Comprehensive Study of Heart Disease Diagnosis Using Data Mining and Soft Computing Techniques

Gayathri. P [1] and N. Jaisankar [2]
Assistant Professor (Senior) [1] and Professor [2]
School of Computing Science and Engineering, VIT University, Vellore – 632014, Tamil Nadu, India.
pgayathri@vit.ac.in [1] and njaisankar@vit.ac.in [2]

Abstract **- Heart disease diagnosis is a challenging task which can offer automated prediction about the heart disease of patient so that further treatment can be made easy. Due to this fact, heart disease diagnosis has received immense interest globally among medical community. Here, artificial intelligence played an important role in diagnosis of heart disease with improved effectiveness. Based on this perspective, several researches have been conducted in the literature recently. So, analyzing those diagnosis techniques can lead to new development in this area. Accordingly, we present a detailed survey of 47 articles published in the standard journals from the year 2005 to 2013. The survey of the papers related to heart disease and also the survey of many categories of heart disease such as coronary heart disease, coronary artery disease, heart failure, ischemic heart disease, cardiovascular disease, congenital heart disease, valvular heart disease and hypoplastic left heart syndrome are presented in this paper. From the survey the finding is that neural network based techniques contribute more effectiveness and some techniques have obtained more than 90% accuracy. Finally, some of the research issue is also addressed to precede the further research on the same direction.**

Keywords - Heart disease, Disease diagnosis, Machine learning, Neural network

I. INTRODUCTION

Among various life- threatening diseases, heart diseases have a great deal of attention in medical research. Also, it has more impact on human health. Various heart diseases was discussed and founded how they lead to heart attack [32]. The number one cause of death in industrialized countries was due to cardiovascular disease. Cardiovascular diseases not only have a major impact on individuals and their quality of life in general, but also on public health costs and the countries' economies. Risk factors for these pathologies include diabetes, smoking, family history, obesity, high cholesterol etc [5]. Health information decision was enabled by particularly knowing about the anatomy and functioning of the heart. A new born infant also have the possibility of heart disease. Some of the symptoms of heart disease in people were chest pain and fatigue. It occurred while the heart does not meet the circulatory demands of the body [22]. The physician takes decision based on the patient's answers to questions and lab results [15].

Blood flow to the heart muscles was decreased when block occurs in coronary arteries. The electrocardiogram recordings were analyzed to detect irregularity of heart beat problems occurred due to cardiovascular diseases [34]. In advance of medical and surgical treatment the patient with heart disease reached adulthood [38]. There are many diseases that affect the heart and arteries but four are particularly prevalent. Myocardial infarction was linked to damage to the coronary arteries in 90% of cases. Strokes occurred as a result of impaired blood flow to the brain linked to a hemorrhage or a blockage of the arteries that supply blood to the brain. Heart failure was mainly linked to various changes in cardiovascular tissues, most often the result of ageing. High blood pressure was defined as the sustained elevation of arterial blood pressure in comparison to what is considered to be the "normal" value of 140/90 millimeters of mercury. There is a wide range of long-term consequences: heart failure, stroke, kidney failure etc.

The different types of heart disease widely in the world are Coronary heart disease, Heart failure, Coronary artery disease, Ischemic heart disease, Cardiovascular disease, Hypoplastic left heart syndrome, Atherosclerosis, Chronic obstructive pulmonary disease, Congenital heart disease, Valvular heart disease. Mostly heart attacks are occurred when the plaque on the artery ruptures and a clot then forms, stopping blood flow. And the diagnosis of heart disease was based on medical knowledge occurred from patients. Correct diagnosis of the heart patient was delayed due to various problems. Diagnosis of heart disease was more costly and optimal decision path finder was used in terms of diagnostic accuracy while minimizing cost in diagnosis [17]. Heart disease can strike suddenly and quick decisions have to be made. Prediction of heart diseases can provide some useful information about the health of patient. The prediction can be done with various computer-aided diagnosis methods.

