the bloodstream to the heart muscle in the coronary arteries. In order to diagnose CAD, there is a gold standard called angiography. However, this method is invasive, risky and expensive. Many researches had been successfully conducted to solve the above problem of angiography by using the data mining technique and fuzzy expert system. This research compares the performance of C4.5, CART, and RIPPER as a fuzzy rules generator to be used on the fuzzy expert system. The results show that the combination between algorithm C4.5 and fuzzy expert system achieves the highest accuracy of 81.82%.

Keywordsi—Coronary Artery Disease (CAD); data mining; fuzzy expert system; rule; fuzzy membership functions

I. INTRODUCTION

Heart and blood vessels diseases are classified as *cardiovascular disease* (CVD), which consists of hypertension, coronary heart disease, cerebrovascular disease (stroke), peripheral vascular disease, heart failure, rheumatic heart disease, congenital heart disease, cardiomyopathies [1]. Based on the data from the World Health Organization (WHO), the primary cause of death in the world is cardiovascular, every year cardiovascular mortality accounted for the highest compared to other causes. In 2012, WHO estimated the number of people died from CVDs as many as 17.5 million people, the entire global death represented by the number of deaths due to CVDs as much as 31%. Amongst these deaths, coronary heart disease is estimated to cause 7.4 million deaths and strokes caused 6.7 million people died. Based on these data, it states that the cardiovascular disease which causes most deaths in 2012 in the world is the artery disease or coronary heart disease that is followed by stroke [1]. Based on the results of basic medical research 2013 [2] based on the interviews, the doctor-diagnosed prevalence of heart disease in Indonesia is 0,5 percent, and the doctor- or symptoms-diagnosed is 1.5 percent. The highest prevalence of doctor-diagnosed coronary heart disease is in Central Sulawesi (0,8%), followed by North Sulawesi, Jakarta, Aceh of 0.7 percent each. The highest prevalence of diagnosed or symptoms-diagnosed coronary heart disease is in East Nusa Tenggara (4,4%), followed by Central Sulawesi (3,8%), South Sulawesi (2,9%), and West Sulawesi (2,6%).

A disease in which the arteries in the heart there is a waxy substance called plaque builds up called with coronary artery disease, whereas the heart muscle supplied oxygen-rich blood to the arteries, so that it is accumulated in the arteries that impairs the blood supply [3]. Many factors that cause plaque builds up in the coronary arteries, the controllable factors; cholesterol, smoking or exposure to cigarette smoke, high blood pressure, obesity, diabetes, depression, and the unchangeable or uncontrollable factors; lineage/family history, age, and gender (more at risk of coronary heart disease in men than women) [4].

The patient's medical and family history, risk factors for coronary heart disease, physical examination, and the results of tests and procedures serve as the basis for the diagnosis of coronary heart disease in a patient. To perform the test, there are many factors that will be considered, which makes the detection difficult and time consuming. Some tests are recommended, such as ECG, Stress Testing, Echocardiography, Chest X-Ray, Blood Test, Coronary Angiography and Cardiac Catheterization that have the accurate results in decision-making [5], but in terms of financing in the treatment of heart disease, it requires considerable time and cost. The cost of the treatment for heart disease is estimated to US $ 14 billion (approximately 42 trillion rupiahs) spent each year for the treatment of coronary heart disease in the United States. [6]. It becomes an obstacle in the treatment of heart disease in Indonesia as a developing country.

In recent decades, some experts have tried to use computer assistance techniques for diagnosing CAD (Coronary Artery Disease), especially the use of fuzzy expert systems, fuzzy expert system is one important application in fuzzy logic. Fuzzy logic has an important role in medicine, to diagnose medical problems effectively, because the fuzzy logic is conceptually simple, easy to understand, and behind the fuzzy logic has a mathematical concepts very easily [7]. Fuzzy rule is one important part of the fuzzy system, fuzzy rule determining the quality of a fuzzy system. There are several ways to generate rules that are used for fuzzy systems, the rules produced by the experts or doctors in a laboratory or hospital and the rules

produced by the classification of patient data with data mining algorithms.

