International Journal of Electrical and Computer Engineering (IJECE)
Vol. 7, No. 2, April 2017, pp. 1023~1031
ISSN: 2088-8708, DOI: 10.11591/ijece.v7i2.pp1023-1031

Hybrid System of Tiered Multivariate Analysis and Artificial Neural Network for Coronary Heart Disease Diagnosis

Wiharto[1], Hari Kusnanto[2], Herianto Herianto[3]
[1]Department of Informatic, Sebelas Maret University, Indonesia
[1,2,3]Department of Biomedical Engineering, Gadjah Mada University, Indonesia
[2]Department of Medicine, Gadjah Mada University, Indonesia
[3]Department of Mechanical & Industrial Engineering, Gadjah Mada University, Indonesia

Article Info

Article history:

Received Aug 28, 2016
Revised Jan 4, 2017
Accepted Jan 19, 2017

Keyword:

Coronary heart disease
Diagnosis
Logistic regression
Multi-layer neural network
Multivariate

ABSTRACT

Improved system performance diagnosis of coronary heart disease becomes an important topic in research for several decades. One improvement would be done by features selection, so only the attributes that influence is used in the diagnosis system using data mining algorithms. Unfortunately, the most feature selection is done with the assumption has provided all the necessary attributes, regardless of the stage of obtaining the attribute, and cost required. This research proposes a hybrid model system for diagnosis of coronary heart disease. System diagnosis preceded the feature selection process, using tiered multivariate analysis. The analytical method used is logistic regression. The next stage, the classification by using multi-layer perceptron neural network. Based on test results, system performance proposed value for accuracy 86.3%, sensitivity 84.80%, specificity 88.20%, positive prediction value (PPV) 90.03%, negative prediction value (NPV) 81.80%, accuracy 86,30% and area under the curve (AUC) of 92.1%. The performance of a diagnosis using a combination attributes of risk factors, symptoms and exercise ECG. The conclusion that can be drawn is that the proposed diagnosis system capable of delivering performance in the very good category, with a number of attributes that are not a lot of checks and a relatively low cost.

Corresponding Author:

Wiharto,
Department of Informatic,
Sebelas Maret University,
Jl. Ir. Sutami No.36A, Kentingan, Surakarta, Central of Java, Indonesia.
Email: wiharto@staff.uns.ac.id

1. INTRODUCTION

Rapid economic growth brought about many changes in the pattern of life, especially in developed and developing countries. Changes in lifestyle can give a bad effect on health. One of the diseases that are influenced by lifestyle is coronary heart disease [1]. To overcome this, prevention and early detection are very important to maintain a good lifestyle. Diagnosis of coronary heart disease requires some kind of inspection. Each inspection need costs. The number of cost is depending on the type of examination performed. The more types of tests will be greater costs that must be issued in the process of diagnosis. In making a diagnosis, a clinician will make the selection of the various types of inspection. Selection is based on whether the investigation has a relative advantage compared to other tests. These advantages can be a value higher diagnosis, examination faster, lower risk, checks that are not expensive and some other clinical considerations.

The development of coronary heart disease diagnosis system has been widely applied, using trending classification algorithms and feature selection. A diagnostic system that has been done can be grouped into two categories, namely using and not using the feature selection process. Feature selection is a preprocessing to choose the features that effect and override feature does not affect in any activity modeling or data analysis [2]. The method of feature selection is divided into two groups, ranking selection and subset selection. Ranking unmatched specifically gave the rank on every existing feature and override feature that does not meet certain standards. Subset selection is the method of selection were looking for a set of features which are considered as the optimal feature. The subset selection is divided into three approaches, namely the wrapper, filter, and embedded approach [2], [3]. Wrapper approaches the selection process along with the implementation of classification. This approach also uses criteria by utilizing the classification rate of classification methods are used. Wrapper and filter is almost similar too, but the filter does not involve the method of classification. While embedded, utilizing a machine learning algorithm, so that feature naturally eliminated, if the machine learning assume these features are not so influential [2], [3].