Generally, artificial intelligence techniques were used in medical diagnosis with an improvement in prediction of heart disease [20]. Machine learning algorithm was used in medical diagnostic problem for heart disease [44]. Case-based reasoning (CBR) was considered as a suitable technique for diagnosis, prognosis and prescription in the medical domain puts more stress on real cases than other domains [18]. Coronary Artery Disease was diagnosed using two techniques called Binary Particle Swarm Optimization (BPSO) and Genetic Algorithm (GA) [20]. For the diagnosis of heart disease various classification and regression processes was used. It provides medical knowledge for diagnosis purpose [21].

II. EXISTING SURVEY OF HEART DISEASE DIAGNOSIS

In this section, 20 research articles related to heart disease have been reviewed.

Carlos Ordonez, [1] introduced an algorithm to reduce the number of rules which used search limitation. The introduced algorithm searches for association rules in a training set and finally validates them on an independent test set. In medical terms, to the degree of disease in four specific arteries, the association rules related heart perfusion measurements and risk factors. Association rules were applied on a real data set containing medical records of patients with heart disease. Search limitations and test set validation importantly reduced the number of association rules and produced a set of rules with high predictive accuracy. Unfortunately, when association rules were applied on a medical data set, they produced an extremely large number of rules. They used the train and test approach which used two disjoint samples from a data set to search and validate rules. To filter rules on the test set, support, confidence and lift have different importance. They opinioned that to validate rules confidence was the most important metric. Based on heart perfusion measurements and risk factors they used association rules to predict the degree of narrowing in four arteries. They presented medically significant rules discovered on medical data set that remain valid in several independent train/test cycles. The two problems were addressed such as large numbers of rules were obtained by the standard association rule algorithm and the validation of rules on an independent set, which was required to eliminate unreliable rules.

K.C. Tan and E.J. Teoh *et al,* [4] have proposed a hybrid approach consist of two conventional machine learning algorithms. Genetic Algorithms (GAs) and Support Vector Machines (SVMs) were the two proposed algorithm combined effectively based on a wrapper approach. Here, by an evolutionary process genetic algorithm component searches for the best attribute data set. Based on the attribute subset represented by GA, the SVM classified the patterns into reduced data set. This cyclic method was known as wrapper approach. UCI machine learning repository provided 5 set of data and it was checked by the proposed GA and SVM hybrid approach. After that the data was combined with some of the established classifier in the data mining community and showed that the collected result of hybrid approach provided a high average classification. Also the consistency of the GA-SVM hybrid was clearly seen from the histogram analysis and box plots. The hybrid approach included the utilization of a correlation measure to improve the average fitness of a chromosome population and the substitution of weaker chromosomes based on the correlation measure improved the ability of hybrid classification. The analysis demonstrated GA-SVM hybrid as a good classifier when the irrelevant attributes were removed. The GA-SVM hybrid approach attained an average accuracy of 76.20% which was relatively high. The robustness of the GA-SVM hybrid in the multi-class domain was showed by the obtained average accuracy 84.07%.

Jesmin Nahar and Tasadduq Imam *et al,* [5] have proposed a computer intelligent based approach for the diagnosis of heart diseases. Apriori, Predictive Apriori and Tertius were the three different rule mining algorithms used to present rule extraction experiment on heart disease data and showed as efficiency algorithm for diagnosis task. Cleveland dataset, a publicly available dataset and widely popular with data mining researchers, have been used for diagnosis because of the privacy problem related to medical data set. Generally diagnosis were costly, time consuming and likely to suffer from error. The analyzed information available on sick and healthy individuals indicted that females have less chance of coronary heart disease than males. Heart disease for both men and women was existed only in the presence of exercise-induced angina and factors such as chest pain were asymptotic. The resting ECG for men and women was different. The risk factors of resting ECG for women were being either normal or hyper and slope being flat. And only a single rule expressing resting ECG being hyper was an important factor for men. Slope being up, number of colored vessels being zero and old peak being less than or equal to 0.56 indicated after comparing the healthy status of men and women.

