

A Tiered Approach on Dimensional Reduction Process for Prediction of Coronary Heart Disease

Wiharto Wiharto¹, Herianto Herianto², Hari Kusnanto³

¹Department of Informatic, Sebelas Maret University, Indonesia

²Department of Mechanical & Industrial Engineering, Gadjah Mada University, Indonesia

³Department of Medicine, Gadjah Mada University, Indonesia

Article Info

Article history:

Received Nov 24, 2017

Revised Jan 9, 2018

Accepted Apr 21, 2018

Keywords:

Dimensional Reduction

Diagnosis

Coronary heart disease

Likelihood ratio

ABSTRACT

The use of dimensional reduction in the diagnostic system model of coronary heart disease, many same of case do not take into account the clinical procedures commonly used by clinicians in diagnosis. This requires that the examination be done thoroughly, thus making the high cost of diagnosis. This study aims to develop a tiered approach model in reducing dimensions for predicting CHD. The method in this research is divided into several stages, namely preprocessing, building the knowledge base and system testing. Preprocessing consists of several processes, namely the removal of missing value data, grouping attributes, and dividing data for training and testing. Knowledge base modeling is divided into three levels. The first level were the risk factor attributes, the second level were the type of chest pain & ECG, and the third were scintigraphy & coronary angiography. The knowledge base was modeled based on fuzzy rules and its inferencing process using Mamdani method. The first, fuzzy rule-based was obtained by using the FRS study. The second and third stage, using the induction rule algorithm to get the rule, then converted to fuzzy rule. The tested algorithm were C4.5, CART, and FDT. The system testing was performed by the 5-folds cross-validation method, with performance parameters based on population and individual. The test resulted using the Cleveland and Hungarian datasets, the FRS+CART combination was capable of reducing the most attributes and the highest likelihood ratio performance parameter, which was 15.96. FRS+C4.5, at least the attributes were reduced, but has an AUC performance of 80.43%, while FRS+FDT, more reduced attributes than FRS+C4.5, and AUC performance parameters are better than FRS+CART. Dimensional reduction model for prediction of CHD, capable of providing better performance than not tiered.

*Copyright © 2018 Institute of Advanced Engineering and Science.
All rights reserved.*

Corresponding Author:

Wiharto Wiharto,
Department of Informatic,
Sebelas Maret University, Indonesia.
Email: wiharto@staff.uns.ac.id

1. INTRODUCTION

The statement from WHO that cardiovascular disease of the heart and blood vessels, especially coronary heart disease (CHD) was still ranked as the leading cause of death in developing countries until 2020. The disease can be prevented by making an early diagnosis. Unfortunately, the cost of a thorough examination for the diagnosis of heart disease is relatively expensive for the size of the developing country. The high cost of such diagnosis brings an economic impact on a country, namely the decline in productivity of the population, a country. These conditions encourage the development of clinical decision support system models for the diagnosis of artificial intelligence based coronary artery disease, which takes into account clinical costs and procedures.

The system model of diagnosis on coronary heart disease is widely developed by using a combination of dimensional reduction and classification. Dimensional reduction in the diagnostic system model is divided into two, namely dimensional reduction taking account of cost and not considering cost. The model of the diagnostic system with no cost consideration, has been largely done from the time period from 2009-2017. Research conducted by Liu et.al [1] proposed the dimensional reduction model, by combining ReliefF and Heuristic Rough Sets algorithms. The study has a number of weaknesses. First, the attributes generated in the dimensional reduction process, when referring to the costing groupings of Feshki and Shijani's [2] research results, there are two costly checks, namely scintigraphy and fluoroscopy. A similar result is also made by Ahmadi et.al [3], a combination model of C5.0 with a neural network, which does not reduce both the costly attributes. Second, dimensional reduction process does not take into account the clinical procedures that clinicians normally perform.