Model feature selection that is widely used is the subset selection with good filters, wrapper and embedded approach. Subanya and Rajalaxmi [3] proposed a system of diagnosis of coronary heart disease, which preceded the process of feature selection. The process is performed by using an algorithm artificial bee colony (ABC). The feature selection algorithms perform against coronary heart disease 13 attributes of the dataset University of California Irvine (UCI). The results produced seven feature selection attributes from 13 attributes. The resulting attribute further classified algorithms using support vector machine (SVM). Use of stage feature selection with ABC algorithm is able to provide better system performance diagnosis. Similar to the study conducted Murthy and Meenakshi [4], in the study before it is classified by the multilayer perceptron backpropagation, preceded the process of feature selection. The feature selection algorithm used is a combination of Neuro-fuzzy and genetic algorithms (Neuro-genetic). Feature selection attributes generates 8 of 13 attributes of coronary heart disease. Similar with research conducted by Mokeddem et.al [5], the feature selection is done by using a genetic algorithm, but the classification is using naive Bayesian. The results of feature selection which using the wrapper approach, obtained 7 attributes from 13 attributes. Genetic algorithms are also used for feature selection in research Santhanam and Ephzibah [6], but the genetic algorithm combined with a fuzzy inference system for classification. Feature selection algorithm results with those obtained 7 attributes of coronary heart disease.

Further studies are associated with the feature selection was proposed by Muthukaruppan & Er [7] in which combining particle swarm optimization (PSO), and classified by fuzzy inference system (FIS). The results of the feature selection attribute obtained 9 of 13 attributes. Not far to the study Marateb & Goudarzi [8] proposed a system of diagnosis by a combination of multiple logistic regression (MLR) with Neuro-fuzzy classifier. The system is able to reduce to 6 attributes from 20 attributes. In addition to using multiple logistic regression, the study also tested the algorithm sequence feature selection (SFS). SFS is able to reduce to 9. Performance attributes resulting from the combination of SFS and Neuro-fuzzy classifier is not better than the combination of multiple logistic regression with Neuro-fuzzy classifier. Arjenaki et.al [9], proposed a system of diagnosis by considering the cost required for diagnosis. The study used a genetic algorithm to do the feature selection with consideration of costs, and classified by naive Bayesian. The results of feature selection produce 8 attributes from 13 attributes are available. The feature selection process is able to reduce costly attributes, namely scintigraphy and flouroscopy examination.

Reduction of dimensions in addition to using feature selection, can also be done with feature extraction. One feature extraction algorithm used is the principle component analysis (PCA). Previous studies on coronary heart disease diagnosis system has been widely used feature extraction algorithm principle component analysis. Model studies conducted Zhang et.al [10], which combine PCA with SVM for the diagnosis of coronary heart disease. PCA feature extraction by generating 9 attributes from 13 attributes. A similar study conducted Bhuvanesawari Amma NG [11], only the classification of research using adaptive neuro fuzzy inference system (ANFIS). In the study were able to reduce the attributes to 7 attributes of the 13 attributes of coronary heart disease.

Diagnosis system model by combining a feature selection or feature extraction in previous studies, generally do not pay attention to the examination stage as that done by the clinician. Clinicians perform the examination with a tiered manner, for each additional inspection should provide added value to diagnose values. Most of the research that has been done feature selection process carried out with the assumption that all attributes have available, for further selection process. Under these conditions, this study proposes a system of diagnosis of coronary heart disease by using the combination of tiered multivariate analysis and multi-layer perceptron neural network (MLP-NN). Tiered analysis done by following the steps in the examination conducted clinician, namely that the addition of the examination should provide significant additional diagnostic value and cost is relatively cheaper. In this study, using the performance parameters commonly used by clinicians, the sensitivity, specificity, positive prediction value, negative prediction value, accuracy and area under the curve.

2. RESEARCH METHOD

This research is conducted using the dataset Clevelands of the University of California Irvine (UCI) and is available online [12]. The dataset consists of 13 attributes that influence the incidence of coronary heart disease, and one attribute is the output result of the diagnosis. The dataset attributes can be grouped into six, namely the risk factors, symptoms, rest ECG, exercise ECG, scintigraphy and Flouroscophy. The risk factors consist of age, gender, systolic blood pressure (restbps), fasting blood suger (fbs) and cholesterol (chol). Further ECG both rest and exercise consisted of Resting ECG (restecg), maximum heart rate Achieved (thalac), exercise induced angina (exang), ST depression induced by exercise relative to rest (oldpeak), and the slope of the ST segment for peak exercise (slope). While scintigraphy to determine the defect type and flouroscopy to detect whether there is a blockage of blood vessels that are classified into single, double, tripple vessel disease. Diagnosis generated categorized into two, namely the blockage of coronary arteries <50% (healthy) and blockage of blood vessels> 50% (sick). The grouping is reinforced in previous research that suggests that, blockage of coronary arteries <50% is not significant, but if it had> 50% or 70% it had a significant [13]. Dataset totaled 303, with 164 compositions obstruction <50% and 139 blockages of> 50%.