In the research conducted by Adeli and Neshat [8], Kumar and Kaur [9], Barman and Choudhury [10], Pal et al [11], the rules obtained from a doctor or hospital or laboratory were used, however the designed system is highly dependent on the ability of doctors or specialists. On the other hand, research [12] used fuzzy and genetic algorithms to optimize the fuzzy membership function, researches [13][14] employed fuzzy and PSO to optimize the fuzzy membership function, and research [15] used fuzzy and ICA to optimize fuzzy membership function. The studys only use the classification results of the CART decision tree which is then extracted into rules of "if-then" and transformed to generate fuzzy rules.

This study uses C4.5, CART, and RIPPER to generate fuzzy rules and imperialist competitive algorithm (ICA) for the optimization of fuzzy membership functions. fuzzy rules and fuzzy membership functions that have been optimized are used in fuzzy expert system for diagnosing coronary heart disease. There are four steps in this study, the first is data preprocessing that consists of filling missing value, removing outlier, and normalization. The second step is using C4.5, CART, and RIPPER to generate fuzzy rules extracted from the classification results. The third step, the rule is transformed into a fuzzy set. Lastly, imperialist competitive algorithm is used to optimize the fuzzy membership functions.

II. MATERIAL AND METHODS

A. Data Set of Coronary Artery Disease

Data sets from Hungarian Institute of Cardiology, Budapest, and the Cleveland Clinic datasets from the University of California at Irvine were used to evaluate the proposed method. There are 294 records from the data sets from Hungarian Institute of Cardiology, Budapest and 303 records from Cleveland data sets, so there were 597 records used in this research. From the attributes used by most of the researchers, in which there were 14 attributes that consists of 13 input and 1 output attributes, although in fact there were 76 attributes in the data sets. 13 input attributes consist of age, blood pressure, serum cholesterol, maximum heart rate, sex, type of chest pain, fasting blood sugar, resting ECG, exercise-induced angina, oldpeak, slope, fluoroscopy, and thallium scan. The output variable is the angiography status.

There is a difference in the angiography status between the data sets of Hungarian and Cleveland. The data sets of Hungarian have the angiography status (0 and 1), while Cleveland (0, 1, 2, 3, and 4), so that the angiography status (1, 2, 3, and 4) for data sets of Cleveland are converted into (1), while the angiography status (0) is fixed.

B. Data Preprocessing

Data preprocessing is the first step in this study. There are a series of steps used in data prerpocessing: filling the missing values, dealing with outliers, normalizing the data and attribute selection. Missing data is an information that is not available for an object. It may occur because the value is not relevant to the particular case, and it could not be recorded when the information is gathered, or overlooked by clients because of protection concerns. Extracting useful information from the data set will be difficult when there are missing values in the data set [16]. There are some methods can be used to handling missing data such as removing attributes, including missing data, and the data imputation. The first and second method can be used in the research when there are a small number of missing values in the data set used. This study used the last method because there are a lot of missing data in the data set used. There are two techniques that can be used in these methods, the statistical technique and machine learning technique [17]. This study uses statistical techniques, mode used to replace the categorical values, and mean to replace the numerical values [18].

Statistics-based approach used to detect outliers in previous studies, in which tuples are modeled as distributions and outliers are tuples that show significant deviations from the distribution assumed. However, such an approach can not build a proper model of high-dimensional data [19]. Tthis study used distance-based outliers method proposed by Knorr and Ng to detect outliers, the k-nearest neighbor and euclidean distance were used [20].

In the normalization step, this study used a method that can change all of the data range to be between 0 and 1, the equation was used to normalize the data can be seen as follows

$$\text{Normalize } (e_i) = \frac{e_i - E_{min}}{E_{max} - E_{min}}, \quad (1)$$

where

E_{min} = the minimum value for variable E.

E_{max} = the maximum value for variable E.

If E_{max} is equal to E_{min} then normalized (e_i) is set to 0.5.

C. Feature Selection

The main purpose of the selection of the feature is to specify the minimum features of a subset of the domain of the problem while retaining high accuracy appropriate in representing the original features. In the real world, the choice of features is a must due to the large number of noise, irrelevance, or misleading the features [21]. In this research, the feature selection is using C4.5, CART and RIPPER algorithm. In addition to a selection of features used to produce features that has significant influence to improve accuracy, it is also used to generate an if-then rule. Decision tree is selected because it has some benefits, that are easy to understand, and easy to be transformed into classification rules [15]. While RIPPER algorithm does not need to be transformed into an if-then rule, because the result of classification of RIPPER algorithm is in the form of if-then rule [22].

