

# System Diagnosis of Coronary Heart Disease Using a Combination of Dimensional Reduction and Data Mining Techniques: A Review

Wiharto Wiharto<sup>\*1</sup>, Hari Kusnanto<sup>2</sup>, Herianto Herianto<sup>3</sup>

<sup>1</sup>Department of Informatics, Universitas Sebelas Maret, Indonesia

<sup>1,2</sup>Department of Biomedical Engineering, Gadjah Mada University, Indonesia

<sup>2</sup>Department of Medicine, Gadjah Mada University, Indonesia

<sup>3</sup>Department of Mechanical & Industrial Engineering, Gadjah Mada University, Indonesia

\*Corresponding author, e-mail: wiharto@staff.uns.ac.id

## Abstract

Coronary heart disease is a disease with the highest mortality rates in the world. This makes the development of the diagnostic system as a very interesting topic in the field of biomedical informatics, aiming to detect whether a heart is normal or not. In the literature there are diagnostic system models by combining dimension reduction and data mining techniques. Unfortunately, there are no review papers that discuss and analyze the themes to date. This study reviews articles within the period 2009-2016, with a focus on dimension reduction methods and data mining techniques, validated using a dataset of UCI repository. Methods of dimension reduction use feature selection and feature extraction techniques, while data mining techniques include classification, prediction, clustering, and association rules.

**Keywords:** coronary heart disease, Dimension reduction, data mining, feature selection, feature extraction

Copyright © 2017 Institute of Advanced Engineering and Science. All rights reserved.

## 1. Introduction

Coronary heart disease is very dangerous in the world. In 2014 in Southeast Asia, especially Indonesia, the recorded death rate reached 35%, about 1.8 million, and most deaths were caused by coronary heart disease. Such disease is closely related to a person's lifestyle [1]. A good lifestyle as well as an early detection are one of the alternatives for the prevention of heart disease. One alternative to early detection can be done with a computer-based model of diagnosis. It has been proven in studies by [2], that the use of the computer is able to provide health service improvements. In addition, the use of clinical decision support system, capable of providing convenience to clinicians in decision making [3].

Diagnosis of coronary heart disease is done by utilizing clinical data analysis results. The clinical data has relatively many attributes. The number of attributes can cause a lot of work for any sub-optimal classification algorithm [4]. It is possible that many attributes could be not relevant or even in contradiction, thus becoming noise. Clinical data may have attributes sometimes difficult to understand and/or to visualize. Besides this, large amounts of attributes can lead to long computation time and large memory. The use of dimensional reduction method is very important in the development of diagnosis system. Selection of the appropriate dimension reduction methods would improve the system accuracy [5]. In addition to the selection method in terms of dimension reduction, the use of data mining techniques was also influential in the diagnosis of system performance, by implementing various combinations of dimension reduction methods with data mining techniques.

The purpose of this paper is to review a number of articles about the combined use of dimension reduction and data mining techniques for the diagnosis of coronary heart disease. Articles in the review range within the period 2009-2016, and focus on the articles using the dataset from the UCI repository heart disease [6]. The dataset of heart disease has 13 attributes related with the incidence of coronary heart disease, and one output (normal or abnormal). The results of the review can be used to trace the development of research on diagnosis system, so it can be used as a reference for future developments.

This paper is composed of several sections. Besides the introduction, the second section is a brief overview of data mining and dimension reduction methods, describing also the comparison of performance of the system diagnosis among articles, with reference to the dimension reduction algorithms and data mining techniques used. The third section discuss the development of research, while the fourth section draws the conclusions.

## 2. Research Method

Most of the coronary heart disease diagnosis systems are based on the combination of data mining techniques and dimension reduction. Data mining is often also called knowledge discovery in databases (KDD). KDD is a computer-assisted process to explore and analyze large amounts of datasets and may extract useful information and new knowledge [7]. KDD can be decomposed into a step-by-step process, i.e. data selection, data cleaning, data reduction, data mining and interpretation evaluation [8]. Data mining techniques used in this study are focused on classification, clustering [9], prediction [10], and association rules [9]. Algorithms in classification techniques can be grouped into two approaches, namely the algorithm with black-box and non black-box approach [11].

Dimension reduction is the process of reducing high-dimensional data into a low dimension of data, with minimal lack of information [8]. The dimension of the dataset is the number of attributes which represent objects in the dataset. The data dimension tends to differ qualitatively lower than data in high dimension. The reduction process can be grouped into two dimensions, namely, feature selection and feature extraction. Feature selection is a process that aims to select a set of data features more suitable to infer an almost complete knowledge about the problem represented by those data. Feature selection usually addresses three types: embedded, filtering and wrapper. Furthermore, feature extraction is a process to make high-dimensional data transformation to a low-dimensional space through the adoption of several mapping [12]. Feature extraction can also create a new feature based on the transformation and combination of original features [13]. Feature extraction can be grouped into two categories, namely linear and non-linear.