Kemal Polat and Salih Gunes, [7] have proposed a feature selection method called Kernel F-score Feature Selection (KFFS) which is used as pre-processing step in the classification of medical datasets. The proposed KFFS method has two phases. In first phase by means of Linear (Lin) or Radial Basis Function (RBF) kernel functions, the features of medical datasets have been transformed to kernel space. Using F-score formula, the F-score values of medical datasets with high dimensional feature space have been calculated. The cause of using kernel functions transformed from non-linearly separable medical dataset to a linearly separable feature space. To test the performance of KFFS method the UCI (University California, Irvine) machine learning

database used were heart disease dataset, SPECT (Single Photon Emission Computed Tomography) images dataset and Escherichia coli Promoter Gene Sequence dataset. The area under ROC curve values (AUC) values obtained from just Least Square Support Vector Machine (LS-SVM) and Artificial Neural Network (LANN) classifiers without KFFS method on the classification of heart disease, SPECT images dataset and E. coli Promoter Gene Sequence dataset were 0.796–0.708, 0.557–0.596 and 0.656–0.679. The AUC values obtained from LS-SVM and ANN classifiers with KFFS (RBF kernel) on the classification of heart disease, SPECT images dataset, and E. coli Promoter Gene Sequence dataset were 0.831–0.765, 0.75–0.634 and 0.647–0.730. They compared and found the best expert system based on the classification used in medical data set.

Pasi Luukka and Jouni Lampinen, [9] have applied classification method based on preprocessing the data first with Principal Component Analysis (PCA) and then applying differential evolution classifier to the diagnosis of heart disease. This method was applied here for predictind diagnosis from clinical data sets with chief complaint of chest pain using classical Electronic Medical Record (EMR), heart data sets which contains demographic properties, clinical symptoms, clinical findings, laboratory test results specific Electro-CardioGraphy (ECG), results pertaining to angina and coronary infarction. Individually they computed results for four different heart data sets and also the results for the case when all data sets were combined together. It was to demonstrate and assess the proposed classification approach. They were considered that the main factor resulting in the good classification accuracy in studied cases was the application of an effective global optimizer, differential evolution, for fitting the classification model instead of local optimization based approaches. With diagnosis of heart disease they found the data preprocessing with PCA. Here higher classification accuracy can be achieved than without preprocessing. The result indicated that prepossessing the data before classification might not only help with the curse of increasing data dimensionality, but also provide a further improvement in classification accuracy. They managed to classify the Switzerland data set with 94.5%±0.4% mean accuracy. They were combined the data sets and achieved the mean accuracy of 82%±0.5%. Here, the classification accuracy yielded by the proposed approach was outperformed when compared with the other corresponding results of several classifiers.

Resul Das and Ibrahim Turkoglu, *et al,* [10] have proposed several tools and various methodologies to develop effective medical decision support system. Diagnosing of the heart disease was one of the important issue and many researchers investigated to develop intelligent medical decision support systems to improve the ability of the physicians. A method was introduced which uses Statically Analysis System (SAS) base software 9.1.3 for diagnosing of the heart disease. In this method neural networks ensemble model was used which enabled an increase in generalization performance by combining several individual neural networks train on the same task. SAS base software 9.1.3 supported all tasks in a within a single, integrated solution while providing the flexibility for efficient collaborations. For heart disease diagnosis the experimental result obtained 89.01% classification accuracy, 80.95% sensitivity and 95.91% specificity values for heart disease diagnosis.

Hongmei Yan and Jun Zheng, *et al,* [15] have proposed a real-coded GA based system to select the critical clinical features essential to the heart diseases diagnosis. It has been proposed to select the critical features and assist the diagnosis of five major heart diseases which were hypertension, coronary heart disease, rheumatic valvular heart disease, chronic pulmonale and congenital heart disease. They used heart disease data with 352 cases and for each case 40 diagnostic features were recorded. Among the 352 cases of heart disease data sets 24 critical diagnostic features have been identified and their corresponding diagnosis weights for supporting or denying the diagnosis of each heart disease have been determined. It provided high accuracy in heart disease diagnosis.

