An Overview of Data Mining Techniques Applied for Heart Disease Diagnosis and Prediction

Salha M. Alzahani, Afnan Althopity, Ashwag Alghamdi, Boushra Alshehri, and Suheer Aljuaid
Dept. of Computer Science, College of Computers and Information Technology
Taif University, Taif, Saudi Arabia
Email: s.zahrani@tu.edu.sa

***Abstract* — Data mining techniques have been applied magnificently in many fields including business, science, the Web, cheminformatics, bioinformatics, and on different types of data such as textual, visual, spatial, real-time and sensor data. Medical data is still information rich but knowledge poor. There is a lack of effective analysis tools to discover the hidden relationships and trends in medical data obtained from clinical records. This paper reviews the state-of-the-art research on heart disease diagnosis and prediction. Specifically in this paper, we present an overview of the current research being carried out using the data mining techniques to enhance heart disease diagnosis and prediction including decision trees, Naive Bayes classifiers, K-nearest neighbour classification (KNN), support vector machine (SVM), and artificial neural networks techniques. Results show that SVM and neural networks perform positively high to predict the presence of coronary heart diseases (CHD). Decision trees after features reduction is the best recommended classifier to diagnose cardiovascular disease (CVD). Still the performance of data mining techniques to detect coronary arteries diseases (CAD) is not encouraging (between 60%-75%) and further improvements should be pursued.**

***Index Terms*—heart disease, data mining, decision tree, naive bayes, K-nearest neighbor, support vector machine**

I. Introduction

Knowledge discovery in data is defined as: "the extraction of hidden previously unknown and potentially useful information about data" [1]. Basically knowledge discovery in data is the process of extracting different features from data in various steps. Fig.1 shows the process of Knowledge discovery from various data sources in a specific domain. Data mining is the heart (core) step, which results in the discovery of implicit but potentially valuable knowledge from huge amount of data. Data mining technology provides the user with the methods to find new and implicit patterns from massive data. In the healthcare domain, discovered knowledge can be used by the healthcare administrators and medical physicians to improve the accuracy of diagnosis, to enhance the goodness of surgical operations and to reduce the harmful effects of drug [2], [3]. It aims also to propose less expensive therapeutic [4].

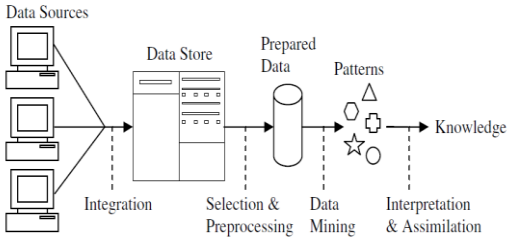

Figure 1. Process of knowledge discovery in data.

The diagnosis of diseases is a difficult but critical task in medicine. The detection of heart disease from "various factors or symptoms is a multi-layered issue which is not free from false presumptions often accompanied by unpredictable effects" [5]. Thus, we can use patients' data that have been collected and recorded to ease the diagnosis process and utilize knowledge and experience of numerous specialists dealt with the same symptoms of diseases. Providing invaluable services with less costs is a major constraint by the healthcare organizations (hospitals, polyclinics, and medical centres). According to [6], "valuable quality service denotes the accurate diagnosis of patients and providing efficient treatment. Poor clinical decisions may lead to disasters and hence are seldom entertained". Besides, it is essential that the hospitals decrease the cost of clinical tests. Using professional and expert computerized systems based on machine-learning and data mining methods should help in one direction or another with achieving clinical tests or diagnosis at reduced risks [7], [8].

This paper aims to provide a survey of current techniques of knowledge discovery using data mining techniques applied to medical research; particularly, to heart disease prediction. Literature studies between 2010 and 2014 are discussed, unless a significant study before that should be mentioned. A number of experiments and research works have been done to compare the performance of predictive data mining techniques like decision tree, Naive Bayes, K-nearest neighbour, support vector machine and artificial neural networks. This paper discussed the results of the state-of-the-art techniques and gives conclusions towards future research.

Manuscript received July 18, 2014; revised December 15, 2014.

doi: 10.12720/lnit.2.4.310-315

II. Heart Diseases: An Overview

The heart can be affected by diverse types of diseases most of them are dangerous on human lives. Coronary heart diseases, cardiomyopathy and cardiovascular diseases are some examples of heart diseases, as shown in Fig. 2.