### 2.1. Feature Selection Embedded Type

The embedded type of feature selection utilizes machine learning in the process. In this selection system a feature is removed if the machine learning assumes this feature as not so influential. The selection process that utilizes machine learning is similar to the system diagnosis using classification algorithms with the non black-box approach. The results of a review of the diagnosis system using embedded feature selection type, is shown in Table 1.

Table 1. Review of Articles For Embedded Feature Selection

Authors	Feature Selection	Description system	Results
[14]	C4.5 and Ensemble Bagging with C4.5	Comparing the performance of the system with C4.5 and bagging ensemble. Both systems were validated using 10-fold cross validation.	Bagging ensemble system is able to provide improved performance compared to using only C4.5.
[15]	C4.5 and Fast Decision Tree (FDT)	FDT, an improvement of algorithm C4.5, FDT has a lower complexity compared to C4.5, so the process requires faster computation.	The test results showed that the average performance of C4.5 is better than FDT. FDT is superior in terms of a reduced number of attributes. At least attribute makes the size of the resulting tree FDT smaller, so the decision process results faster.
[16]	Decision tree (DT)	In a study comparing a number of classification algorithms, one of which is a decision tree	The results show that the decision tree is better than logistic regression algorithm, but it loses well with the neural network.
[17]	C4.5, CART, RIPPER and Fuzzy inference system (FIS)	Fuzzy-based diagnosis system, but the rule extraction is done using the algorithm C4.5, CART, and RIPPER then transformed into a fuzzy rule. Membership functions used in fuzzy rules optimization with algorithm Imperialist competitive algorithm (ICA).	The resulting performance indicates that the use of extraction rule with C4.5 is better. Unfortunately, in these studies is not explained anywhere the attributes reduced.

Authors	Feature Selection	Description system	Results
[18]	Fuzzy Inference System (FIS) with weighted fuzzy rule (WFR)	The system is divided into two phases, namely making a weighted fuzzy rule and build a fuzzy rule-based decision support system. The WFR-making process is done by mining techniques and selected attributes. As for the decision-making process, it is done by using FIS.	The test results provide accuracy performance for the dataset Cleveland 62.3%, Hungarian 46.93%, and Switzerland 51.22%. Unfortunately, it is not shown the reduced number of attributes.
[19]	Decision tree with fuzzy inference system (FIS)	The method used is divided into several processes, namely, conducting mining techniques, selection attributes and creates a tree diagram with reference to the decision tree algorithm. The process of tree-rule result then converted into a fuzzy system. The result of such conversion will be obtained by fuzzy rules. The decision-making process is done by FIS models.	The test results showed the performance parameters of accuracy by using a dataset Cleveland of 68.35%, while the dataset Hungarian of 49.9%. The results also showed improvement from past studies. Unfortunately, also not shown the reduced number of attributes.
[20]	The combination of k-means, Weighted Associative Classifier (WAC) and C5.0.	The system leverages the k-mean algorithm, for clustering of data. Furthermore, the WAC to produce a number of rules for classification. Furthermore, the classification process by using C5.0.	The use of K-mean and WAC are able to provide improved performance C5.0. Unfortunately, in these studies did not indicate the number of attributes generated from the dimension reduction process.
[21]	Fuzzy Inference System (FIS) and C4.5 with tiered approach	The model system uses a tiered approach, based on three levels. They are prediction using risk factors, diagnosis with reference of symptoms and Electrocardiography (ECG), and the specialist examination. At the level of the 2nd and 3rd, the extraction rule is done with C4.5 then transformed into a fuzzy rule. While level-1 refers to the Framingham risk score.	The reduction process occurs also at each level. The advantage of tiered models the possibility to perform dynamic dimension reduction. The reduction showed that, when a diagnosis is measuring a certain negative level, then no further examination is required. It is also able to reduce the cost of inspection.

## 2.2. Feature Selection through Filtering

In such cases the process is filtered by providing scores for each attribute, then selecting the attributes based on the assigned score. Feature selection filtering type has several advantages: the process is fast, scalable, independent from the classification algorithms and reduces the computational complexity [22]. The weakness of this type is the lack of interaction with the classification algorithm. Examples of types of filtering algorithms are: information gain [23], genetic search [12], correlation-based feature selection [12] and rough sets [8]. Review Articles on coronary heart disease diagnosis systems that uses this type are divided into two groups, namely diagnosis systems with binary outputs (normal and abnormal) and multiclass (abnormal output is divided into 4 types/levels). The binary approach in the learning process is preceded by the conversion from multiclass to binary class, whereas multiclass approach is performed without any conversion process. Diagnosis systems with binary output are shown in Table 2, while the multiclass systems are reported in Table 3.