Akin Ozcift and Arif Gulten, [16] have constructed a Rotation Forest (RF) ensemble classifier. It was constructed to improve the accuracy of machine learning algorithm. It was absolutely necessary in designing of high performance computer aided diagnosis system. Here rotation forest (RF) ensemble classifiers of 30 machine learning algorithms to evaluate their classification performances using Parkinson's, diabetes and heart diseases data sets. Using correlation based feature selection algorithm three data sets were reduced and then performances of 30 machine learning algorithms were calculated for three data sets and constructed based on RF algorithm. The performance of respective classifier was accessed with the same disease data and 60 algorithms were evaluated using three metrics. Average accuracy for diabetes, heart and Parkinson's data sets were 72.15%, 77.52% and 84.43%. And for the proposed RF ensemble classifiers have accuracy of 74.47%, 80.49% and 87.13%. The accuracy of RF was improved and used for designing advance systems.

Chih-Lin Chi and W. Nick Street, *et al,* [17] have proposed Optimal Decision Path Finder (ODPF) which was machine learning based expert system. It was used because of informative in terms of diagnostic accuracy in case of minimizing the time and money spent on diagnostic testing. Two tests were considered here. In first the immediate result was obtained in blood pressure test and the second was more costly and the test result was delayed. The proposed method ODPF focused on second test because of the time delayed result in test. It takes pre-test probability, interaction of variables and the cost of each test into account to generate an individualized test sequence to avoid delay test result in costly test. Lazy-learning classifiers, confident

diagnosis and Locally Sequential Feature Selection (LSFS) were the methods used to identify the sequence of diagnostic test. Among this LSFS provide cost saving accuracy of heart disease, thyroid disease, diabetes and hepatitis datasets and test saving accuracy by combining four different data sets. After comparing the results test saving provide more information and accuracy on patients available information than cost saving.

Yoon-Joo Park and Se-Hak Chun, *et al*, [18] have suggested Cost-Sensitive Case-Based Reasoning (CSCBR), a new knowledge extraction technique. It included unequal misclassification cost into conventional case based reasoning. To classify the absence and presence of disease genetic algorithm was used. An effort was taken to minimize misclassification error costs into CBR by the best classification of boundary point and number of neighbor. A fixed number of nearest neighbors in CBR was overcome by CSCBR. The absence and presence of disease was classified by adjusting the optimal cut-off classification point and cut-off distance point for selecting best neighbors. The CSCBR technique was applied in five medical data sets and then compared the result with C5.0 and CART. The total misclassification cost of CSCBR was lower than other cost-sensitive methods and was originally designed to classify binary case.

Jesmin Nahar and Tasadduq Imam, *et al*, [21] have examined the fact of computational intelligent techniques in heart disease diagnosis. Because early detection of heart disease was essential to save lives. Cleveland data was used to perform comparison with six well known classifiers. The potential of medical knowledge-driven feature selection was showed by comparing with computational intelligent based technique. And the imbalance data issue created by publicly available cleaved data was identified. For most classifiers and majority data set the performance was improved by the use of Motivated Feature Selection (MFS). It was because of the conversion of Cleveland data set for binary classification. The experimental results demonstrated that the use of MFS noticeably improved the performance, especially in terms of accuracy, for most of the classifiers considered and for majority of the datasets. MFS with Computer Feature Selection (CFS) was a promising technique used in heart disease diagnosis.