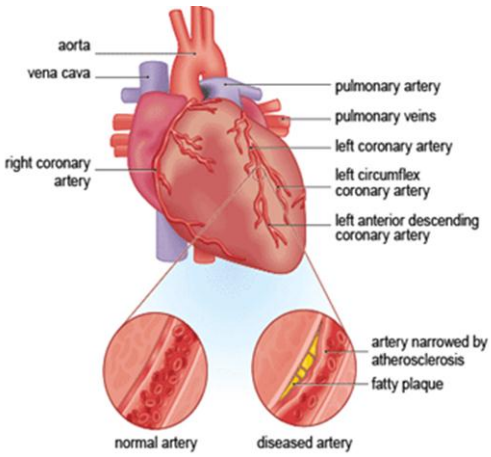

Figure 2. Heart arteries miseases.

The most common type of these diseases is *Coronary Arteries Disease* (CAD) wherein coronary arteries hard and tight [9]. The term *Cardiovascular Disease* (CVD) denotes a wide range of conditions that affect the heart and the blood vessels, and the manner in which the blood is pumped and circulated throughout the body [4], [10]. CVD results in severe illness, disability, and is most likely to cause death. Narrowing of the coronary arteries causes the reduction of oxygen supplied to the heart and leads to the so-called *Coronary Heart Disease* (CHD) [4], [10]. A sudden blockage of a coronary artery is generally due to a blood clot which may cause a heart attack. Chest pains arise when the blood received by the heart muscles is inadequate and unconnected [4].

There are many remarks and symptoms used by the physicians to diagnose heart diseases. Age, sex, chest pain type, blood pressure, cholesterol, fasting blood sugar, maximum heart rate, and hereditary are meaningful symptoms [6], [11]. Besides, other habits might be used including stress, overweight, smoking, alcohols intake and less exercise [6].

III. Data Mining Techniques in Health Care

After exploring some types of heart arteries diseases and symptoms that when have certain values denote a heart disease, in this section we will explore different data mining techniques applied generally to healthcare. Symptoms and patient records can be used as features and such huge amounts of data can be used for knowledge discovery in the health care domain. A general framework proposed by [2] for medical data mining is shown in Fig. 3. The framework starts with a specific medical problem wherein a dataset should be pre-processed and cleaned before mining the data using one of the available data mining tools. Knowledge evaluation comes at the last and expertise from the medical domain should involve.

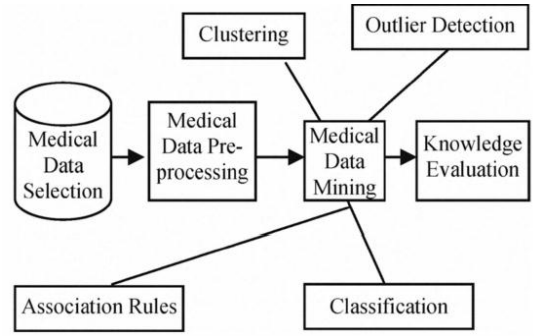

Figure 3. Framework for medical data mining

A. Neural Networks

Neural networks are biologically inspired highly interconnected cells that simulate the human brain [1]. The perceptron is the simplest architecture which has one neuron and a learning method. More sophisticated architecture is multi-layer neural networks (MLP) which one or more neurons connected at different layers. Neural networks can be trained to learn a classification task and to predict diseases [6], [12].

B. K-nearest Neighbor Algorithm (KNN)

K-nearest neighbour classification algorithm is a well-known method for classifying an unseen instance using the classification of the instances closest to it [1]. Basic KNN classification algorithm works by finding K training instances that are close to the unseen instance using distance measures such as Euclidean, Manhattan, maximum dimension distance, and others. Then, the algorithm decides the class for the unseen instance by taking the most commonly occurring class in the nearest K instances [1].

C. Decision Tree Classification Algorithm

Decision trees are powerful classification algorithms used alternatively as decision/classification rules [1]. Popular decision tree algorithms include Quinlan's ID3, C4.5, C5, and Breiman *et al.*'s CART [13]-[15]. As its name implies, this classification technique works by recursively constructing branches of the tree based on certain observations (or variables). A well-known algorithm to construct the branches is called top-down induction of decision trees (TDIDT) [1]. Decision trees are easily constructed with binary or categorical variables but the mission becomes harder with numerical variables. A corresponding threshold value must be specified for the later based on some mathematical or observational considerations in order to be able to construct the branches of the tree. This step is repeated at each leaf node until the complete tree is constructed ending with leaves which gives one of the classes or predictions in the dataset. The objective of the splitting algorithm is "to find a variable-threshold pair that maximizes the homogeneity of the resulting two or more subgroups of samples" [16].