A system of diagnostics with cost-based dimension reduction has been performed by Arjenaki et.al [4] and also by Feshki and Shijani [2]. Both of these studies in dimensional reduction using a genetic algorithm and particle swarm optimization with its fitness function is a cost function. Both studies are capable of reducing costly attributes. Unfortunately the model does not consider the stages of examination attributes, consequently at the time of the diagnosis process requires examination of all attributes dimensional reduction results. This makes the absence of a reduction process again when the diagnosis is done. Dimensional reduction models capable of reducing costly attributes have also been undertaken by Wiharto et.al [5]. The study used a combination of tiered approaches with logistic regression algorithms, unfortunately, the tiered approach is only used in dimensional reduction process, but at the time of the diagnosis process is not used, so no dimensional reduction occurs when the diagnosis system is used.

The use of tiered concepts has also been done by Wiharto et.al [6]. The study grouped the attributes according to the stages in the process of diagnosis, but in each stage without dimensional reduction process. Dimensional reduction process occurs only when the diagnostic process is done, ie if the diagnosis in the early stages has been declared negative, no further examination of the attributes at the next stage is required. The same concept has also been used by Wiharto et.al [7], the study used the C4.5 algorithm for induction of the rule as well as dimensional reduction. Unfortunately, in the dimension reduction results generated when using the C4.5, only one attribute is reduced. This makes the model less able to reduce the cost of the examination, especially attributes that costly in the examination. Another disadvantage is that the performance measurement is only population-based, whereas individual-based [8] with likelihood ratio parameters has not been done, but the testing method does not use k-folds cross-validation.

Referring to a number of studies that have been developed, most dimensional reduction occurs only when prior to classification, while the dimensional reduction at the time of the diagnostic system which used was still minimal. In addition, these studies of reduced attributes were still small, resulting in costly consequences that must be incurred in the diagnostic process. Furthermore, previous studies have mostly conducted measurements of population-based on performance. There was still minimal individual-based on performance measurement, that likelihood ratio parameter. Based on this, in this study proposed a dimensional reduction model that occurred when dimensional reduction process and diagnostic process. This model can make the system model more provide cost savings in the diagnosis of CHD.

2. RESEARCH METHOD

2.1 Material and Data

The study was developed using the Cleveland and Hungarian datasets of the UCI repository, which can be accessed online [9]. The dataset used consists of 14 attributes, which consist of 13 independent attributes and 1 dependent attribute. The complete attributes are shown in Table 1. The number of the Cleveland datasets was 303, while the Hungarian 294. The dataset has two outputs, healthy which symbolized by 0, and sick symbolized by 1.

Table 1. Attributes in the diagnosis of coronary heart disease

No	Attributes	Description
1	age	Age
2	sex	Gender
3	cp	Chest pain type
4	restbps	Resting systolic blood pressure
5	chol	Cholesterol in mg/dl
6	fbs	Fasting blood sugar
7	restecg	Resting ECG
8	thalac	Maximum heart rate achieved
9	exang	Exercise induced angina
10	oldpeak	ST depression induced by exercise relative to rest
11	slope	The slope of the ST segment for peak exercise
12	ca	Number of major vessels colored by angiography Coronary (0-3)
13	thal	Defect type (Scintigraphy)
14	output	Level Heart disease (0/1)

2.2 Method

The method used in this study was as shown in Figure 1, which was divided into three parts. The first is preprocessing that performs the missing value data removal process, attributes grouping, data sharing into training and testing data, with 5-folds cross-validation method. The grouping of attributes was divided into three groups. The first groups are risk factors, both chest pain & ECG, and scintigraphy & coronary angiography. The second is the process of building a knowledge base for each attribute group. The three groups at the same time reflect the examination level. Development of knowledge base for level 1, by modeling table score in study Framingham risk score(FRS)[10], into fuzzy rule-based. In the second and third level, it was preceded by induction process using non-black-box classification algorithm. The algorithm were C4.5, CART, and FDT. After induction of the rule then converted to fuzzy rule-based. Third, perform the testing process with the model, as shown in Figure 1. At each level in making a decision using fuzzy inference system with Mamdani method.