The method used in the proposed system, divided into several phases, as shown in Figure 1. The first step is to perform the categorization of coronary heart disease attributes that have continuous data into discrete. The second stage doing feature selection, is divided into four stages of the process. The first process bivariate analysis, the analysis is done by using a significance value of 95%. Attribute with a p-value <0.25 will do multivariate tiered analysis process. The second process is a tiered multivariate analysis using logistic regression. Attribute the results of the logistic regression that has a p-value <0.05 can be used as a criterion in the diagnosis of coronary heart disease. The third process of calculating the AUC values of the results of logistic regression. The next step, do any recurrence of the first process. All these processes are carried out with a combination of risk factors, symptoms, rest ECG, exercise ECG, scintigraphy and flouroscopy. Having done all the combinations, the fourth process is performed interpretation AUC values. Interpretation is done using a statistical approach and clinical considerations. Interpretation AUC values with statistical approach is to classify the AUC values. The classification is being very weak at 50% - 60%, weaker at 60% - 70%, medium at 70% - 80%, good at 80% - 90% and very good at 90% - 100% [14]. Use clinical judgment is the cost of each inspection attribute group.

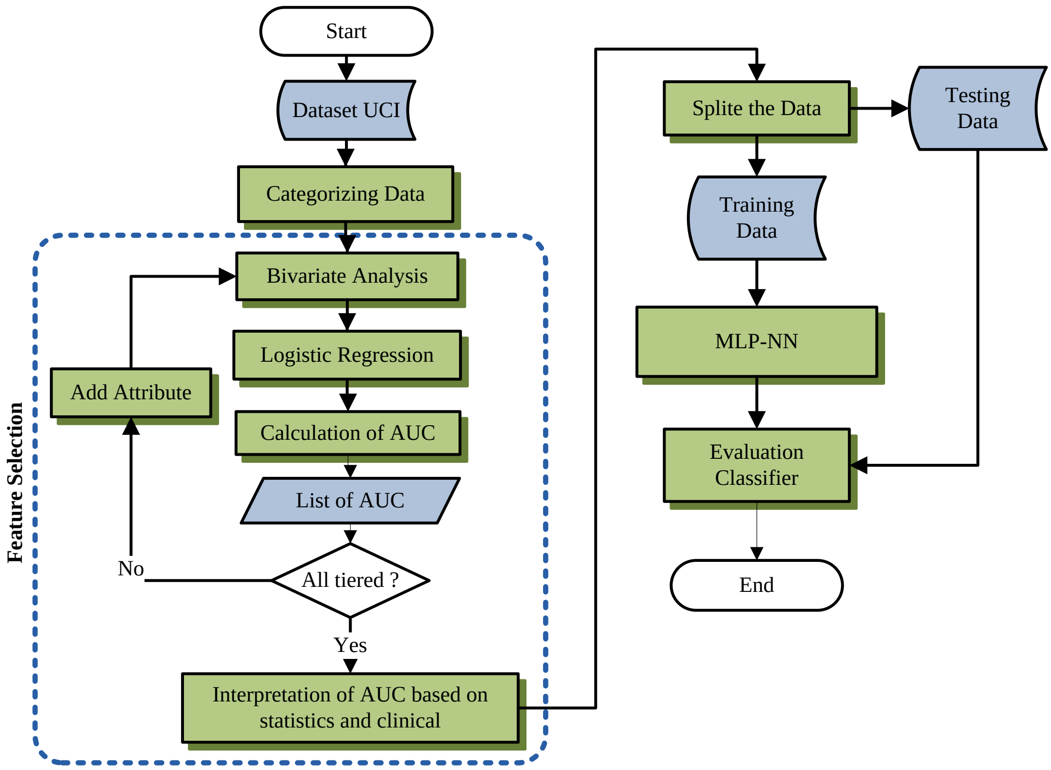

Figure 1. The Proposed system model

The third stage is the process of classification by using multi-layer perceptron neural network (MLP-NN). Attribute that is in the process stages are attributes result previous feature selection stage. Multi-layer perceptron neural netwok trained using backpropagation gradient descent algorithm with momentum. The algorithm is a development of backpropagation gradient descent algorithm, is to perform updates on changes in weight. The addition of variable weights momentum of change can accelerate the convergence in training, compared gradient descent [15],[16]. The following equation (1) and (2), are weight changes during training:

$$w_{jk}(t+1) = w_{jk}(t) + \alpha\delta_k z_j + \mu\left[w_{jk}(t) - w_{jk}(t-1)\right] \tag{1}$$

and

$$v_{jk}(t+1) = v_{jk}(t) + \alpha\delta_k z_j + \mu\left[v_{jk}(t) - v_{jk}(t-1)\right] \tag{2}$$

where μ is the momentum parameter, which has a value between 0-1, w is the weight of the input layer and v is the weight of the hidden layer.