1) C4.5 classification algorithm

One of the best decision tree algorithms is C4.5. This algorithm can manage continuous data in numerical forms using pruning algorithms which aim to simplify the

classification rules without any loss of prediction accuracy. Only the most important features are kept whereby they lowered the error rates. One of its factors is denoted by *M* which indicates the minimum instances that a leaf should have. *C* means the confidence threshold which is considered for pruning. By changing these two factors, the accuracy of algorithm can be increased and the error can be decreased [9].

2) RIPPER classification algorithm

RIPPER stands for Repeated Incremental Pruning to Produce Error Reduction. This classification algorithm which was proposed by Cohen [17], is based on association rules with reduced error pruning (REP), a very common and effective technique found in decision tree algorithms [16]. To generate association rules using REP algorithm, the training data is divided into a growing set and a pruning set. The growing set is the initial association rules which can be generated purely from the dataset using some heuristic methods. The growing set contains a huge set of rules that should be repeatedly simplified to form the pruning set. Thus, the simplification is done using typical pruning operators which may allow to delete a term from any single rule or from different association rules. The pruning operator chosen for simplification should give the most accurate rule with the greatest reduction of error. The simplification process terminates when applying the pruning operator would increase the error value on the pruning set [16].

D. Support Vector Machine (SVM)

SVM is a state-of-the-art maximum margin classification algorithm rooted in statistical learning theory [16]. It is a method for classification of both linear and non-linear data. The training data is converted into n-dimensional data using non-linear transformation method. Then, the algorithm searches for the best hyper-plane to separate the transformed data into two different classes. SVM performs classification tasks by maximizing the margin of the hyper-plane separating both classes while minimizing the classification errors.

E. Naïve Bayes Algorithm

One of the Bayesian methods is Naïve Bayes classifiers which uses the probabilistic formula:

$$P(A \mid B) = \frac{P(B \mid A) \times P(A)}{P(B)} \alpha + \beta = \chi \qquad (1)$$

where *A* and *B* are two events (e.g. the probability that the train will arrive on time given that the weather is rainy). Such Naïve Bayes classifiers use the probability theory to find the most likely classification of an unseen (unclassified) instance [1]. The algorithm performs positively with categorical data but poorly if we have numerical data in the training set [9].

IV. Data Mining Techniques for Heart Disease Diagnosis and Pprediction

In this section, a survey of medical data mining techniques applied for diagnosis and prediction of some types of heart diseases is presented. Literature studies from 2010 and above are discussed, unless a significant study before that should be mentioned. Symptoms denoting a heart disease are processed. Standard datasets for each heart disease are prepared including certain features each with a predefined range of values. Such datasets have been used for knowledge discovery of heart problems in several research studies [4]-[11], [15], [16], [18]-[20].

As mentioned in Section II, the most common type of heart diseases is CAD (*Coronary Arteries Disease*). For diagnosing CAD, different classification algorithms have been implemented [9]-[20]. In these studies, three features of vessel's stenosis have been employed namely *Left Anterior Descending* (LAD), *Left Circumflex* (LCX) and *Right Coronary Artery* (RCA). The ones whom LAD, LCX or RCA vessel is clogged are classified as CAD patients, others as healthy. Table I summarises the results obtained from different classification algorithms for diagnosis of CAD specifically.

In [20], neural networks algorithm was used for prediction of stenosis of each vessel separately. Ten-fold cross-validation methods were involved to measure the accuracy. A multi-layered perceptron neural network was employed for the classification. The accuracy reached 73%, 64.85% and 69.39% for LAD, LCX and RCA vessels, respectively.

Another study [9] applied Naïve Bayes, C4.5, and KNN classification algorithms using more sophisticated features that have not been applied in [20]. The study aimed to diagnose CAD via the stenosis of each LAD, LCX or RCA vessel separately. RapidMiner tool was used and the accuracy was measured using 10-fold cross validation technique. The best accuracy was obtained by C4.5 wherein achieved accuracy were 74.20%, 63.76%, and 68.33%, respectively. The accuracy obtained using C4.5 is the ideal one for diagnosing CAD via LAD stenosis and was not achieved by previous studies.