D. C4.5

C4.5 is one of the algorithms of the decision tree, C4.5 has been proposed by J. R. Quinlan in 1987 to improve the weaknesses of ID3 [23] which is not good in dealing with the continuous valued features. Decision tree will display knowledge using graphs such as trees or decision model. Each node in the tree represents a decision, while branch represents the possible result. Leaf node is a sample class. The advantage of the decision tree is that it is simple and easy to understand [24]. C4.5 divides the real value in an appropriate interval based

on the ratio to obtain the information which is the ratio between the gain information and intrinsic value defined as (2) :

$$\text{GainRatio} = \frac{\text{Gain(S,A)}}{\text{SplitInfo(S, A)}}, \quad (2)$$

where SplitInfo(S, A) denotes the split's information defined as (3) and Gain (S, A) as (4) :

$$SplitInfo(S,A) = \sum_{i=1}^{m} \frac{|s_i|}{|s|} \; x \; log^{\frac{|s_i|}{|s|}}, \quad (3)$$

where m is the total of feature A values, Si is the total of samples in the subsets resulted from the original separating defined in accordance with the features of A and S that indicate the sample in the defined original. Splitlnfo is the output information generated by the testing dataset D when D is divided into part v according to the attribute A. v is a part of the values. It can represent a broad and uniform data sets S Split by attributes A, C4.5 can handle continuous values by using

$$Gain(S,A) = Entropy(S) - \sum_{V_e}^{Value(A)} \frac{|S_v|}{S} * Entropy(S_v), \quad (4)$$

where Sv shows the subset of S for which attributes A that have a v value, S is the set of data that is currently being counted, entropy is s attribute and value (A) of all of the A attributes. As a result, a feature by getting greater information is preferred, and Entropy(S) defined as (5) :

$$\text{Entropy(S)} = -\sum_{i=1}^{n} P(u_i) \log^{P(u_i)}, \quad (5)$$

where P (u_i) is the possibility of the emergence of features i, and n represents the total of samples. the larger the entropy data the more the clutter.

E. CART

CART is part of the decision tree algorithm, in addition to C4.5. CART was firstly developed by Leo Breiman, Jerome H Friedman, Richard A. Olshen and Charles J. Stone in 1980. Decision tree will be generated by the CART if the output variable is in the form of category and will generate regression trees if the output variable is a numeric [25].

CART measures the level of impurity for the given data and the construction of a binary tree where each node is an internal output exactly for two classes for the given attribute. Gini index is calculated for each attribute and then attribute with the lowest gini index selected as attributes of the divide. The tree is built recursively to select attribute with the lowest gini index. If the probabilities of pi for the class of k target for certain attributes, then the Gini index is defined as

$$\text{Gini(t)} = 1 - \sum_{j=1}^{J} P_j^2, \quad (6)$$

where pi is the probability of class i.

F. RIPPER

Ripper (Repeated Incremental Pruning to Produce Error Reduction) is the learner based on rules that establishes the set of rules to identify the class with minimize the number of errors [26]. RIPPER was the development of the algorithm developed by REP. William d. Cohen in 1995 [27].

G. Fuzzy Expert System

The expert system as a computer application program is an artificial intelligence (AI) expert system. The role of expert system is to emulate a human expert decision-making process. Fuzzy logic is a method, which is called as the theory of obscurity, problem solving since it does not remain true and false traditional concepts of logic reasoning, it is associated with truth values ranging between 0 and 1. Fuzzy sets were first developed by Zadeh in 1965 [28]. Inaccuracy input can be handled by obtaining the degree of membership value for each of the functions defined in fuzzy expert systems, fuzzy set of equations can be seen in as follows

$$A = \left(x.\ \mu_A(x)\right) \ |x \in X.\ \mu(x) : X \rightarrow [0.1], \quad (7)$$

where A is the fuzzy sets, μA is the membership function, and X is the universe of discourse.