Table 2. Review of Articles using Filtering Feature Selection with binary output

Authors	Feature Selection	Algorithm in Data Mining Techniques	Results
[23]	Information Gain (IG)	Artificial Neural Network	The dimensional reduction process produces 8 attributes on 13 available. The resulting performance is better than using all the attributes.
[24]	Rough Sets Theory (RST)	Fuzzy Decision Support System (FDSS), MLP-ANN, k-NN, C4.5 and RIPPER	FDSS performs the learning system of rules generated by RST, as well as for MLP-ANN, C4.5, and RIPPER. FDSS better performance and efficient than algorithm-ANN MLP, k-NN, C4.5, and RIPPER.
[25]	Rough Sets Feature selection (RSFS), Logistic regression (LR), Multivariate adaptive regression splines (MARS)	Neural Network	RSFS reduces attributes to 10, LR to 12 and MARS 6 attributes from original 13 attributes. The resulting performance are improved if compared with no use of FS, and the combination of MARS and NN provide better performance with accuracy 82.14%.

Authors	Feature Selection	Algorithm in Data Mining Techniques	Results
[26]	Multiple Logistic Regression (MLR) and Sequential feature selection (SFS)	Neuro-Fuzzy Classifier	MLR is able to reduce the space to 6 attributes. The best performance was reached by the combination of MLR & Neuro-Fuzzy Classifier, with an accuracy of 84%.
[27]	Logistic Regression	C5.0, Support Vector Machine, k-NN, Neural Network	LR is able to reduce attributes to 6, and the best performance is reached when combined with C5.0, with an accuracy of 93.02%. When compared with the LR itself, C5.0 also performs well, by looking at the value of the area under the curve (AUC).
[28]	Fuzzy entropy with Mean Selection (MS), Half selection (HS) and Neural Network for threshold selection (NNTS)	Radial basis function neural network (RBF-NN)	The use of MS reduces the attributes to 6. HS reduces to 7 attributes and NNTS to 3 attributes. The resulting performance of the combination NNTS and RBF-NN, is with an accuracy of 84.46%.
[29]	Best Search First[30], The algorithm has four stages, namely the subset generation, evaluation subset, Stop criterion and validation subset.	Fuzzy Inference System (Type Mamdani) Decision Tree (J48/C4.5) Neural Network	Feature selection resulted in 6 attributes from 13. Fuzzy inference system is capable of delivering performance with better accuracy than Decision Tree and Neural Network.
[31]	CFS subset, Chi-squared, consistency subset, filtered attributes, filtered subset, gain ratio, information gain, latent semantic, one attribute and relief	Naive Bayesian, Decision Tree, k-NN and Neural Network	The study shows a naive Bayesian algorithm, able to provide the best performance by using a feature selection algorithm of CFS subset.
[32]	Sequential Floating Forward Selection (SFFS)	Fuzzy Rule Based on Multi-Objective Optimization by a Genetic algorithm, and then use the bagging ensemble classification.	Generates 4 attributes influencing coronary heart disease. The bagging system model ensemble provides better performance.
[33]	Statistic Feature Selection (SFS)	Supervised Fuzzy C-Mean Classification (SFCM) and in combination with Generalized Minkowski Metrics (GMM) to overcome the object with features that different measurement scale. Parameters in SFCM and GMM optimization with Differential Search Algorithm (DSA)	SFS is able to reduce the use of attributes, and with a combination of SFS + SFC + GMM and optimized with DSA is able to deliver improved system performance with an accuracy of 87%.
[34]	Correlation Coefficient Method, R-Squared Method, and Weighted Least Squared Method	Fuzzy inference system with optimization of fuzzy rules and membership functions using Dynamic -Particle Multi-Swarm Optimization (DMS-PSO)	The combined use of fuzzy logic with DMS-PSO provides more effective diagnosis system and is able to provide improved accuracy. In addition, the system is better than a number of other research.
[35]	Correlation Coefficient Method, R-Squared Method, and Weighted Least Squared Method.	Fuzzy inference system with fuzzy weighted rule is generated using a genetic algorithm (GA)	The combination is capable of providing more effective diagnosis system and is able to provide improved accuracy. In addition, the system is able to reduce attributes.
[36]	Integrate the Relief approach with the Rough Set method (RFRS)	Ensemble classifier based on C4.5 algorithm	Dimension reduction is done in two stages: first with relief algorithm, while the second stage is a rough sets heuristic algorithm. It obtains 7 attributes, with an accuracy of 92.59%.
[37]	Logistic regression with tiered approach	Multilayer perceptron neural network with gradient descent	The tiered approach can reduce the dimension to two complex attributes, namely scintigraphy, and fluoroscopy. The proposed system performance, by using classification algorithms based on MLP, is better, especially without using these two expensive attributes.