The fourth stage, to evaluate the performance of the system. System performance measured by reference diagnosis matric confusion, as shown in Table 1, the performance parameters used are

1. When a clinician will do an examination, raised two questions, which if positive patients suffering from a disease, how the ability diagnosis system produces positive results?. The question relates to the value of sensitivity. Secondly, if the patient is not suffering from a disease, what is the diagnosis system's ability to generate negative results? The question relates to the value of specificity. Both, the equation (3) and (4), are the value of sensitivity and specificity can be formulated as follows:

$$\text{Sensitivity} = \frac{\text{TP}}{\text{TP+FN}} \tag{3}$$

$$\text{Specificity} = \frac{\text{TN}}{\text{TN+FP}} \tag{4}$$

2. When a clinicians have obtained the diagnosis is positive, then the question arises, how much positive results really positive. This is related to Positive Prediction Value (PPV). Whereas if you get a diagnosis result is negative, then how much negative result really negative. It is associated with a Negative Prediction Value (NPV). Both of these parameters can be formulated at the equation (5) and (6) which shown as follows:

$$\text{Positive Prediction Value (PPV)} = \frac{\text{TP}}{\text{TP+FP}} \tag{5}$$

$$\text{Negative Prediction Value (NPV)} = \frac{\text{TN}}{\text{TN+FN}} \tag{6}$$

3. The performance parameters of area under the curve (AUC), the value of this parameter indicates if there are a number of patients who carried the diagnosis using the system, then how many patients can be diagnosed correctly by the system, so the AUC is the percentage of patients who are diagnosed correctly.
4. The performance parameters of accuracy which can be formulated at the equation (7) as follows:

$$\text{Accuracy} = \frac{\text{TP+TN}}{\text{TP+FP+TN+FN}} \tag{7}$$

Table 1. Confusion Matrics

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

3. RESULTS AND ANALYSIS

In this study proposes a model of hybrid systems for the diagnosis of coronary heart disease. The system combines the two processes, namely, tiered multivariate analysis and multi-layer perceptron neural network (MLP-NN). Tiered multivariate analysis process using logistic regression algorithm, a process aimed at obtaining stretcher attribute that significantly affect coronary heart disease. The process is called the process of feature selection. Tiered method is a reflection of the stages in acquiring attribute, as is done by the clinician. The AUC value calculation is done for each level in order to determine the effect of each level of examination. The calculation of AUC values for each level by using multivariate logistic regression analysis. Analysis done using a 95% confidence level.

The calculation result tiered analysis with logistic regression mutltivariate as shown in Table 2. The table shows AUC values based on a combination of attribute groups. Interpretation of the AUC values does with the two approaches, namely statistically and clinically approach. The use of the two approaches in order to obtain a large selection of attribute combinations that produce optimal AUC. The addition of the risk factors to attribute symptoms to the level-2 produced the AUC values change very significantly from 72.2% to 85.0%. The results with statistical considerations, suggests the addition of these attributes are very good, and when seen from cost considerations, examination of symptoms is relatively simple and not costly. The next level is the addition of rest ECG examination, AUC values generated relatively little increase, namely 0.4%. As for the addition of exercise ECG is able to give rise AUC value of 4.1% (89.1%). This means that checks the ECG provides additional AUC values are relatively high, when the examination of an ECG done during exercise. Refers to a combination of risk factors, symptoms and exercise ECG, scintigraphy examination addition, the AUC will provide additional value by 1.1%. Unfortunately, the value addition should require relatively expensive cost of inspection [17]. So taking into account the statistical and clinical approaches, the addition of such checks do not provide optimal diagnostic value, when it is only for the diagnosis stage. Things to do similar if in addition flouroscopy examination, or a combination of both does not provide a significant increase. Referring to the analysis by considering new interpretations AUC values, the statistical approach and clinicians, the attributes for diagnosis of coronary heart disease can be reduced, which is as shown in Table 3.