Laercio Brito Gonçalves and Marley Maria Bernardes Rebuzzi Vellasco, *et al*, [23] have determined that the Inverted Hierarchical Neuro-Fuzzy Binary Space Partitioning ($HNFB^{-1}$) was based on the Hierarchical Neuro-Fuzzy Binary Space Partitioning Model (HNFB) which gave an idea that recursive partitioning of the input space. It was able to generate its own structure automatically and allowed a greater number of inputs. The classification task of $HNFB^{-1}$ has been evaluated with different benchmark databases such as heart disease data sets. They introduced an Inverted Hierarchical Neuro-Fuzzy BSP System. It was a neuro-fuzzy model which has been specifically created for record classification and rule extraction in databases. It allowed the extraction of knowledge in the form of interpretable fuzzy rules. Fuzzy accuracy and Fuzzy coverage were the two fuzzy evaluation measures defined for the process of rule extraction in the $HNFB^{-1}$ model. The $HNFB^{-1}$ model had showed better classification performance when compared with several other pattern classification models and algorithms and the processing time converged by $HNFB^{-1}$ was very less.

Kemal Polat and Salih Gunes, [24] have presented a hybrid approach based on feature selection, fuzzy weighted preprocessing and Artificial Immune Recognition System (AIRS) to medical decision support systems. The hybrid approaches based on feature selection have two stages. The dimensions of heart disease and hepatitis disease datasets were reduced to 9 from 13 and 19 in the feature selection (FS) sub-program by means of C4. 5 decision tree algorithm. The second stage was heart disease and hepatitis disease datasets were normalized in the range of [0, 1] and were weighted via fuzzy weighted pre-processing. The obtained classification accuracies of system were 92.59% and 81.82% using 50–50% training-test split for heart disease and hepatitis disease datasets. AIRS have showed an effective performance on several problems such as machine learning benchmark problems and medical classification problems like breast cancer, diabetes and liver disorders classification. They have used the heart disease and hepatitis disease datasets taken from UCI machine learning database as medical dataset.

Kemal Polat and Seral S ahan *et al*, [25] have applied k-nearest neighbour (k-nn) weighting preprocessing and fuzzy resource allocation mechanism with AIRS on the task of diagnosis of heart disease. Here, diagnosis of heart disease was conducted with a machine learning system. In this system, a new weighting scheme based on k-nearest neighbour (k-nn) method was utilized as a preprocessing step before the main classifier. It was evident that the usages of machine learning methods in disease diagnosis have been increased gradually. While conducting this study, they first applied the k-nn based weighting process to the dataset and weighted it in the interval [0, 1]. The results strongly suggested that k-nn weighted preprocessing and fuzzy resource allocation mechanism with AIRS can aid in the diagnosis of cardiac arrhythmias. 87% of classification accuracy was obtained by their system.

Humar Kahramanli and Novruz Allahverdi, [26] have developed a hybrid neural network which included Artificial Neural Network (ANN) and Fuzzy Neural Network (FNN). The proposed method accuracy, sensitivity and specificity measures were evaluated which were used commonly in medical classification. The aim of classification was to increase the reliability of the results obtained from the data. Here a new method was presented for classification of data of a medical database. The proposed algorithm achieved the highest accuracy

rate when comparing the records in the UCI web site and related previous studies for diabetes dataset. The proposed method achieved accuracy values of 84.24% and 86.8% for Pima Indians diabetes dataset and Cleveland heart disease dataset respectively. The classification accuracies obtained by the proposed hybrid neural network were one of the best results compared with the results reported in the literature.

P.K. Anooj, [29] presented a weighted fuzzy rule-based Clinical Decision Support System (CDSS) for computer-aided diagnosis of the heart disease. The proposed CDSS for risk prediction of the heart patients contains two steps such as: generation of weighted fuzzy rules and developing of a fuzzy rule-based decision support system. Here, data preprocessing was applied on the heart disease data set for removing the noisy information and to find missing values. After that using the frequent attribute categories, the deviation range and relevant attributes were computed in this method. According to the deviation range, the attributes were selected if any deviation exists or not and also the deviation range was used to construct the decision rules. Those decision rules were scanned in the learning database to find out its frequency. As per its frequency the weight age was calculated for every decision rule obtained and by the help of fuzzy membership function, the weighted fuzzy rules were obtained. The automatic procedure to generate the fuzzy rules was an advantage of the proposed system and the weighted procedure introduced in the proposed work was additional advantage for effective learning of the fuzzy system. These weighted fuzzy rules were used to build the CDSS using Mamdani fuzzy inference system.