The performance of the proposed diagnostic system uses two approaches, namely population-based and individual-based. The population-based performance, using a number of parameters, is referring to the matrix confusion table, as shown in Table 2. The parameters are sensitivity, specificity, accuracy, and area under the curve (AUC) [11].

Table 2. Confusion Matric

Actual Class	Prediction Class	
	Positive	Negative
Positive	TP (True Positif)	FN (False Negative)
Negative	FP (False Positif)	TN (True Negative)

The second approach was individual-based, ie performance parameters to interpret patients individually, ie by the likelihood ratio parameter [8]. Referring to the likelihood ratio, it can then be used to calculate the pre-test and post-test probability. The performance parameters can be formulated in the Equation (1-8).

$$\text{sensitivity} = \frac{TP}{TP+FN} \tag{1}$$

$$\text{specificity} = \frac{TN}{TN+FP} \tag{2}$$

$$LR+= \text{sensitivity}: (1 - \text{specificity}) \tag{3}$$

$$LR-= (1 - \text{sensitivity}): \text{specificity} \tag{4}$$

$$\text{pre - test probabily} = \frac{TP+FN}{TP+FN+FP+TN} \tag{5}$$

$$\text{pre - test odds} = \text{pre - test probability}: (1 - \text{pre - test probability}) \tag{6}$$

$$\text{post - test odds} = \text{pre - test odds} * LR + \tag{7}$$

$$\text{post - test probability} = \text{post - test odds} : (1 + \text{post - test odds}) \tag{8}$$

2.3 Induction Algorithm

Algorithm C4.5 is one of the algorithms in decision tree learning. The C4.5 algorithm is a development of the ID3 algorithm [12]. The ID3 algorithm attempts to build a top-down decision tree, by specifying the root first. The root determination process is done by evaluating all attributes by using a statistical measure (information gain) to measure the effectiveness of attributes in classifying the sample data. The difference in a C4.5 algorithm with ID3 was in the defining attributes as root. The calculation of the gain in the C4.5 algorithm as shown in the equations (9) and (10) [12] [13].

$$\text{GainRatio}(S, A) = \frac{\text{Gain}(S,A)}{\text{SplitInfo}(S,A)} \tag{9}$$

$$\text{SplitInfo}(S, A) = \sum_{i=1}^c \frac{S_v}{S} \log_2 \left(\frac{S_v}{S} \right) \tag{10}$$

The FDT algorithm is an improvement of the C4.5 algorithm developed by Jiang Su and Harry Zhang [14]. FDT has a process of constructing a decision tree that is better and faster, which indicated by its complexity. The decision tree in FDT is built using independent information gain (IIG), as a criterion for separation [14]. IIG was calculated for all candidate attributes. The attribute with the highest IIG value will be selected as the root node.