Table 2. Tiered Multivariate Analysis with Logistic Regression

Tiered	Test Result Variables	AUC	*p-value*	*p-value* 95% Confidence Interval	
				Lower Bound	Upper Bound
1	Risk	,722	,000	,665	,779
2	Risk+Symtoms	,850	,000	,807	,893
3	Risk+Symtoms+Rest ECG	,854	,000	,812	,896
4	Risk+Symtoms+ Exercise ECG	,891	,000	,856	,927
5	Risk+Symtoms+Rest ECG +Exercise ECG	,893	,000	,858	,928
6	Risk+Symtoms+ Exercise ECG+Scintigraphy	,902	,000	,868	,936
7	Risk+Symtoms+ Exercise ECG +Flouroscophy	,907	,000	,875	,940
8	Risk+Symtoms+ Exercise ECG +Scintigraphy+ Flouroscophy	,915	,000	,883	,946

Table 3. List of Attribute Coronary Heart Disease

No	Attribute	Group	Selected
1	Age	Risk Factor	✓
2	Gender	Risk Factor	✓
3	Chest Pain Type	Symptoms	✓
4	Systolic Blood Pressure (mmHg)	Risk Factor	
5	Cholesterol (mg/dl)	Risk Factor	
6	Fasting Blood Sugar	Risk Factor	
7	Resting ECG	Rest ECG	
8	Maximum heart rate achieved	Exercise ECG	✓
9	Exercise induced angina	Exercise ECG	✓
10	ST Depression induced by exercise relative to rest	Exercise ECG	✓
11	The slope of the ST segment for peak exercise	Exercise ECG	✓
12	Number of Major vessel colored by Flouroscophy	Floroscopy	
13	Defect Type by Scintigraphy	Scintigraphy	

Attributes resulted in the feature selection process, will now be entered in the classification process. The classification process is done using MLP-NN algorithm. The composition of the data used is 70% to 30% for training and testing, the division of the composition done randomly. The MLP-NN architecture used consisted of several parts, the first part comprises 7 inputs layer and 21 neurons. The second part, amounting 2 hidden layers, each hidden layer neuron having 8 and 6 neurons with sigmoid activation function. The third part consists of a second layer output neuron with sigmoid activation function. Based on test results show that, the proposed system provides the performance to value sensitivity 84.80%, specificity 88.20%, positive prediction value (PPV) 90.03%, negative prediction value (NPV) 81.80%, accuracy 86,3% and area under the curve (AUC) of 92.1%.

System performance for AUC parameters can be represented in a graphic which y-axis is sensitivity and x-axis as specificity, as shown in Figure 2. Referring to the statistical interpretation, the performance of the proposed system with AUC values of 92.1%, including in the very good category (90% -100%).

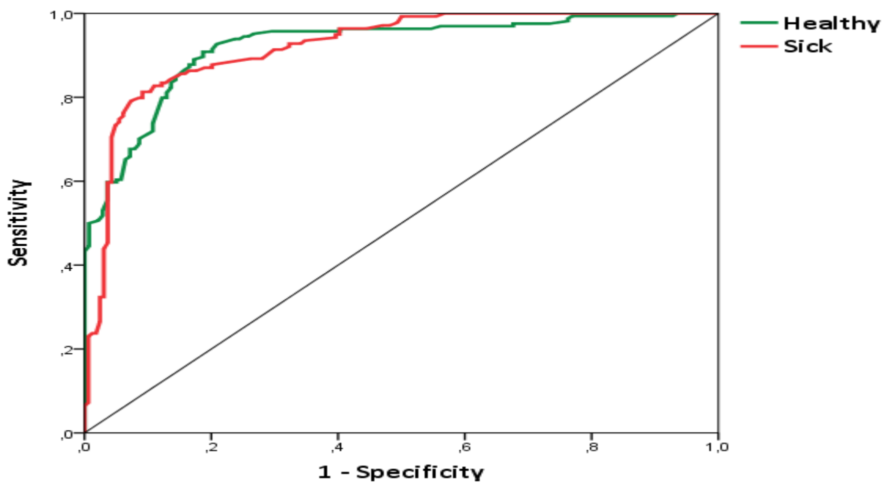

Figure 2. Graphic of AUC

In this study can be shown the percentage of the degree of urgency of each attribute which has been reduced as shown in Table 4. The order of the importance level of the highest attribute is a type of chest pain, ST depression induced by exercise relative to rest, the slope of the ST segment for peak exercise, gender, age, maximum heart rate and exercise-induced angina achieved. Chest pain is a symptom of that type is very important and typical as a sign of coronary heart disease. Chest pain type consists of typical angina, atypical angina, non-angina pain and asymptomatic. This type can be identified from three things: the discomfort or pain is felt behind the breastbone with the quality and length of a typical, pain triggered by activity or emotional stress and pain subside when a break or are given nitroglycerin [18]. The next attribute that has a percentage of 80% above the level of urgency is whether or not there either when resting ST depression, exercise and gender risk factors. ST depression with certain conditions is the description that the cells of the myocardium began to lack of oxygen. Further attributes gender, for men have a greater risk of heart attack and happened earlier than in women [19], while morbidity in men, two times greater than women, and this is the case almost 10 years earlier in men than women [20]. The risk factors have percentage levels of urgency age above 60%, lower than gender. Age person more susceptible to coronary heart disease, but rarely cause serious disease before 40 years and increased 5-fold at the age of 40-60 years [21]. Attributes that have the lowest percentage of 7 attribute is exercise-induced angina, which is 41.9% or below 50%. The low percentage of the degree of urgency of the system diagnosis is possible only attributed describes the presence or absence of induced angina in the examination.