Another types of heart attacks are CVD and CHD explained briefly in Section II. Different studies have explored the prediction of these heart diseases but unlike previous studies, without considering the stenosis of each LAD, LCX or RCA vessel separately. Table II shows the results obtained from different studies to diagnose CVD and CHD diseases. In [10], CVD and CHD were predicted using three different supervised machine learning algorithms namely Naïve Bayes, KNN, decision list techniques. Tanagra tool was used to classify the data, the experiments were conducted using 10-fold cross validation, and the results were compared. The accuracy from Naïve Bayes, KNN, and decision list were 52.33%, 52%, 45.67%, respectively. Naive Bayes algorithm showed slightly better performance compared with the other algorithms.

Moreover, CVD heart disease prediction were further analysed using four data mining classification techniques namely RIPPER classifier, decision tree, artificial neural network, and SVM [16]. The results were compared using 10-fold cross validation method and the accuracy obtained were as follows: 81.08%, 79.05%, 80.06%, and 84.12%, respectively. Their analysis shows that out of these four classification models, SVM predicts cardiovascular

disease with the highest accuracy. A study by [19] used genetic algorithms (GA) to reduce the actual data size by reducing the number of attributes used in previous studies. Thus, it aims to get the optimal subset of attributed sufficient for CVD prediction. The study then applied three classifiers namely Naïve Bayes, decision tree, and classification via clustering. As clustering is the process of grouping similar instances into different groups or clusters, it was used as a pre-processing step before feeding the data to the classifying model. Experiments were conducted using WEKA 3.6.0 tool. The results obtained from the three classifiers outperform previous results in the literature reaching 96.5%, 99.2%, and 88.3, respectively. Another study [21] used Naïve Bayes and decision trees to detect CVD and achieved 62.03%, and 60.40% of accuracy, respectively.

TABLE I. ACCURACY OF NEURAL NETWORK, NAÏVE BAYES, C4.5 AND KNN CLASSIFICATION ALGORITHMS FOR DIAGNOSIS OF CORONARY ARTERIES DISEASE (CAD)

Data Mining Techniques	Accuracy (LAD)	Accuracy (LCX)	Accuracy (RCA)	Ref.
Neural Network	73%	**64.85%**	**69.39%**	Babaoglu *et al.* [20]
Naïve Bayes	51.81%	62.73%	67.29%	Alizadehsani *et al.* [9]
C4.5	**74.20%**	63.76%	68.33%	
KNN	59.65%	61.39%	59.11%	

a. Studies that consider LAD, LCX, and RCA vessel separately to diagnose CAD

TABLE II. ACCURACY OF DECISION TREE, NAÏVE BAYES AND CLASSIFICATION VIA CLUSTERING ALGORITHMS FOR DIAGNOSIS OF CARDIOVASCULAR DISEASE (CVD) AND CORONARY HEART DISEASE (CHD)

Data Mining Techniques	Accuracy	Ref.
Naïve Bayes	52.33%	Rajkumar and Reena [10]
KNN	52%	
Decision List	45.67%	
RIPPER Classifier	81.08%	Kumari and Godara [16]
Decision Tree	79.05%	
Artificial Neural Network	80.06%	
SVM	84.12%	
Naïve Bayes	96.5%	Anbarsi *et al.* [19]
Decision Tree	99.2%	
Classification via Clustring	88.3%	
Naïve Bayes	62.03%	Sitar-Taut, *et al.* [21]
Decision Tree	60.40%	
Hybrid Genetic Neural Network	89%	Amin *et al.* [22]
C4.5	82.5%	Srinivas *et al.* [23]
Multilayer Neural Network (MLP)	89.7%	
Bayesian Classifier	82%	
SVM	82.5%	
C5 Classifier	89.6%	AbuKhousa and Campbell [12]
MLP	91.0%	
SVM	92.1%	

a. Studies that do not consider LAD, LCX, and RCA vessel separately to diagnose CVD and CHD diseases.

A recent study [22] used a hybrid genetic neural network method to detect CVD heart disease and achieved 89% of accuracy.