Fuzzy expert system is a knowledge-based system in which fuzzy logic is used as a tool to develop the relationship between the input and output data. Fuzzy expert system is one of the important applications of fuzzy logic, in medicine, fuzzy logic plays an important role for effective diagnosis of medical problems because Fuzzy Logic (FL) is conceptually simple and easy to understand and the mathematical concepts behind fuzzy logic is very easy. Besides being used to diagnose coronary heart disease, fuzzy expert system has been used to diagnose thyroid disease [29], to detect the level of autism [30], to diagnose multiple sclerosis [31].

The decision tree obtained from C4.5 algorithm and CART, was extracted into the "if-then" rules which would then be transformed into fuzzy rules based on fuzzy membership function, while the results obtained from RIPPER algorithm does not need to be extracted into "if-then" rules for transformed into the fuzzy rule. Prior to transform "if-then" rules into a fuzzy rule, the fuzzy models must be designed.

Fuzzy expert system has three main parts, they are; fuzzification, inference and defuzzification. Sugeno fuzzy inference system has been used in this research. The process of fuzzification is a transformation crisph input into a set of fuzzy input, the values of the input variables are used to determine the level of these values suitable for each membership functions used by fuzzy rules, variables or attributes used to build the fuzzy membership functions are determined from the results of feature selection by using C4.5, CART, and RIPPER algorithms. There are some types of membership function, such as triangular, trapezoidal, and gaussian, This research used a triangular membership function. The specification of the membership function is defined as

$$\text{triangle}(x : a, b, c) = \max(\min(x-a)/(b-a),(c-x)/(c-d)),0)) \quad (8)$$

where the parameters a and b denote the lower and upper bounds respectively while c locates the peak of the triangle.

The inference process that used fuzzy input is defined to determine the fuzzy output set using fuzzy rules defined in the knowledge base and membership functions. The fuzzy rules

used are the results of the transformation rules generated from C4.5, CART and RIPPER algorithms to be the fuzzy rules. The process of defuzzification counted the crisp output using fuzzy sets generated by the process of composition, weighted average method has been used in this research, the equation of the weighted average method is defined as

$$Z^{*}=\frac{\sum \mu_c(\bar{z}) \cdot \bar{z}}{\sum \mu_c(\bar{z})}, \quad (9)$$

where $\bar{z}$ Is the centroid of each symmetric membership function. Fig. 3 illustrates the fuzzy expert system proposed, prior to the data entered into the system of fuzzy inference and then going on classified data entered in the preprocessing step that consists of filling in missing data, removing outliers, and normalization.

H. Membership Function Optimization

In the fuzzy model there are very important phase to be designed: the fuzzy membership functions and fuzzy rules, because the system performance is directly affected by both. Membership functions can be optimized by using an evolutionary algorithms including intelligent methods, i.e. imperialist competitive algorithm and particle swarm optimization.

In 2007, ICA was first proposed. Imperialist competition inspired the formation of ICA. As with other evolutionary algorithms, starting with the initial population. The individual residents of the mentioned countries, consist of two types: colonies and imperialism, which together form the kingdoms [32]. The competition between these kingdoms forms the basis of the ICA. Powerful ones will dominate the weak and crumbling empire during this competition. There will be only one empire remaining after this algorithm ends. ICA was first used to diagnose coronary artery disease carried out in 2014 [15], ICA was used to optimize the fuzzy membership functions.

In addition to ICA, the PSO algorithm was developed by Kennedy and Eberhart in 1995 also refers to a family that is relatively new on the evolutionary algorithms, and inspired by the social behavior of certain animals, such as flocking of birds, schools of fish, swarms of bees, and even human behavior [33]. In contrast to the ICA, the PSO is already widely used to diagnose coronary heart disease, research [13] [34] used PSO to optimize fuzzy membership functions, a research by Hedeshi et. Al. [34] used PSO-based approach of the new ensemble to extract a set of rules for the diagnosis of coronary heart disease. The research by Muthukaruppan And Er [13] designed a computerized method to predict heart disease with the help of data mining and optimization techniques such as Genetic and PSO.

Each triangular fuzzy membership function has three parameters. Parameters Membership is shown in figure 1: C (center), L (left) and R (right) in accordance with the membership function originals, where the center, left and right are referred by C', L', and R' of the customized membership functions respectively.