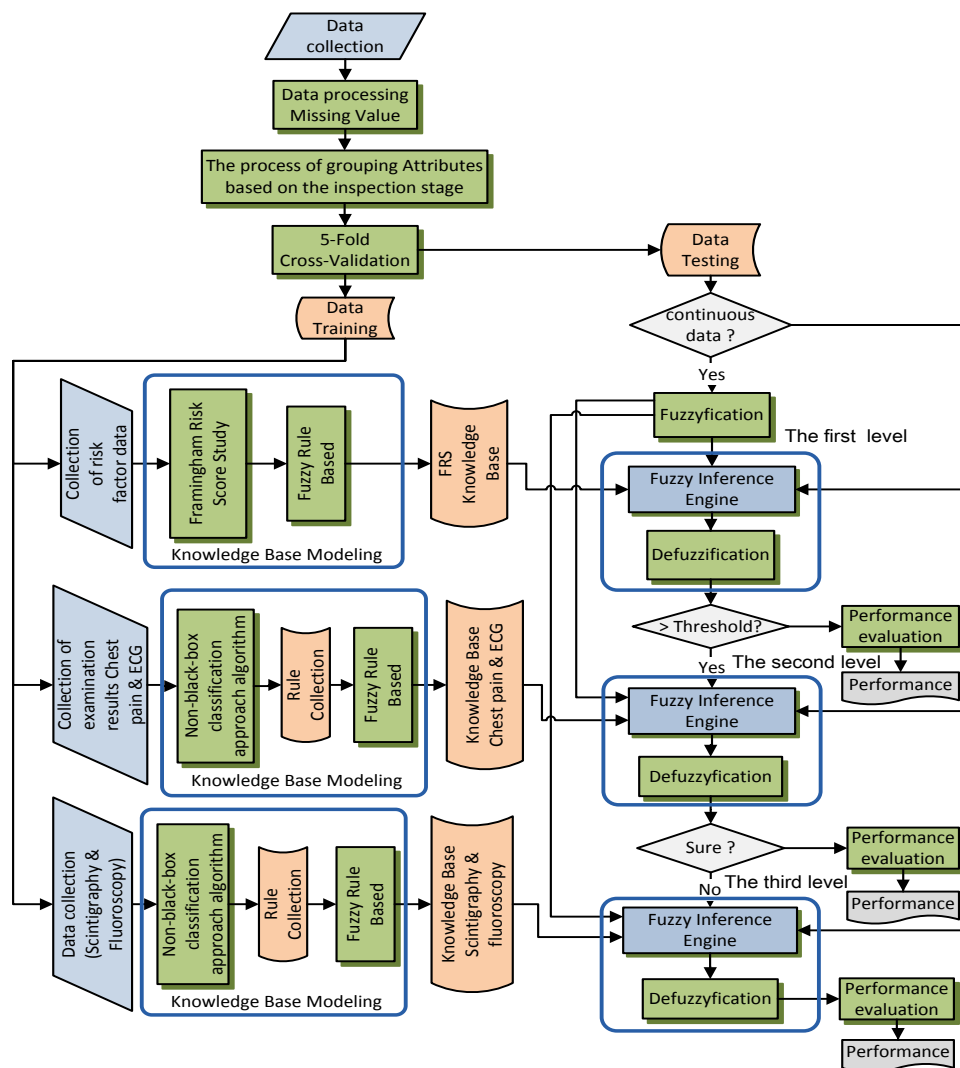


Figure 1. The proposed of method

The CART algorithm was a classification algorithm based on the decision tree, introduced by Leo Breiman et al. in 1984 [15]. CART produces a classification tree if the response variable was categorical, and produces a regression tree if the response variable was continuous. The CART algorithm has the following steps:

- a. Prepare the branch candidates for all predictor variables, each divided into 2, the left and the right branch candidates.
- b. Assess the overall prospective branch that is on the list of candidates for the latest branches. The performance of each branch candidate will be measured by a measure called conformity.
- c. Determine which branch candidate will actually be a branch.
- d. If no more decision nodes, the implementation of the CART algorithm will be terminated. However, if there is still a decision node, go back to the second step, by first throwing the branch candidate who has successfully become a branch.

The main concept of the fuzzy logic theory was to place an input space into the output space using IF-THEN rules, and the concept can be used to solve the uncertainty problem [16]. Mapping was done by using Fuzzy Inference System (FIS). FIS evaluates all rules simultaneously to generate conclusions. FIS has a component consisting of fuzzification, membership function, inference process and defuzzification [17].

3. RESULTS AND DISCUSSION

3.1 Results

The dimensional reduction model with a tiered approach in the case of prediction of coronary heart disease was divided into three levels, as shown in Figure 1. The process of the diagnosis was done by tiered model, if at the level of 1 declared prediction results have a high risk, then go to the level of 2, but if not then the diagnosis was complete. If the diagnosis results at the second level, it is possible to make a diagnostic decision for the clinician, there is no need for a diagnosis at the third level, but if not, then the third level diagnosis is done. At level 1, gives the output of the percentage of coronary heart disease events in the next 10 years. A follow-up diagnosis to the second level will be performed if the generated procession passes a certain threshold.