On the other hand, a number of studies have explored CHD heart diseases particularly. A study to predict CHD were employed using WEKA and applying four data mining techniques namely C4.5, multilayer neural network (MLP), Bayesian classifier, and SVM [23]. The results obtained showed that out of these four models, MLP obtained the highest accuracy (89.7%) compared with the others. Another three prediction models for CHD were implemented in [12] namely C5 classifier, MLP and SVM. The results obtained from 10-forld cross validation method showed that SVM was the best predictor gaining the accuracy of 92.1%.

TABLE III. EFFECTIVENESS OF DATA MINING TECHNIQUES USED FOR HEART DISEASE DIAGNOSIS AND PREDICTION

Data Mining Techniques	Purpose of study	*Maximum Accuracy*	Ref.
Neural Network	To diagnose the presence of CHD.	91.0%	AbuKhousa and Campbell [12]
Naïve Bayes Classifier *+GA Feature Reduction*	To diagnose the presence of CVD.	96.5%	Anbarsi *at el.* [19]
Decision Tree *+GA Feature Reduction*	To diagnose the presence of CVD.	**99.2%**	Anbarsi *at el.* [19]
RIPPER Classifier	To diagnose the presence of CVD.	81.08%	Kumari and Godara [16]
C4.5 Classifier	To diagnose CAD via stenosis of LAD vessel	**74.20%**	Alizadehsani *et al.* [9]
	To diagnose the presence of CHD.	82.5%	Srinivas *et al.* [23]
C5 Classifier	To diagnose the presence of CHD.	89.6%	AbuKhousa and Campbell [12]
SVM	To diagnose the presence of CHD.	**92.1%**	AbuKhousa and Campbell [12]
Clustering	To diagnose the presence of CVD.	88.3%	Anbarsi *at el.* [19]
KNN	To diagnose CAD via stenosis of LCX vessel.	61.39%	Alizadehsani *et al.* [9]
Hybrid Genetic Neural Network	To diagnose the presence of CVD.	89%	Amin *et al.* [22]

a. Maximum accuracy obtained by each technique to diagnose a certain type of heart disease.

V. DISCUSSION

As you have seen, the performance of different data mining algorithms on the heart disease datasets has been studied and compared. The accuracy of these techniques have been measured using 10-fold cross validation method on a standard dataset of each disease. Almost the same features have been employed for the prediction of each type of disease. Therefore, these techniques can be analysed and compared on a larger scale as you will see below.

In this section, several medical data mining techniques for heart diseases diagnosis focusing on three famous types namely CAD, CVD and CHD are discussed. The effectiveness of different techniques are shown in Table

III. For each technique, the maximum accuracy rate obtained for a certain disease is shown. For example, the maximum accuracy rate obtained using neural network is 91.0% to predict CHD by AbuKhousa and Campbell [12] whereas Naïve Bayes classifier achieved the maximum accuracy of 96.5% to predict CVD by Anbarsi *at el.* [19]. The table shows ten different techniques and their maximum accuracy rate for heart disease prediction (i.e. patient classification as healthy or affected by certain disease).

Based on such evaluation, the highest and uppermost accuracy to diagnose each type of disease using a certain data mining technique can be investigated (shown in **bold** font in the table). We can recommend the following: (i) C4.5 classifier perform better than other data mining techniques to diagnose CAD via stenosis of the LAD vessel, followed by KNN via the stenosis of LCX vessel. (ii) SVM and neural networks perform comparably and positively high; therefore, they can be utilised to predict the presence of CHD. (iii) Decision trees method after the reduction and optimization of features using GA is the best recommended classifier to diagnose CVD heart disease.

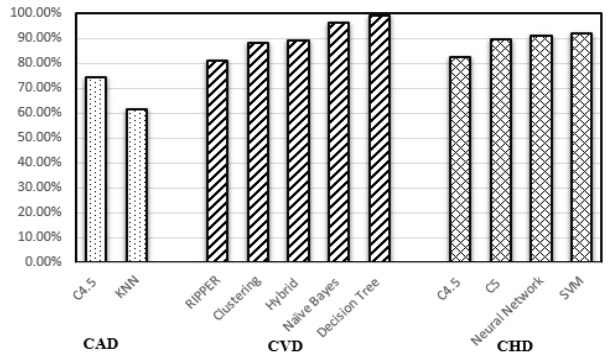

Figure 4. Comparison of different medical data mining techniques for heart diseases prediction.