The membership function of the fuzzy adjusted with the following equation:

$$C' = (C + k_i) - w_i \quad (10)$$

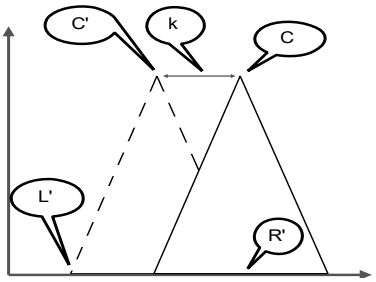

Fig. 1. Triangular Fuzzy Membership Function Parameters, adapted from [13]

$$L' = (L + k_i) - w_i \quad (11)$$

$$R' = (R + k_i) - w_i \quad (12)$$

where, adjustment coefficient is represented by ki and wi. Ki moves each membership function to the left or right. wi parameter passed to shrinkage and expansion of the membership function. The values of the ki and wi parameters were found by using ICA.

The flowchart of step by step in the proposed system in this research is illustrated in figure 2.

III. Result and Discussion

MATLAB 7.12 (R2011a) was used to design the proposed fuzzy system. The rules were obtained from the C4.5, CART, and RIPPER algorithms using weka 3.6. Fuzzy rules obtained from the C4.5, CART, and RIPPER algorithms are counted to 76, 8, 10 respectively. To generate the rules, the training data amounted 411 were used, with 122 healthy instances, and 289 instances with coronary heart disease.

The testing data amounted 176 are used to evaluate the designed fuzzy system. Confusion matrix is used to evaluate the performance of the designed fuzzy system, as follows

$$CM=\begin{pmatrix} TP & FP \\ FN & TN \end{pmatrix} \quad (13)$$

where TP (True Positive) is the sum of the results of the prediction shows that the patients are sick and the data actually show that the patients are sick. TN (True Negatives) is the sum of the results of the prediction shows that the patients are not sick and the data actually showed no pain. FN (False negatives) is the sum of the results of the prediction shows that the patients are not sick, but the actual data of the patients are sick. FP (False Positives) is the sum of the results of the prediction shows that the patients are sick, but actual data of the patients are health.

With the Confusion Matrix (CM) system performance can be calculated in terms of accuracy, sensitivity, specifity, and precision, as follows