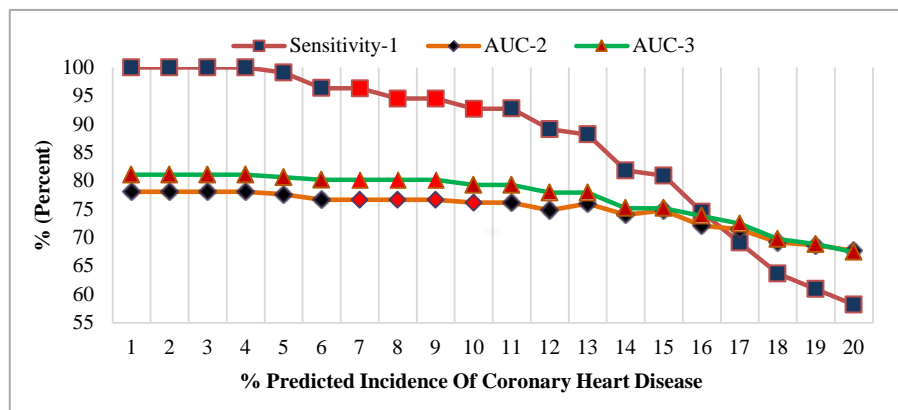


Figure 2. The performance of the diagnostic system at the levels 1, 2 and 3

Determination of threshold, which was the percentage of the prediction of coronary heart disease refers to several considerations that were:

- a. the influence of a threshold value on the performance at level 1, level 2 and level 3,
- b. At the first stage is the screening stage, thus emphasizing the sensitivity performance parameter [8],
- c. refers to the percentage predicted incidence of coronary heart disease in the next 10 years, if values are <10% said to have low risk [18].

A number of these considerations used to analyze the results of testing at level 1. Testing to determine threshold is done by using training data. The resulting performance as shown in Figure 2, referring to a number of these considerations and the performance of the test results, the percentage value of coronary heart disease events, which was used as a threshold was > 7%. This value indicates that if the predicted output at level 1 was > 7%, then the diagnosis was required at the second level.

Tiered system model test was done by using 5-fold cross-validation method. Performance parameters used were population-based and individual based. The results of the tests performed when using the threshold at level 1 are > 7%, the resulting population-based performance as shown in Table 3. The next performance parameter was individual-based parameters, namely the likelihood ratio. The parameters can be used to calculate the pre-test probability, post-test probability, and the difference. These values indicate how much influence the use of a prediction model of coronary heart disease. The resulting performance as shown in Table 4.

Table 3. System performance for population-based for each level

Algorithm	Level	Sensitivity (%)	Specificity (%)	Accuracy (%)	AUC (%)
FRS	1	98,58	7,64	50,67	53,11
C4.5	2	75,89	82,17	79,19	79,03
	3	65,96	94,90	81,21	80,43
FRS	1	98,58	7,64	50,67	53,11
CART	2	71,63	78,98	75,50	75,31
	3	60,99	96,18	79,53	78,59
FRS	1	98,58	7,64	50,67	53,11
FDT	2	81,56	73,89	77,52	77,72
	3	63,83	94,90	80,20	79,37

The system performance shown in Table 3 and Table 4 were the system performance derived from the use of a number of dimension reduction attributes. Dimension reduction using feature selection method type embedded. The method was also used for dimension reduction and induction rule. The algorithm for the induction rule used, including within the decision tree family. The algorithms are C4.5, FDT, and CART. The result of the dimensional reduction process as shown in Table 5.