Table 3. Review of Articles using Filtering Feature Selection with multiclass output

Authors	Feature Selection & Classification	Algorithm in Data Mining Techniques	Results
[38]	Correlation-based feature selection (CFS) and motivated feature selection (MFS)	Naive Bayesian (NB), IBK, AdaBoostM1 C4.5 (J48) and PART (Binary approach)	Use of CFS or MFS provides improved accuracy, but low in the True Positive Rate
[39]	Correlation-based feature selection (CFS) and MFS	NB, IBK, AdaBoostM1, C4.5, and PART. The iteration process performed up to 100 times. (Binary approach)	Use of CFS or MFS provides improved accuracy, and also a true positive rate improvement.
[40]	Correlation-based feature selection (CFS) and motivated feature selection (MFS)	Naive Bayesian, IBK, AdaBoostM1 C4.5 (J48) and PART. The addition of randomizing process before processing the feature selection (Binary class Approach)	Use of CFS or MFS provides improved accuracy, but low in the True Positive Rate when not using feature selection. True Positive Rate is better when combined with feature selection, either CFS or MFS.
[41]	CFS, MFS, Rough Sets Feature Selection (RSFS) and Wrapper based feature selection (WBFS)	J48 (C4.5) and Naive Bayesian (Binary class Approach)	Accuracy is best known for WBFS combination with J48 (C4.5), also WBFS with naive Bayesian. Performance accuracy of the resulting average is 86.76% and 85.67% respectively.
[42]	Information Gain (IG) and Genetic Search (GS)	Naive Bayesian, Support Vector Machine, Logistic Regression, Multilayer Perceptron, Decision Tree, Adaboost (C4.5) (Multiclass approach)	The IG method produces 5 attributes, and GS 8 attributes. The resulting performance is relatively low, i.e. the accuracy is still below 70%
[43]	Information Gain	Resample, Remove duplicate, Synthetic Minority Over-Sampling Technique (SMOTE), C4.5. (multiclass approach)	Being able to reduce the 3 attributes from 19 attributes, but it can provide improved accuracy and true positive rate, either using feature selection or not, with an average accuracy of 90%.

### 2.3. Feature Selection Based on Wrappers

The feature selection based on wrapper systems works by selecting concurrently with the modeling. Selections of this type using a criterion that utilizes classification rate of classification methods are used. The process of wrapping has several advantages: it is simple, interacts with the classification algorithm, is dependent with algorithm classification and provides good classification accuracy. This type also has the disadvantage of requiring intensive computational processes [22]. Feature selection based on wrappers makes use of commonly used algorithms, like genetic algorithm (GA) [12,44-46], artificial bee colony [47] and particle swarm optimization (PSO) [48]. The review of a number of articles that combine wrapping feature selection with data mining techniques, for the diagnosis of coronary heart disease, can be grouped into two categories, by considering the cost of inspection for each attribute or not considering the cost of the examination. The results of such review are shown in Table 4 and Table 5, respectively.

Table 4. Review of Articles For Wrapping Feature Selection without cost examination

Authors	Feature Selection	Algorithm in Data Mining Techniques	Results
[44]	Genetic algorithm (GA)	Naive Bayesian, support vector machine, C4.5 and multilayer perceptron (MLP).	The reduction process produces 7 from 13 attributes, and the best accuracy performance parameters are given from the combination of GA and Naive Bayesian
[45]	GA with multi-object function	Neural Network (NN): Back-propagation	The combination of GA with NN produces 8 attributes. The performance accuracy is 89.58%.
[46]	Genetic Algorithm	Fuzzy Inference System	It produces 7 attributes and obtains 86% of accuracy, using a validation method 10-fold cross validation
[47]	Artificial Bee Colony (ABC)	Support vector machine	Feature selection resulted in 7 attributes and the resulting accuracy is 86.76%.

Authors	Feature Selection	Algorithm in Data Mining Techniques	Results
[48]	Particle swarm optimization (PSO) and CFS	MLP, Multinomial Logistic Regression (MLR), fuzzy unordered rule induction algorithm (FURIA) and C4.5	Feature selection resulted in 7 attributes. An improvement of the diagnosis accuracy is done by combining with k-means, whereas the classification uses multiple algorithms such as MLR, MLP, FURIA, and C4.5. The test results using a 10-fold cross validation, a model with a combination of CFS, PSO, k-mean and MLP obtain an improvement of 11.4% compared to using 13 attributes.
[49]	Associative Rule	Multi-Layer Perceptron (MLP)	Combination association rule and the MLP are able to provide an accuracy of 84.97%, compared with no use of association rule.

Table 5. Review of Articles For Wrapping Feature Selection with cost examination

Authors	Feature Selection	Algorithm in Data Mining Techniques	Results
[50]	Genetic algorithm using a fitness function that takes into account the cost of inspection	Naive Bayesian (NB), Support Vector Machine	The reduction process produces 8 attributes from 13. The use of expensive attributes, namely of scintigraphy and fluoroscopy, is reduced. Performance accuracy is produced using the same relatively costly attributes, which is 85.18% with the naive bayesian algorithm.
[51]	PSO using a fitness function that takes into account the cost of inspection	Feed Forward Neural Network (FFNN)	Results of dimension reduction are 8 attributes, with expensive attributes, namely scintigraphy and fluoroscopy. The resulting performance is better than a number of previous studies.