Table 4. Independent attribute importance

No Attribute	Importance	Normalized Importance
1	0,119	61,1%
2	0,163	83,9%
3	0,194	100,0%
8	0,102	52,7%
9	0,081	41,9%
10	0,176	90,4%
11	0,164	84,7%

The proposed system, we compare with previous studies using several parameters such as, the accuracy and the number of attributes. Comparisons are grouped into two, namely, the comparison with the research group that has lower accuracy performance, and higher than the proposed system. The studies will be compared as shown in Table 5. The study was conducted Anooj [22], Detrano [23], Mokeddem et.al [24], Bashir et.al [25], Marateb & Gouddarzi [8], Santhanam & Aphizibah [6], Khemphila & Boonjing [26] and Abdar et.al [27], providing performance accuracy was considerably lower, compared to the proposed system. Besides having lower accuracy performance, these studies also use one or two attributes costly in conducting the investigation, namely flouroscopy and scintigraphy [17]. Furthermore, if it refers to the number of attributes in research conducted by Anooj [22] which using Fuzzy Inference System is able to reduce attributes into six attributes. These six attributes are the attributes with costly examination, namely flouroscophy. Furthermore Mokeddem et.al [24] by using a wrapper feature selection, which is implemented by genetic algorithm and C4.5, produce a number of attributes that are less than the proposed system, but weaknesses attributes generated, has two attributes costly in examination.

The next comparison with research conducted Muthukruppan & Er [7], Abdar et.al [27], Wiharto et.al [28] and Subanya & Rajalaxmi [3]. These studies are able to provide higher accuracy performance as compared to the proposed system. Unfortunately, the high accuracy should still require costly attribute, namely scintigraphy examination and flouroscopy. In addition the number of attributes required in research Muthukruppan & Er [7] and Wiharto et.al [28] to produce an accuracy above 90% require a relatively large number of attributes compared to the proposed system. While the research conducted Abdar et.al [27], using logistic regression attribute is able to reduce from 13 to 6 attributes, and with C5.0 algorithms capable of generating an accuracy above 90%. Unfortunately, they still need about six attributes which costly, even negate the ECG examination. ECG examination costs are relatively cheaper compared to two flouroscopy examination and scintigraphy. In the proposed system has a number of attributes that even more than in research Abdar et.al [27], but these attributes can be found in the examination of risk factors, symptoms and ECG. While the research Abdar et.al [27] requires the examination of risk factors, symptoms, flouroscophy and scintigraphy, resulting from the examination group more.

The use of a tiered method in the process of feature selection with logistic regression, is able to provide AUC values were better than without using a tiered approach. This is shown AUC value of 0.891 tiered, while not using a tiered approach as in research Abdar et.al [27] is 0.835. Furthermore, when combined with classification algorithms, for logistic regression with tiered approach capable of providing AUC value of 0.921, while research by Abdar et.al [27] which combined with C5.0 only able to provide AUC value of 0.869. Furthermore, if the tiered logistic regression method with a combination of 7 attributes provides AUC 0.891, whereas when combined with MLP-NN, capable of providing a higher AUC is 0.921. Tiered logistic regression method using all the attributes is also capable of providing AUC value of 0.915, the value is still lower than the proposed system.