Table 4. System performance for individual-based for each level

Algorithm	Level	LR+	Probability		differences (%)
			Pre-test (%)	Post-test (%)	
FRS	1	1,07	47,32	48,94	1,63
C4.5	2	4,26	47,32	79,26	31,94
	3	12,94	47,32	92,08	44,76
FRS	1	1,07	47,32	48,94	1,63
CART	2	3,41	47,32	75,37	28,06
	3	15,96	47,32	93,48	46,16
FRS	1	1,07	47,32	48,94	1,63
FDT	2	3,12	47,32	73,72	26,40
	3	12,53	47,32	91,84	44,52

Table 5. The ability to reduce attributes at each level

Algorithm	level	the amount is reduced	number of attributes	Attributes
FRS	1	1	4	age, sex, restbps, chol
C4.5	2	1	5	cp, thalac, exang, oldpeak, slope
FDT	2	3	3	cp, oldpeak, thalac
CART	2	5	1	cp
C4.5	3	0	2	ca, thal
FDT	3	0	2	ca, tha
CART	3	0	2	ca, thal

3.2 Discussions

The dimensional reduction model with a tiered approach in the prediction system of coronary heart disease, using two approaches. First, the dimensional reduction approach to the system as a whole and dynamic. This approach occurred when the system was used for diagnosis. The second, dimension reduction approach in each level, especially the second level and the third level. The approach occurs when the induction process of the rule at each level, using the method of feature selection type embedded [19].

The overall dimensional reduction approach can be explained by referring to Figure 1. The process of diagnosis in the system begins with a prediction at level 1. If the percentage of the output risk of coronary heart disease is low at the level 1, then further diagnosis is not necessary. In these conditions, it was not

necessary to attribute examination at level 2 and level 3. It shows that there has been a dimensional reduction of all attributes at level 2 and level 3. If the percentage output in the level 1 was high so the risk of coronary heart disease was high, then further diagnosis is required at the second level. If the diagnosis at the second level was sufficient to make a decision for the clinician, further diagnosis to the level 3 was not necessary. The cessation of diagnosis at level 2, indirect dimension reduction, ie all attributes at level 3 do not need examination.

The second approach, namely dimensional reduction at each level, especially level 2 and level 3. Dimensional reduction process at both levels occurs simultaneously with the induction process rule. Referring to the results of the test shown in Table 5, at the first level there is a dimensional reduction, ie one attribute, due to unaccommodated attribute fbs in the score table in the FRS study [20]. At the second level, all algorithms for induction rule can reduce some attributes. The CART algorithm can provide the highest dimensional reduction, compared to other algorithms. The CART algorithm is also capable of delivering performance for the highest post-test probability parameters compared to C4.5 and FDT, ie by using only one cp attribute at level 2. In population-based performance parameters, the performance of C4.5 and FDT is relatively better than the CART.

At the level 3, the use of C4.5, CART, and FDT algorithms is not capable of reducing attributes. The absence of reduced attributes indicates that both attributes at the third level have a high contribution to the result of the diagnosis, so that it can not be reduced. The resulting performance at the 3rd level, for population-based performance parameters is the C4.5 induction algorithm providing the highest performance, while for individual-based is a CART. Dimensional reduction results of the three algorithms, CART algorithm able to reduce the attribute in the number of the most. Referring to the number of attributes that can be reduced, post-tests and pre-test probability, and AUC, the FRS + CART-based fuzzy rule based combinations can provide better performance than FRS + C4.5 and FRS + FDT.

The proposed dimension reduction model is also capable of reducing checking costs. The ability is divided into two approaches. First, the overall approach to the system is dynamic. Second, cost reduction at every level, especially the second and third level. Feshki and Shijani [2] classify the required inspection fees for each attribute into 4, in order from the low-cost to the expensive. Grouping these attributes can be shown as follows:

1. sex, age
2. chol, restbps, fbs, restecg
3. cp, thalac, exang, oldpeak, slope.
4. thal, ca

Referring to the classification of inspection fees, the cost of the examination level 1 < level 2 < level 3. In the proposed system model, if the examination of a patient at level 1 was stated that the risk of coronary heart disease is low, there is no need for examination at level 2 and level 3. Do not do the examination at the level 2 and level 3, making the cost to be reduced. Whereas, if at the first stage it is stated that it needs a further diagnosis, it can be examined at level 2, and if at that level is enough to make a conclusion, then the level 3 examination was not needed anymore.