#### 2.4. Feature Extraction

Feature extraction has the sense to make high-dimensional data transformation into a low-dimensional space, through the adoption of several mappings [12]. Feature extraction can be classified into two categories, depending on the type of transformation used, such as linear and non-linear. Linear feature extraction algorithms are Principle component analysis (PCA) [12] and linear discriminant analysis (LDA) [12], while for non-linear category, the locally linear embedding and ISOMAP [8]. ISOMAP is a special type of linear feature extraction algorithm known as multidimensional scaling [8]. The most often used in the case of coronary heart disease is the PCA.

The authors [52] proposed a diagnosis model by combining PCA with SVM. PCA performs a transformation to a particular domain, which in turn produced a principal component of influence. The results of the performed rotation principal component are mapped to the original attributes. The results of the PCA produces 8 attributes. The classification using 8 attributes and SVM produced a performance with an accuracy of 84.1%, using a linear kernel function. In addition, if the kernel function used in SVM is the Radial Basis Function (RBF), the method is able to increase the accuracy to 88.6%.

PCA is also used in research by [53], which combine PCA with adaptive neuro-fuzzy inference system (ANFIS). Results of dimension reduction using PCA obtained 7 attributes from the original 13 attributes. The test demonstrated an accuracy of 93.2%. The authors [54] also use PCA combined with ANN for diagnosis of coronary heart disease. This study generated 10 principal components, after transformation. The test results with ANN classification obtained the highest accuracy performance (86.1%) when using 5 principal components. Unfortunately, the study did not mention specifically the number of attributes that can be reduced from the use of the 5 principle components.

Feature extraction is also widely approached with Fisher Discriminant Ratio, also known as LDA. The authors [55] proposed a diagnosis system using a hybrid system. The proposed system consists of several stages. First, undergo a ranking of the attributes by using LDA. Second, the top ranking attributes are used as the initial population of the genetic algorithm. Genetic algorithms are used to provide an array of optimal attributes. Third, a modified k-NN is applied to improve the performance of the back-propagation neural network. The test results show that the performance of the system is better than a number of classification algorithms, such as ANFIS, ID3, NB, and radial basis function.

### 3. Results and Analysis

Reviewed articles related to the research on coronary heart disease diagnosis system, using a combination of dimension reduction and data mining techniques, taken over a period of years 2009-2016, are 41. The distribution of these articles per year is shown in Table 6. Referring to Table 6, the feature selection is dominated by filtering methods, while the embedded and wrapper type are lower. The advantage of embedded type is the dimensional reduction process that is attached to the classification process. The classification algorithm used is non-black box approach, such as fuzzy system. This approach makes knowledge based modeled in the form of IF-THEN, so it is easily understood by clinicians. Most studies rarely give reasons for selecting an algorithm taking into account the ease of understanding the processes in the system by clinicians. Most research on the diagnosis system, do not consider the convenience of the clinician in understanding the process in diagnosis, to choose the classification algorithm. Such considerations are only used by some studies as performed by [21], [26], and [43].

Table 6. Distribution of Articles Per Year

Year	Feature Selection			Feature Extraction	Total
	Embedded	Filtering	Wrapper		
2009	1	1	0	0	2
2010	1	0	0	0	1
2011	0	1	0	0	1
2012	1	2	0	1	4
2013	1	2	1	2	6
2014	0	5	3	1	9
<b>2015</b>	<b>1</b>	<b>5</b>	<b>2</b>	<b>0</b>	<b>8</b>
<b>2016</b>	<b>3</b>	<b>2</b>	<b>3</b>	<b>0</b>	<b>8</b>
<b>2017</b>	<b>0</b>	<b>2</b>	<b>0</b>	<b>0</b>	<b>2</b>
Total	8	20	9	4	41

Data mining techniques used in the development of a model of coronary heart disease diagnosis system, within the range of 2009-2016 can be shown in Table 7. The most widely used data mining technique is classification, whereas the classification algorithm which mostly used are C4.5, ANN and Fuzzy system. The algorithm C4.5 almost every year as an option in developing a system of diagnosis of coronary heart disease. The use of data mining clustering techniques is the least used, it is caused in the case of coronary heart disease is a classification case, although clustering techniques can also solve classification cases. Classification process solving with clustering technique is not optimal, the technique is mostly used to optimize the performance of classification algorithms such as those done by [48] and [20]. The prediction techniques in data mining are mostly also used for the optimization of classification algorithms, and are even used as dimensional reduction algorithms, such as research conducted by [27] and [26]. The latter technique is association rule in data mining, the algorithm in this technique is also mostly used as a method for dimensional reduction, such as research conducted by [49].