$$\text{Accuracy}=\frac{(TP+TN)}{(TP+TN+FN+FP)} \quad (14)$$

$$\text{Sensitivity (recall)}=\frac{TP}{(TP+FN)} \quad (15)$$

$$\text{Specificity}=\frac{TN}{(TN+FP)} \quad (16)$$

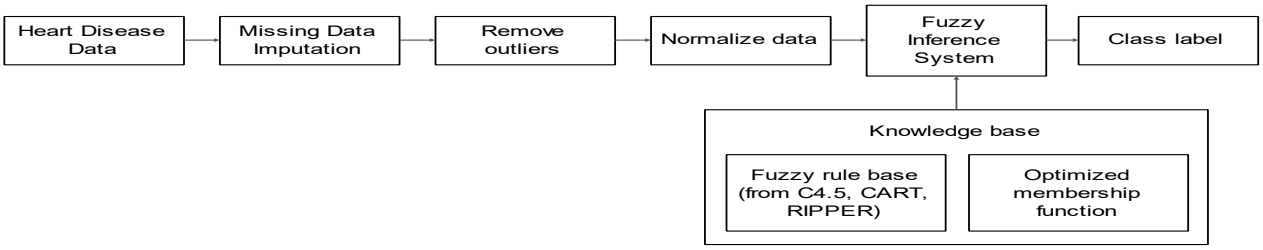

Fig. 2. Proposed Fuzzy Expert System

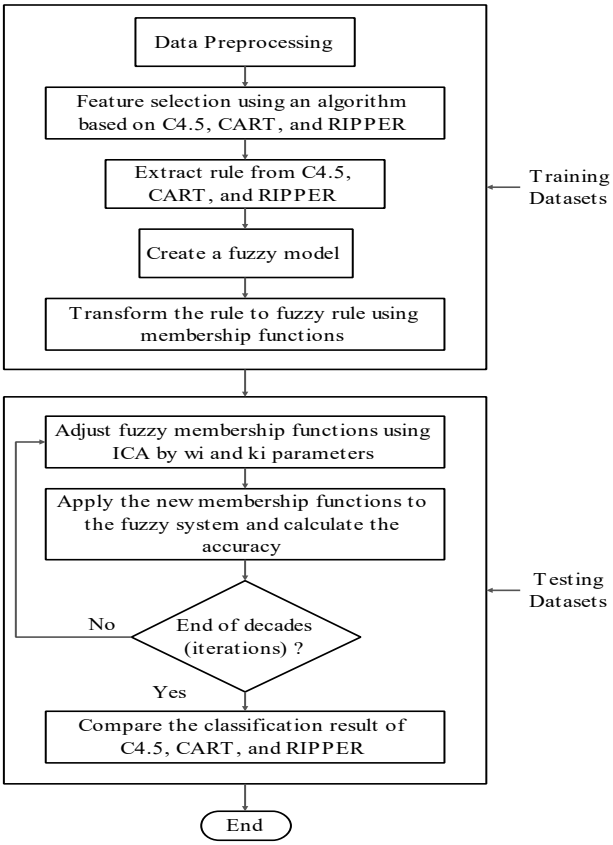

Fig. 3. Flowchart of Step by Step in the Proposed System

$$\text{Precision}=\frac{TP}{(TP+FP)} \quad (17)$$

$$\text{F-measure}=2*\frac{\text{Precision}*\text{Recall}}{\text{Precision}+\text{Recall}} \quad (18)$$

Table 1-3 illustrates the confusion matrix from the results of the classification of fuzzy system, fuzzy rules derived from C4.5, CART, and RIPPER algorithms. Table 4 illustrates the comparison of the performance of the proposed fuzzy system, fuzzy rules which derived from C4.5, CART, and RIPPER algorithms.

TABLE I. CONFUSION MATRIX OF FUZZY SYSTEM, FUZZY RULES FROM C4.5

	Heart disease	Normal
Heart disease	60	16
Normal	16	84

TABLE II. CONFUSION MATRIX OF FUZZY SYSTEM, FUZZY RULES FROM CART

	Heart disease	Normal
Heart disease	31	45
Normal	4	98

TABLE III. CONFUSION MATRIX OF FUZZY SYSTEM, FUZZY RULES FROM RIPPER

	Heart disease	Normal
Heart disease	50	26
Normal	15	85

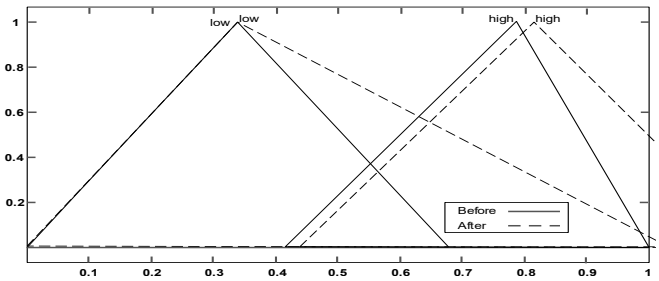

Fig. 4. The Membership Functions of oldpeak (Fuzzy rule base RIPPER)

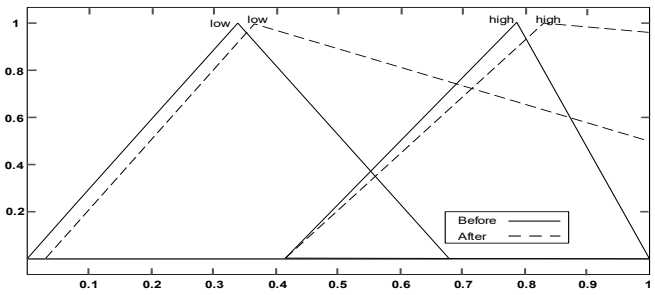

Fig. 5. The Membership Functions of oldpeak (Fuzzy rule base CART)

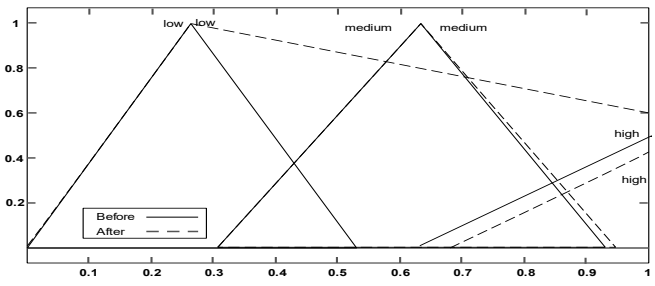

Fig. 6. The Membership Functions of thalac (Fuzzy rule base C4.5)