The second cost reduction concept, was the cost reduction that occurs at each level, especially at level 2 and level 3. This is due to the second and third level there was a reduction in the number of attributes, resulting in reduced costs. For example, when using the CART rule induction algorithm, at the second level it was able to reduce the attributes of ECG examination at rest and exercise, ie restecg, exang, thalac, oldpeak, and slope. Referring to the groupings by Feshki and Shijani [2], these attributes belong to the second and the third groups, or cost in relatively expensive categories. The large number of attributes that can be reduced by the CART induction algorithm rule, making the total cost of examination at level 2 is relatively cheap, compared to using C4.5 and FDT algorithms.

The dimensional reduction with cost consideration has also been done in some previous studies, only in previous studies did not use a tiered approach. Feshki and Shijani [2] research, using particle swarm optimization with fitness functions, consider cost, so as to reduce the two costly attributes, ca and thal. The same is done by Arjenaki et.al [4], only that the study uses a genetic algorithm. Both studies have some disadvantages, namely the use of classification algorithms with the black-box approach, making it difficult to understand the process of diagnosis by clinicians. Second, do not use a tiered approach, so in the process of diagnosis must check all attributes of the reduction results, before being used for input diagnosis system.

The result of comparison of system model performance diagnosis of coronary heart disease with the tiered and non-tiered approach in dimension reduction process can be shown in Table 6. In Table 6, it shows that the performance resulted from the system model with the tiered approach by using the black-box classification algorithm better than the tiered model. This is indicated by the parameters of population-based performance, ie, accuracy and AUC > 80%, on the NB and SVM algorithms. Model of the diagnostic system with the tiered approach is better than tiered by using black-box classification algorithm, that is for

individual-based performance parameters, that is, the mean value of post-test probability and pre-test probability >43,80%. In the non-tiered model, the parameter value is 35.02%, so there is a difference of 8.78%. It shows a better tiered model when used for the diagnosis of each individual. The same is true for non-black-box classification algorithms, more tiered approaches, both for population-based and individual-based performance parameters.

Table 6. Comparison of tiered system performance with no tiered

Algorithm	The algorithm approach	Tiered	Accuracy (%)	AUC (%)	Differences of post & pre-test Probability (%)
NB	Black-box	-	83,89	83,65	37,78
SVM	Black-box	-	82,55	82,21	37,29
MLP	Black-box	-	77,52	77,32	29,99
C4.5	Non-Black-box	-	78,19	77,79	32,43
CART	Non-Black-box	-	78,52	78,06	33,53
FDT	Non-Black-box	-	75,84	75,41	29,88
C4.5+FIS	Non-Black-box	√	81,21	80,43	44,76
CART+FIS	Non Black-box	√	79,53	78,59	46,16
FDT+FIS	Non-Black-box	√	80,20	79,37	44,52

4. CONCLUSION

The tiered dimension reduction model in the predicted system of coronary heart disease was capable of providing two dimensional reduction processes at once. The dimensional reduction that occurs at each level and dimension reduction is dynamically when used for the diagnosis of coronary heart disease. Dimensional reduction at each level, especially the second and the third level, FRS+CART algorithm can provide the highest dimension reduction and the best individual-based performance. The best population-based performance was the FRS+C4.5 algorithm, but the resulting dimensional reduction was lowest.