Table 7. Distributed Algorithm Techniques of Data Mining (Dm)

DM Techniques	Algorithm	2009	2010	2011	2012	2013	2014	2015	2016	2017	Total	
Classification	PART					1	2				3	
	FIS, NFC, FURIA, ANFIS	1			1	2	2	3	4		13	
	SVM				1	1	1	3			6	
	C4.5	2	1		1	3	5	2	4	1	19	
	C5.0							1	1		2	
	ID3							1			1	
	FDT							1			1	
	RIPPER	1								1		2
	CART									1		1
	ANN	1		1	1	3	5	2	3	1	17	
	k-NN	1			1	1	2	1				6
	NB				1	2	5	2				10

DM Techniques	Algorithm	2009	2010	2011	2012	2013	2014	2015	2016	2017	Total
	Bagging	1			1						2
	Adaboost					1	2	1			4
Clustering	SFCM							1			1
Prediction	LR							1			1
	MLR								1		1
Association Rule	WAC								1		1

The development of a diagnostic system model with a combination of dimensional reduction and data mining techniques is dominated with the aim of achieving the best accuracy performance with few attributes. It has not used much consideration of the cost of examining each attribute. The research which have used these considerations are conducted by [50] and [51]. Research by [50], using the genetic algorithm for dimensional reduction with its fitness function is a cost function. This algorithm, combined with the Bayesian naïve classification algorithm or the support vector machine which able to reduce the two attributes that require expensive fees in the examination. If viewed from the performance, the system has a performance that is not much different when using these two attributes. The same concept was also carried out by [51]. The research uses particle swarm optimization algorithm with its fitness function is a cost examination function. The research combined with Feed Forward Neural Network classification algorithm. Both studies, research by [50] and [51] use a classification algorithm with the black-box approach. It was making difficult to understand the process of system work by the clinicians. Another disadvantage is not using the approach, which is commonly used in the diagnostic process, namely a tiered approach.

#### 4. Conclusion

Review of a number of articles about the system diagnosis of coronary heart disease by using the combination of dimension reduction and data mining techniques provides several important points. First, the development of a diagnostic system focuses on improving the performance parameters of accuracy and the number of reduced attributes. Secondly, there has not been much research that included consideration of ease of understanding of the diagnosis process for clinicians, in choosing a classification algorithm. Thirdly, the diagnostic system approach used has not been much to refer to the tiered diagnostic approach, as is commonly used by clinicians. Fourth, the development of many diagnostic systems has not used the cost variables in dimensional reduction. Referring to these points, the model of the diagnostic system can be developed in the direction of the diagnostic system model with a tiered approach, using non-black-box classification algorithms, and Dimensional reduction that considers the cost of attribute inspection.

#### References

- [1] Kim J, Lee J & Lee Y. Data-Mining-Based Coronary Heart Disease Risk Prediction Model Using Fuzzy Logic and Decision Tree. *Healthcare Informatics Research*, 2015; 21(3): 167–174.
- [2] Anderson C, Henner T & Burkey J. Tablet computers in support of rural and frontier clinical practice. *International Journal of Medical Informatics*. 2013; 82(11): 1046–1058.
- [3] Garg AX, Adhikari NK, McDonald H, Rosas-Arellano MP, Devereaux PJ, Beyene J, ... Haynes RB. Effects of computerized clinical decision support systems on practitioner performance and patient outcomes: a systematic review. *Jama*. 2005; 293(10): 1223–1238.
- [4] Wang S, Li D, Zhao L & Zhang J. Sample cutting method for imbalanced text sentiment classification based on BRC. *Knowledge-Based Systems*. 2013; 37(-): 451–461.
- [5] Forman G. An extensive empirical study of feature selection metrics for text classification. *Journal of Machine Learning Research*. 2003; 3(2003): 1289–1305.
- [6] Detrano R, Jonasi A, Steinbrunn W & Pfisterer M. Heart Disease Dataset. California: University California Irvine. 1988.
- [7] Zainudin S, Jasim DS & Bakar AA. Comparative Analysis of Data Mining Techniques for Malaysian Rainfall Prediction. *International Journal on Advanced Science, Engineering and Information Technology*. 2016; 6(6): 148–1153.