Fig. 4 shows the performance evaluation metrics of data mining techniques for heart disease prediction. It is observed that the accuracy of various classification techniques for CVD diagnosis is highly encouraging (between 85% and 99%). Consequently, diagnosis systems that employs classifiers or clusters can assist the medical professionals in making decision about CVD early diagnosis. Moreover, data mining techniques perform positively well with diagnosing CHD (achieving accuracy between 82% and 92%). Still the performance of classification methods to detect CAD diseases is not encouraging (between 60%-75%) whether or not the LAD, LCX, and RCA vessels are considered separately. Therefore, further research should be carried out using more sophisticated features and hybrid algorithms to improve the prediction of CAD diseases.

Due to the successful implementation of data mining techniques for heart diagnosis, various heart disease prediction systems that employed the aforementioned data mining techniques such as *Intelligent Heart Disease Prediction System (IHDPS)* [7] have been proposed. However, there some limitations in these systems. One of these weaknesses is the fact that physicians do not feel comfortable or having a good knowledge of using computerized systems for diagnosis. A recent survey [24] showed that existing systems are computational expensive and do not achieve optimum accuracy nor 100% reliable results.

VI. Conclusion and Future Work

Early diagnosis of heart diseases may save humans from heart attacks. This paper reviews the state-of-the-art data mining techniques applied for diagnosing three heart diseases namely CAD, CVD, and CHD. Among the famous and sever diseases is CAD which can be diagnosed via the stenosis of blood vessels. Such disease is data rich but unfortunately the obtained accuracy from CAD classifiers is poor. Data mining techniques applied for CVD and CHD are promising. Results showed that the optimization and feature reduction utilising GA or principle component analysis (PCA) for a certain disease may strongly increase the accuracy of a classifier. It is found that decision trees and Naïve Bayes classifiers are recommended for CVD diagnosis with an accuracy reaching more than 95%. Further, C5, SVM, and neural networks are the best recommended classifiers for CHD prediction. It is observed that the prediction results of various data mining classification techniques are strongly encouraging and would assist the physicians to do early diagnosis and make more accurate decisions. Nonetheless, data mining techniques do not achieve 100% accuracy for heart diseases prediction and hence cannot be utilised solely for diagnosis.

Future works should focus on improving the predication of CAD diseases utilising more features and separate combinations of vessel stenosis. Furthermore, feature reduction should be utilised in various ways to achieve better accuracy results with all diseases. New classifiers should be developed for other heart diseases and problems such as coronary microvascular diseases, pulmonary, and cyanotic heart diseases. This work can be further extended by working with different heart related datasets from health care organizations and agencies using all the available techniques and also using a combination of them.

References

[1] M. Bramer, *Principles of Data Mining*: Springer-Verlag, 2007.

[2] Z. Jitao and W. Ting, "A general framework for medical data mining," presented at the International Conference on Future Information Technology and Management Engineering (FITME), 2010.

[3] S. Tsumoto, "Problems with mining medical data," presented at the The 24th Annual International Computer Software and Applications Conference, 2000.

[4] K. Srinivas, B. K. Rani, and A. Govardhan, "Applications of data mining techniques in healthcare and prediction of heart attacks," *International Journal on Computer Science and Engineering*, vol. 2, pp. 250-255, 2010.

[5] S. B. Patil and Y. S. Kumaraswamy, "Extraction of significant patterns from heart disease warehouses for heart attack prediction," *International Journal of Computer Science and Networks Security*, vol. 9, pp. 228-235, 2009.

[6] S. Oyyathevan and A. Askarunisa, "An expert system for heart disease prediction using data mining technique: Neural network,"

International Journal of Engineering Research and Sports Science, vol. 1, pp. 1-6, 2014.

[7] S. Palaniappan and R. Awang, "Intelligent heart disease prediction system using data mining techniques," presented at the International Conference on Computer Systems and Applications, 2008.

[8] G. Subbalakshmi, "Decision support in heart disease prediction system using naive bayes," *Indian Journal of Computer Science and Engineering*, vol. 2, pp. 170-174, 2011.

[9] R. Alizadehsani, J. Habibi, B. Bahadorian, H. Mashayekhi, A. Ghandeharioun, and R. Boghrati, *et al.*, "Diagnosis of coronary arteries stenosis using data mining," *J Med Signals Sens*, vol. 2, pp. 153-9, Jul 2012.