TABLE IV. COMPARISON OF FUZZY SYSTEM PERFORMANCE

Fuzzy Rule Based	Accuration	Sensitivity	Specificity	Precision	F-measure
C4.5	81.82 %	78..95 %	84.00 %	78.95 %	78.95 %
RIPPER	76.70 %	76.92 %	76.58 %	65.79 %	70.92 %
CART	72.47 %	88.57 %	68.53 %	40.80 %	55.86 %

IV. CONCLUSION

The combination of data mining and fuzzy expert systems have been successfully carried out in this research to diagnose coronary heart disease. The comparison between the data mining algorithms are C4.5, CART, and RIPPER has been successfully carried out, as shown in Table 4. It shows the highest accuracy generated by C4.5 as the algorithm used to extract the rules used in fuzzy expert system. It has been proved that the rule is an important part in the fuzzy expert system, a good rule will generate good accuracy on the application of fuzzy expert system. In addition, this research also successfully applied Imperialist Competitive Algorithm (ICA) as an algorithm to optimize the fuzzy membership functions.

ACKNOWLEDGMENT

The research was supported by Intelligent System group research of Electrical Engineering and Information Technology Department, Gadjah Mada University.

REFERENCE

[1] WHO, "Cardiovascular diseases (CVDs)," *Nov*-2015. [Online]. Available: http://www.who.int/mediacentre/factsheets/fs317/en/. [Accessed: 19-Nov-2015].

[2] Agency for Health Research and Development of the Republic of Indonesia, "Riset Kesehatan Dasar 2013," The Health Ministry of the Republic of Indonesia, 2013.

[3] National Heart, Lung, and Blood Institute, "What Is Coronary Heart Disease?," *Nov*-2015. [Online]. Available: http://www.nhlbi.nih.gov/health/health-topics/topics/cad. [Accessed: 20-Nov-2015].

[4] Heart Foundation, "Coronary heart disease.," National Heart Foundation of Australia , 2013.

[5] National Heart, Lung, and Blood Institute, "How Is Coronary Heart Disease Diagnosed?," Nov-2015. [Online]. Available: http://www.nhlbi.nih.gov/health/health-topics/topics/cad/diagnosis. [Accessed: 18-Nov-2015].

[6] G. H. Martono and T. B. Adji, "Penggunaan Principal Component Analysis dan Pohon Keputusan untuk Mendeteksi Penyakit Jantung Koroner," Unpublished thesis, Dept. Elect. Eng., Universitas Gadjah Mada, Yogyakarta, 2012.

[7] N. Mishra, "A Review on the Applications of Fuzzy Expert System for Disease," *Int. J. Adv. Res. Eng. Appl. Sci.*, vol. 3, no. 12, pp. 28–43, 2014.

[8] A. Adeli and M. Neshat, "A Fuzzy Expert System for Heart Disease Diagnosis," *Proc. Int. MultiConference Engineeers Comput. Sci.*, vol. I, pp. 1–6, 2010.

[9] S. Kumar and G. Kaur, "Detection of Heart Diseases using Fuzzy Logic," *Int. J. Eng. Trends Technol.*, vol. 4, no. 6, pp. 2694–2699, 2013.

[10] M. Barman, "A Fuzzy Rule Base System for the Diagnosis of Heart Disease," *Int. J. Comput. Appl.*, vol. 57, no. 7, 2012.

[11] D. Pal, K. M. Mandana, S. Pal, D. Sarkar, and C. Chakraborty, "Fuzzy expert system approach for coronary artery disease screening using clinical parameters," *Knowledge-Based Syst.*, vol. 36, pp. 162–174, 2012.

[12] Y. N. Devi and S. Anto, "An Evolutionary-Fuzzy Expert System for the Diagnosis of Coronary Artery Disease," vol. 3, no. 4, pp. 1478–1484, 2014.