- [8] Jensen R. *Combining rough and fuzzy sets for feature selection*. PhD Thesis. Scotland: University of Edinburgh; 2005.
- [9] Gorunescu F. *Data Mining Concepts, Models and Techniques*. Berlin, Heidelberg: Springer Berlin Heidelberg. 2011; 12.
- [10] Krishnaiah V, Narsimha G & Chandar NS. Heart Disease Prediction System using Data Mining Techniques and Intelligent Fuzzy Approach: A Review. *International Journal of Computer Applications*. 2016; 136(2): 43–51.
- [11] Duda RO, Hart PE & Stork DG. *Pattern Classification*. New York: John Wiley. 2012.
- [12] Cateni S, Vannucci M, Vannocci M & Coll V. Variable Selection and Feature Extraction Through Artificial Intelligence Techniques. In L. Freitas (Ed.), *Multivariate Analysis in Management, Engineering and the Sciences*. Rijeka, Croatia: InTech; 2013.
- [13] Jain A & Zongker D. Feature selection: Evaluation, application, and small sample performance. *IEEE Transactions on Pattern Analysis and Machine Intelligence*. 1997; 19(2): 153–158.
- [14] Tu MC, Shin D & Shin D. Effective diagnosis of heart disease through bagging approach. The 2nd International Conference on Biomedical Engineering and Informatics. Tianjin, China. 2009.
- [15] El Bialy R, Salamay MA, Karam OH & Khalifa ME. Feature Analysis of Coronary Artery Heart Disease Data Sets. *Procedia Computer Science*. 2015; 65: 459–468.
- [16] Khemphila A & Boonjing V. *Comparing performances of logistic regression, decision trees, and neural networks for classifying heart disease patients*. International Conference on Computer Information Systems and Industrial Management Applications (CISIM). Krakow, Poland, Poland. 2010.
- [17] Baihaqi WM, Setiawan NA & Ardiyanto I. *Rule extraction for fuzzy expert system to diagnose Coronary artery disease*. The 1st International Conference on Information Technology, Information Systems and Electrical Engineering (ICITISEE). Yogyakarta, Indonesia. 2016.
- [18] Anooj PK. Clinical decision support system: Risk level prediction of heart disease using weighted fuzzy rules. *Journal of King Saud University - Computer and Information Sciences*. 2012; 24(1): 27–40.
- [19] Anooj PK. *Implementing Decision Tree Fuzzy Rules in Clinical Decision Support System after Comparing with Fuzzy based and Neural Network based systems*. The International Conference on IT Convergence and Security (ICITCS). Macao, China. 2013.
- [20] Rajalakshmi K & Nirmala K. Heart Disease Prediction with MapReduce by using Weighted Association Classifier and K-Means. *Indian Journal of Science and Technology*. 2016; 9(19): 1-7.
- [21] Wiharto W, Kusnanto H & Herianto H. Tiered Model Based On Fuzzy Inference System For The Diagnosis of Coronary Heart Disease. *Far East Journal of Electronics and Communications*. 2016; 16(4): 985–1000.
- [22] Kumari B & Swarnkar T. Filter versus Wrapper Feature Subset Selection in Large Dimensionality Micro array: A Review. *International Journal of Computer Science and Information Technologies*. 20011; 2(3): 1048–1053.
- [23] Khemphila A & Boonjing V. *Heart Disease Classification Using Neural Network and Feature Selection*. The 21st International Conference on Systems Engineering (ICSEng). Las Vegas, NV, USA. 2011.
- [24] Setiawan NA, Venkatachalam PA & Hani AFM. *Diagnosis of coronary artery disease using artificial intelligence based decision support system*. Presented at the International conference on man-machine systems (ICoMMS). Batu Ferringhi, Penang. 2009.
- [25] Shao YE, Hou CD & Chiu CC. Hybrid intelligent modeling schemes for heart disease classification. *Applied Soft Computing*. 2014; 14: 47–52.
- [26] Marateb HR & Goudarzi S. A noninvasive method for coronary artery diseases diagnosis using a clinically-interpretable fuzzy rule-based system. *Journal of Research in Medical Sciences*. 2015; 20(3): 214–223.
- [27] Abdar M, Kalhori SRN, Sutikno T, Subroto IMI & Arji G. Comparing Performance of Data Mining Algorithms in Prediction Heart Diseases. *International Journal of Electrical and Computer Engineering (IJECE)*. 2015; 5(6): 1569–1576.
- [28] Jaganathan P & Kuppuchamy R. A threshold fuzzy entropy based feature selection for medical database classification. *Computers in Biology and Medicine*. 2013; 43(12): 2222–2229.
- [29] Oad KK & DeZhi X. A Fuzzy Rule based Approach to Predict Risk Level of Heart Disease. *Global Journal of Computer Science and Technology: C Software & Data Engineering*. 2014; 14(3): -.
- [30] Dash M & Liu H. Feature selection for classification. *Intelligent Data Analysis*. 1997; 1(3): 131-156.
- [31] Peter (T Jhon) & Somasundaram K. *An Empirical Study On Prediction of Heart Disease Using Classification Data Mining Techniques*. The International Conference on Advances in Engineering, Science and Management. Nagapattinam, Tamil Nadu, India. 2012.
- [32] Lahsasna A, Ainon RN, Zainuddin R & Bulgiba A. Design of a Fuzzy-based Decision Support System for Coronary Heart Disease Diagnosis. *Journal of Medical Systems*. 2012; 36(5): 3293–3306.