[10] A. Rajkumar and G. S. Reena, "Diagonsis of heaer disease using data mining algorithm," *Global Journal of Computer Science and Technology*, vol. 10, pp. 38-43, 2010.

[11] M. A. Jabbar, B. L. Deekshatulu, and P. Chandra, "Graph based approach for heart disease prediction," in *Proc. Third International Conference on Trends in Information, Telecommunication and Computing*, vol. 150, 2013, pp. 465-474.

[12] Y. W. Xing, J. Wang, Z. H. Zhao, and Y. H. Gao., "Combination data mining methods with new medical data to predicting outcome of coronary heart disease," presented at the International Conference on Convergence Information Technology, 2007.

[13] L. Rokach and O. Maimon, "Top-down induction of decision trees classifiers - a survey," *IEEE Transactions on Systems, Man, and Cybernetics, Part C: Applications and Reviews*, , vol. 35, pp. 476-487, 2005.

[14] S. R. Safavian and D. Landgrebe, "A survey of decision tree classifier methodology," *IEEE Transactions on Systems, Man and Cybernetics*, vol. 21, pp. 660-674, 1991.

[15] S. Ranganatha, H. R. P. Raj, C. Anusha, and S. K. Vinay, "Medical data mining and analysis for heart disease dataset using classification techniques," presented at the National Conference on Challenges in Research & Technology in the Coming Decades, 2013.

[16] M. Kumari and S. Godara, "Comparative study of data mining classification methodsin cardiovascular disease prediction," *IJCST* vol. 2, pp. 304-305, 2011.

[17] W. W. Cohen, "Fast effective rule induction," presented at the Twelfth International Conference on Machine Learning, 1995.

[18] R. Alizadehsani, J. Habibi, Z. Alizadeh Sani, H. Mashayekhi, R. Boghrati, and A. Ghandeharioun, *et al.*, "Diagnosing coronary artery disease via data mining algorithms by considering laboratory and echocardiography features," *Res Cardiovasc Med*, vol. 2, pp. 133-139, 2013.

[19] M. Anbarsi, E. Anupriya, and N. Iyengar, "Enhanced prediction of heart disease with feature subset selection using genetic algorithm," *International Journal of Engineering Science and Technology*, vol. 2, pp. 5370-5376, 2010.

[20] I. Babaoglu, O. K. Baykan, N. Aygul, K. Ozdemir, and M. Bayrak, "Assessment of exercise stress testing with artificial neural network in determining coronary artery disease and predicting lesion localization," *Expert Systems with Applications*, vol. 36, pp. 2562-2566, 2009.

[21] V. A. Sitar-Taut, D. Zdrenghea, D. Pop, and D. A. Sitar-Taut, "Using machine learning algorithms in cardiovascular disease risk evaluation," *Journal of Applied Computer Science & Mathematics*, vol. 5, pp. 29-32, 2009.

[22] S. U. Amin, K. Agarwal, and R. Beg, "Genetic neural network based data mining in prediction of heart disease using risk factors," presented at the IEEE Conference on Information & Communication Technologies, 2013.

[23] K. Srinivas, G. R. Rao, and A. Govardhan, "Analysis of coronary heart disease and prediction of heart attack in coal mining regions using data mining techniques," presented at the 5th International Conference on Computer Science and Education, 2010.

[24] E. AbuKhousa and P. Campbell, "Predictive data mining to support clinical decisions: An overview of heart disease prediction systems," presented at the International Conference on Innovations in Information Technology, 2012.

Dr. Salha Alzahrani has obtained her Doctor of Philosophy (Computer Science) from the University of Technology Malaysia in 2012, and has conducted a research visit to the University of Oxford, UK in 2011. She obtained her MSc with excellent pass (among the top 10 students) from the University of Technology Malaysia in 2009, and BSc with honours (first rank) from Taif University in 2004. She is currently an assistant professor at Taif University. Her contribution to the field of Plagiarism Detection is illustrated by a number of indexed publications in high-impact journals and IEEE peer-reviewed conferences.

Afnan Althopity is working as a teaching assistant with the Department of Computer Science, Taif University. She obtained her BCs from Taif University with excellent pass. She is teaching various lab courses and introductory courses to computer science. She is also serving as a volunteer for the commnuty works such as Big Sister Little Sistser conducted in her college.

Ashwag Alghamdi, Boushra Alshehri, and **Suheer Aljuaid** are graduates from the Department of Computer Science, Taif University.