Table 5. Comparation research

Author	Method	Nomor Attribute	Accuracy
Anooj [22	Fuzzy Inference System (FIS)	1,4,5,8,10,12	68.35%
Khemphila & Boonjing [26]	Decision Tree	1,2,3,4,5,6,7,8,9,10,11,12,13	73,30%
Detrano et.al [23]	Probability Theory (Logistic Regression)	1,2,3,4,5,6,7,8,9,10,11,12,13	77,00%
Khemphila & Boonjing [26]	Logistic Regression	1,2,3,4,5,6,7,8,9,10,11,12,13	77,70%
Mokaddem et.al [5]	GA Wrapper + C4.5	3,6,12,13	78,54%
Mokaddem et.al [5]	GA Wrapper + MLP-NN	1,2,3,4, 11,12,13	79,86%
Khemphila & Boonjing [26]	Neural Network	1,2,3,4,5,6,7,8,9,10,11,12,13	80,20%
Abdar et.al [27]	Logistic Regression + Neural Network	2,3,4,9,12,13	80,23%
Bashir et.al [25]	Majority Voting (Naive Bayesian, Decision Tree, SVM)	1,2,3,4,5,6,7,8,9,10,11,12,13	81,82%
Mokaddem et.al [5]	GA Warpper + SVM	1,3,9,10,11,12,13	83,82%
Marateb & Goudarzi [8]	Multiple Logistic Regression + Neuro-Fuzzy Classifier	1,8,9,10,12,13	84,00%
Mokaddem et.al [5]	GA Wrapper + Naive Bayesian	2,3,7,10,11,12,13	85,50%
Santhanam & Aphzibah [6]	GA+Fuzzy Inference System (FIS)	2,5,8,10,12,13	86.00%
Abdar et.al[27]	Logistic Regression + Support Vector Machine	2,3,4,9,12,13	86,05%
Subanya & Rajalaxmi [3]	Artificial Bee Colony + SVM	1,5,6,7,8,11,12	86,76%
Abdar et.al[27]	Logistic Regression + kNN	2,3,4,9,12,13	88,37%
Wiharto et.al [28]	k-start	1,2,3,4,5,6,7,8,9,10,11,12,13	92,02%
Abdar et.al[27]	Logistic Regression + C5.0	2,3,4,9,12,13	93,02%
Muthukaruppan & Er [7]	Particle Swarm Optimization+ Fuzzy Inference	3,4,5,7,8,10,11,12,13	93,27%
Arjenaki et.al [9]	GA + Naive Bayesian	1,3,4,6,7,8,9,10	85,18%
Porposed	Tiered Multivariate Analysis+MLP-NN	1,2,3,8,9,10,11	86,30%

Subsequent research conducted by Arjenaki et.al [9], which combines genetic algorithm with naive Bayesian. Fitness function used in the genetic algorithm is a function of the examination fee. The combination of these two attributes can reduce costly for the examination is scintigraphy and flouroscopy. On research by Arjenaki et.al [9], still need laboratory tests for fasting blood sugar attribute determines. In this proposed system does not require laboratory examination. In the study of Arjenaki et.al [9] also still has a number of attributes more than the proposed research. In the course of a study Arjenaki et.al [9] amounted to 8 attributes, whereas the proposed amount to 7 attributes. Furthermore, the resulting accuracy performance of the proposed system with 7 attributes better than using 8 attributes in research Arjenaki et.al [9].

4. CONCLUSION

Diagnosis using hybrid system tiered multivariate analysis model and multi-layer perceptron neural network, capable of delivering the performance of accuracy 86.3%, sensitivity 84.80%, specificity 88.20%, PPV 90.03%, NPV 81.80% and AUC 92,1%. The performance, when viewed from the relative accuracy of better than some previous studies, with the examination cost is relatively cheap, fast and examination results are obtained at low risk. In addition the number of attributes the results of feature selection are relatively little that is 7 attributes, with the percentage of the highest urgency level attribute type of chest pain and exercise-induced angina lows. Referring to the value of the parameter AUC, the proposed system included in the category of very good.