REFERENCES

- [1] Liu L *et al.* A Hybrid Classification System for Heart Disease Diagnosis Based on the RFRS Method. *Comput. Math. Methods Med.* 2017: 1–11.
- [2] Feshki MG and Shijani OS. *Improving the Heart Disease Diagnosis by Evolutionary Algorithm of PSO and Feed Forward Neural Network.* Artificial Intelligence and Robotics (IRANOPEN), Qazvin, Iran, 2016: 48–53.
- [3] Ahmadi E, Weckman GD, and Masel DT. Decision making model to predict presence of coronary artery disease using neural network and C5.0 decision tree. *J. Ambient Intell. Humaniz. Comput.* 2017: 1-13.
- [4] Arjenaki HG, Shahraki MHN, and Nourafza N. A low cost model for diagnosing coronary artery disease based on effective features. *Int. J. Electron. Commun. Comput. Eng.* 2015; 6(1): 93–97.
- [5] Wiharto W, Kusnanto H, and Herianto H. Hybrid System of Tiered Multivariate Analysis and Artificial Neural Network for Coronary Heart Disease Diagnosis. *Int. J. Electr. Comput. Eng. IJECE.* 2017; 7(2): 1023–1031.
- [6] Wiharto W, Herianto H, and Kusnanto H. The Analysis of Performace Model Tiered Artificial Neural Network for Assessment of Coronary Heart Disease. *Int. J. Electr. Comput. Eng. IJECE.* 2017; 7(4): 2183–2191.
- [7] Wiharto W, Kusnanto H, and Herianto H. Tiered Model Based On Fuzzy Inference System For The Diagnosis of Coronary Heart Disease. *Far East J. Electron. Commun.* 2016; 16(4): 985–1000.
- [8] Hidayat R and Primasari I. Metodologi Penelitian Psikodiagnostika. *Bul. Psikol.* 2011; 19(2): 81-92.
- [9] Detrano R, Jonasi A, Steinbrunn W, and Pfisterer M. Heart Disease Dataset. Unversity California Irvine, California, 1988.
- [10] Wilson PWF, D'Agostino RB, Levy D, Belanger AM, Silbershatz H, and Kannel WB. Prediction of Coronary Heart Disease Using Risk Factor Categories. *Circulation.* 1998; 97(18): 1837–1847.
- [11] Ramentol E, Caballero Y, Bello R, and Herrera F. SMOTE-RSB *: a hybrid preprocessing approach based on oversampling and undersampling for high imbalanced data-sets using SMOTE and rough sets theory. *Knowl. Inf. Syst.* 2012; 33(2): 245–265.
- [12] Hssina B, Merbouha AB, Ezzikouri H, and Erritali M. A comparative study of decision tree ID3 and C4.5. *Int. J. Adv. Comput. Sci. Appl. IJACSA.* 2014; 4(2): 13–19.
- [13] Baihaqi WM, Setiawan NA, and Ardiyanto I. *Rule extraction for fuzzy expert system to diagnose Coronary artery disease.* International Conference on Information Technology, Information Systems and Electrical Engineering (ICITISEE). Yogyakarta, Indonesia, 2016: 136–141.
- [14] Su J and Zhang H. A fast decision tree learning algorithm. *The 21st national conference on Artificial intelligence.* Boston, Massachusetts, 2006, 1: 500–505.
- [15] Breiman L, Friedman J, Stone CJ, and Olshen RA. *Classification and Regression Tree.* New York: Chapman And Hall, 1984.

- [16] Kim J, Lee J, and Lee Y. Data-Mining-Based Coronary Heart Disease Risk Prediction Model Using Fuzzy Logic and Decision Tree. *Healthc. Inform. Res.* 2015; 21(3): 167-174.
- [17] Ezhilmaran D and Joseph PRB. *Fuzzy inference system for finger vein biometric images*. International Conference on Inventive Systems and Control (ICISC). India, 2017: 1-4.
- [18] Pokharel D *et al.* Estimation of 10-year risk of coronary heart disease in nepalese patients with type 2 diabetes: Framingham versus United Kingdom prospective diabetes study. *North Am. J. Med. Sci.* 2015; 7(8): 347-355.
- [19] Jensen R and Shen Q. *Computational Intelligence and Feature Selection : Rough and Fuzzy Approaches*. USA: IEEE Press, 2008.
- [20] Schisterman EF and Whitcomb BW. Coronary age as a risk factor in the modified Framingham risk score. *BMC Med. Imaging.* 2004; 4(1): 1-9.