[13] S. Muthukaruppan and M. J. Er, "A hybrid particle swarm optimization based fuzzy expert system for the diagnosis of coronary artery disease," *Expert Syst. Appl.*, vol. 39, no. 14, pp. 11657–11665, 2012.

[14] P. P. I, P. Gayathri, and N. Jaisankar, "A Fuzzy Optimization Technique for the Prediction of Coronary Heart Disease Using Decision Tree," vol. 5, no. 3, pp. 2506–2514, 2013.

[15] Z. Mahmoodabadi and M. S. Abadeh, "CADICA : Diagnosis of Coronary Artery Disease Using the Imperialist Competitive Algorithm," *J. Comput. Sci. Eng.*, vol. 8, no. 2, pp. 87–93, 2014.

[16] V. H. Umathe and G. Chaudhary, "Imputation methods for incomplete data," *2015 Int. Conf. Innov. Information, Embed. Commun. Syst.*, pp. 1–4, 2015.

[17] S. Sajja, "Data Mining of Medical Datasets with Missing Attribute from Different sources," Youngstown State University, 2010.

[18] S. Sajja, "Data Mining of Medical Datasets with Missing Attributes from Different Sources," Youngstown State University, 2010.

[19] M. Bai, X. Wang, J. Xin, and G. Wang, "An efficient algorithm for distributed density-based outlier detection on big data," *Neurocomputing*, vol. 181, pp. 19–28, 2015.

[20] E. M. Knorr and R. T. Ng, "Algorithms for Mining Datasets Outliers in Large Datasets," *24th Int. Conf. Very Large Data Bases*, pp. 392–403, 1998.

[21] M. S. Raza and U. Qamar, "An incremental dependency calculation technique for feature selection using rough sets," *Inf. Sci. (Ny).*, vol. 343–344, pp. 41–65, 2016.

[22] J. Han, M. Kamber, and J. Pei, *Data mining: concepts and techniques*. 2012.

[23] J. R. Quinlan, "Simplifying decision trees," *Int. J. Man. Mach. Stud.*, vol. 27, no. 3, pp. 221–234, 1987.

[24] H. Qian and Z. Qiu, "Feature Selection Using C4.S Algorithm For Electricity Price Prediction," *Int. Conf. Mach. Learn. Cybern.*, 2014.

[25] A. F. A. Pinem and E. B. Setiawan, "Implementation of Classification and Regression Tree(CART) and Fuzzy Logic Algorithm for Intrusion Detection System," *J. Chem. Inf. Model.*, vol. 53, pp. 1689–1699, 2013.

[26] B. Seerat and U. Qamar, "Rule Induction Using Enhanced RIPPER Algorithm for Clinical Decision Support," *Int. Conf. Intell. Control Inf. Process.*, vol. 33, 2015.

[27] W. W. Cohen, "Fast effective rule induction," *Proc. Twelfth Int. Conf. Mach. Learn.*, pp. 115–123, 1995.

[28] L. A. Zadeh, "Fuzzy sets," *Inf. Control*, vol. 8, no. 3, pp. 338–353, 1965.

[29] S. A. Biyouki, "Fuzzy Rule-based Expert System for Diagnosis of Thyroid Disease," *Conf. Comput. Intell. Bioinforma. Comput. Biol.*, 2015.

[30] N. R. M. Isa, M. Yusoff, N. E. Khalid, N. Tahir, and A. binti Nikmat, "Autism severity level detection using fuzzy expert system," *Robot. Manuf. Autom. (ROMA), 2014 IEEE Int. Symp.*, pp. 218–223, 2014.

[31] M. A. Ghahazi, "Fuzzy Rule based Expert System for Diagnosis of Multiple Sclerosis," pp. 1–5, 2014.

[32] E. Atashpaz-Gargari and C. Lucas, "Imperialist competitive algorithm: An algorithm for optimization inspired by imperialistic competition," *2007 IEEE Congr. Evol. Comput.*, pp. 4661–4667, 2007.

[33] Q. Bai, "Analysis of Particle Swarm Optimization Algorithm," *Comput. Inf. Sci.*, vol. 3, no. 1, pp. 180–184, 2010.

[34] N. G. Hedeshi and M. S. Abadeh, "An Expert System Working Upon an Ensemble PSO-Based Approach for Diagnosis of Coronary Artery Disease," no. December, pp. 14–16, 2011.