REFERENCES

[1] J. Kim, *et al.*, "Data-Mining-Based Coronary Heart Disease Risk Prediction Model Using Fuzzy Logic and Decision Tree," *Healthcare Informatic Research (HIR)*, vol/issue: 21(3), pp. 167-174, 2015.
[2] A. Young, "Handbook of Pattern Recognition and Image Processing 1st Edition," *Academic Press*, 1986.
[3] B. Subanya and D. R. R. Rajalaxmi, "Feature Selection using Artificial Bee Colony for Cardiovascular Disease Classification," *International Conference on Electronics and Communication System. Coimbatore*, pp. 1-6, 2014.
[4] H. S. N. Murthy and M. Meenakshi, "Dimensionality Reduction Using Neuro-Genetic Approach for Early Prediction of Coronary Heart Disease," *International Conference on Circuits, Communication, Control and Computing (I4C)*, pp. 329-332, 2014.
[5] R. O. Duda, *et al.*, "Pattern Classification," New York: John Wiley & Sons, 2012.
[6] T. Santhanam and E. P. Ephzibah, "Heart Disease Prediction Using Hybrid Genetic Fuzzy Model," *Indian Journal of Science and Technology*, vol/issue: 8(9), pp. 797–803, 2015.
[7] S. Muthukaruppana and M. J. Er, "A hybrid particle swarm optimization based fuzzy expert system for the diagnosis of coronary artery disease," *Expert Systems with Applications*, vol/issue: 39(14), pp. 11657–11665, 2012.
[8] H. R. Marateb and S. Goudarzi, "A noninvasive method for coronary artery diseases diagnosis using a clinically-interpretable fuzzy rule-based system," *Journal of Research in Medical Sciences*, vol/issue: 20(3), pp. 214-223, 2015.
[9] H. G. Arjenaki, *et al.*, "A Low Cost Model for Diagnosing Coronary Artery Disease Based On Effective Features," *International Journal of Electronics Communication and Computer Engineering*, vol/issue: 6(1), pp. 93-97, 2015.
[10] Y. Zhang, *et al.*, "Studies on application of Support Vector Machine in diagnose of coronary heart disease," *Sixth International Conference on Electromagnetic Field Problems and Applications (ICEF), Dalian, Liaoning*, pp. 1-4, 2012.
[11] B. Amma N. G., "An Intelligent Approach Based on Principal Component Analysis and Adaptive Neuro Fuzzy Inference System for Predicting the Risk of Cardiovascular Diseases," *Fifth International Conference on Advanced Computing (ICoAC)*, pp. 241-245, 2013.
[12] UCI Archive, "Machine Learning Repository," https://archive.ics.uci.edu/ml/machine-learning-databases/heart-disease/ (accessed 24 August 2016).
[13] D. Mannering, *et al.*, "Accurate detection of triple vessel disease in patients with exercise induced ST segment depression after infarction," *Br Heart J.*, vol/issue: 57(2), pp. 133-138, 1987.
[14] F. Gorunescu, "Data Mining Concepts, Models and Techniques," *Verlag Berlin Heidelberg: Springer*, 2011.
[15] L. Fausett, "Fundamentals of Neural Networks Architecture, Algorithms, and Applications," New Jersey: Pentice-Hall Inc, 1994.
[16] S. H. Jung, *et al.*, "Prediction Data Processing Scheme using an Artificial Neural Network and Data Clustering for Big Data," *International Journal of Electrical and Computer Engineering (IJECE)*, vol/issue: 6(1), pp. 330-336, 2016.
[17] D. T. Mangano, *et al*., "Dipyridamole Thallium-201 Scintigraphy as a Preoperative Screening Test A Reexamination of Its Predictive Potential," *Circulation*, vol/issue: 8(2), pp. 493-502, 1991.
[18] L. K. Hermann, *et al.*, "Comparison of Frequency of Inducible Myocardial Ischemia in Patients Presenting to Emergency Department With Typical Versus Atypical or Nonanginal Chest Pain," *The American Journal of Cardiology*, vol/issue: 105(11), pp. 1561-1564, 2009.
[19] T. F. M. v. Berkel, *et al.*, "Impact of smoking cessation and smoking interventions in patients with coronary heart disease," *European Heart Journal*, vol/issue: 20(24), pp. 1773–1782, 1999.
[20] H. H. Gray, "Lecture notes," Jakarta: Erlangga Medical Series, 2002.
[21] S. A. Price and L. M. Wilson, "Pathophysiology: Clinical Concepts of Disease Processes 6e," *Mosby*, 2002.

[22] P. K. Anooj, “Implementing Decision Tree Fuzzy Rules in Clinical Decision Support System after Comparing with Fuzzy based and Neural Network based systems,” *International Conference on IT Convergence and Security (ICITCS). Macao,* pp. 1-6, 2013.
[23] R. Detrano, *et al.*, “International Application of a New Probability Algorithm for the Diagnosis of Coronary Artery Disease,” *The American Journal of Cardiology,* vol/issue: 64(5), pp. 304-310, 1989.
[24] S. Mokeddem, *et al.*, “Supervised Feature Selection For Diagnosis of Coronary Artery Disease Based on Genetic Algorithm,” *Computer Science & Information Technology (CS & IT)*, vol/issue: 3(3), pp. 41-51, 2013.
[25] S. Bashir, *et al.*, “An Ensemble based Decision Support Framework for Intelligent Heart Disease Diagnosis,” *International Conference on Information Society (i-Society),* pp. 259-264, 2014.
[26] A. Khemphila and V. Boonjing, “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),* pp. 193 -198, 2010.
[27] M. Abdar, *et al.*, “Comparing Performance of Data Mining Algorithms in Prediction Heart Diseases,” *International Journal of Electrical and Computer Engineering (IJECE)*, vol/issue: 5(6), pp. 1569-1576, 2015.
[28] W. Wiharto, *et al.*, “Intelligence System for Diagnosis Level of Coronary Heart Disease with K-Star Algorithm,” *Healthcare Informatic Research (HIR)*, vol/issue: 22(1), pp. 30-38, 2016.