- [33] Negahbani M, Joulazadeh S, Marateb HR & Mansourian M. Coronary Artery Disease Diagnosis Using Supervised Fuzzy C-Means with Differential Search Algorithm-based Generalized Minkowski Metrics. *Peertechz Journal of Biomedical Engineering*. 2015; 1(1): 006-014.
- [34] Paul AK, Shill PC, Rabin MRIR, Kundu An & Akhmad MAH. Fuzzy Membership Function Generation using DMSPSO for the Diagnosis of Heart Disease. The 18th International Conference on Computer and Information Technology (ICCIIT). Dhaka, Bangladesh. 2015.
- [35] Paul AK, Shill PC, Rabin MRIR & Akhmad MAH. *Genetic Algorithm Based Fuzzy Decision Support System for the Diagnosis of Heart Disease*. The 5th International Conference on Informatics, Electronics and Vision (ICIEV). Dhaka, Bangladesh. 2016.
- [36] Liu X, Wang X, Su Q, Zhang M, Zhu Y, Wang Q & Wang Q. A Hybrid Classification System for Heart Disease Diagnosis Based on the RFRS Method. *Computational and Mathematical Methods in Medicine*. 2017; 2017: 1–11.
- [37] Wiharto W, Kusnanto H & Herianto H. Hybrid System of Tiered Multivariate Analysis and Neural Network for Coronary Heart Disease Diagnosis. *International Journal of Electrical and Computer Engineering*. 2017; 7(2): 1023-1031.
- [38] Nahar J, Imam T, Tickle KS & Chen YPP. Computational intelligence for heart disease diagnosis: A medical knowledge driven approach. *Expert Systems with Applications*. 2013; 40(1): 96–104.
- [39] Akrami S, Tabatabaee H & Raz NR. Application of Computational Intelligence In Intensifying The Accuracy of Heart Disease Prognosis Using A Mutual Validation. *Indian Journal of Fundamental and Applied Life Sciences*. 2014; 4(S4): 1042–1053.
- [40] Prabowo DW, Setiawan NA & Nugroho HA. A study of data randomization on a computer based feature selection for diagnosing coronary artery disease. *Adv Intell Syst*. 2014; 53: 237–248.
- [41] Setiawan NA, Prabowo DW & Nugroho HA. *Benchmarking of feature selection techniques for coronary artery disease diagnosis*. The 6th International Conference on Information Technology and Electrical Engineering (ICITEE). Yogyakarta, Indonesia. 2014.
- [42] Dominic V, Gupta D & Khare S. An effective performance analysis of machine learning techniques for cardiovascular disease. *Applied Medical Informatics*. 2015; 36(1): 23-32.
- [43] Wiharto W, Kusnanto H & Herianto H. Interpretation of Clinical Data Based on C4.5 Algorithm for the Diagnosis of Coronary Heart Disease. *Healthcare Informatics Research*. 2016; 22(3): 186-195.
- [44] Mokaddem S, Atmani B & Mokaddem M. *Supervised Feature Selection for Diagnosis of Coronary Artery Disease Based on Genetic Algorithm*. The First International Conference on Computational Science and Engineering (CSE-2013). Dubai. 2013.
- [45] Murthy HN & Meenakshi M. *Dimensionality reduction using neuro-genetic approach for early prediction of coronary heart disease*. The International Conference on Circuits, Communication, Control and Computing (I4C). Bangalore, India. 2014.
- [46] Santhanam T & Ephzibah EP. Heart Disease Prediction Using Hybrid Genetic Fuzzy Model. *Indian Journal of Science and Technology*. 2015; 8(9): 797–803.
- [47] Subanya B & Rajalaxmi RR. *Feature selection using Artificial Bee Colony for cardiovascular disease classification*. The International Conference on Electronics and Communication Systems (ICECS). Coimbatore, India. 2014.
- [48] Verma L, Srivastava S & Negi PC. A Hybrid Data Mining Model to Predict Coronary Artery Disease Cases Using Non-Invasive Clinical Data. *Journal of Medical Systems*. 2016; 40(7): 1–7.
- [49] Srinivas K, Reddy BR, Rani BK & Mogili R. Hybrid Approach for Prediction of Cardiovascular Disease Using Class Association Rules and MLP. *International Journal of Electrical and Computer Engineering (IJECE)*. 2016; 6(4): 1800–1810.
- [50] Arjenaki HG, Shahraki MHN & Nourafza N. A Low Cost Model for Diagnosing Coronary Artery Disease Based On Effective Features. *International Journal of Electronics Communication and Computer Engineering*. 2015; 6(1): 93–97.
- [51] Feshki MG & Shijani O S. Improving the Heart Disease Diagnosis by Evolutionary Algorithm of PSO and Feed Forward Neural Network. The Artificial Intelligence and Robotics (IRANOPEN). Qazvin, Iran. 2016.
- [52] Zhang Y, Liu F, Zhao Z, Li D, Zhou X & Wang J. Studies on application of Support Vector Machine in diagnose of coronary heart disease. The Sixth International Conference on Electromagnetic Field Problems and Applications (ICEF). Dalian, Liaoning, China. 2012.
- [53] NG BA. An intelligent approach based on Principal Component Analysis and Adaptive Neuro Fuzzy Inference System for predicting the risk of cardiovascular diseases. Presented at the Fifth International Conference on Advanced Computing (ICoAC), 2013. IEEE.
- [54] Shilaskar S & Ghatol A. Dimensionality Reduction Techniques for Improved Diagnosis of Heart Disease. *International Journal of Computer Applications* 2013; 61(5): 1–8.
- [55] Salari N, Shohaimi S, Najafi F, Nallappan M & Karishnarajah I. A novel hybrid classification model of genetic algorithms, modified k-Nearest Neighbor and developed backpropagation neural network. *PLoS One*. 2014; 9(11): 1–50.