Nazri Mohd Nawi and Rozaida Ghazali *et al,* [31] have proposed a novel method to improve the efficiency of back propagation neural network algorithms. In Gradient Descent with Momentum and Adaptive Gain (GDM/AG) proposed algorithm, for each node the gain value was changed adaptively to modify the initial search direction. The modification enhanced the computational efficiency of training process and can be implemented in optimization process. The convergence speed of the proposed algorithm was evaluated using classification matrix. The heart disease of the patient was predicted efficiently. The algorithms were strongly constructed and have the ability to enhance the computational efficiency.

Evanthia E. Tripoliti *et al,* [44] have proposed a dynamic determination of the number of trees in random forests algorithm, a computerized diagnosis of diseases based on sorting. They have addressed the dynamic purpose of the optimum number of fundamental classifiers making up the random forests. Their proposed technique is different from most of the techniques presented in the literature. They dogged the number of classifiers during the growing procedure of the forest. Their proposed technique produces an ensemble not only accurate but also diverse ensures the two essential properties that distinguish an ensemble classifier. Their technique is derived from online fitting procedure and it is calculated using eight biomedical datasets and five versions of random forest algorithm.

Zhihua Cui *et al,* [46] have proposed a training artificial neural network by exploiting Artificial Photosynthesis and Phototropism Mechanism (APPM). They used a stochastic optimization algorithm that stirs the plant growing process. In their algorithm each entity is called as branch and the sampled points are contemplated as branch growing trajectory. They have applied the APPM algorithm to instruct the connection weights for artificial neural network. They have used two real world issues which are Cleveland heart disease categorization issue and sunspot number foreseeing issue to evaluate the performance of their APPM trained ANN. Their outcome showed that their technique increased the performance significantly contrast to other sophisticated machine learning techniques.

P.K. Anooj, [47] has proposed a weighted fuzzy rule-based CDSS for the diagnosis of heart disease. It automatically obtains the knowledge from the patient's clinical data. The proposed CDSS for risk prediction of heart patients consists of two phases such as automated approach for generation of weighted fuzzy rules and decision tree rules and the second is, developing a fuzzy rule-based decision support system. An example of a medical domain application was a detection system for heart disease based on computer-aided diagnosis methods, where the data was obtained from some other sources and was evaluated by computer based applications. Up to now, computers have usually been used to build knowledge based clinical decision support systems which used the knowledge from medical experts and transferring this knowledge into computer algorithms was done manually. The performance of the proposed CDSS improved the risk prediction when compared with the neural network-based clinical support system.

III. SURVEY OF VARIOUS CATEGORIES OF HEART DISEASE

In this section, the articles are reviewed based on different categories of heart disease such as coronary heart disease, coronary artery disease, heart failure, ischemic heart disease, cardiovascular disease, congenital heart disease, valvular heart disease and hypoplastic left heart syndrome.

A. Coronary Heart Disease (CHD)

In this section, 9 research articles related to coronary heart disease have been reviewed.

Mu-Jung Huang and Mu-Yen Chen *et al,* [3] have proposed a method by combining data mining and CBR to prognosis and diagnosis of chronic diseases. The implicit meaningful rules from health examination data was discovered by the process of adopting data mining techniques. The prognosis of chronic disease was identified by the extracted rules. Then diagnosis and treatment was supported by employing CBR. For the convenience of chronic diseases knowledge creating, organizing, refining and sharing the process was expanded to work within the system. After prognosis the suffering probability of new case chronic diseases was discovered by rules basically and then it trigger CBR. The mechanism of CBR was to retrieve most similar case from the case library. MJ health screening center collected health examination data and implemented through the system for prognosis and diagnosis of heart diseases and it was helpful reference for doctors and patients in chronic disease treatments.

N.A. Setiawan, *et al,* [8] have applied the three imputation methods namely Artificial Neural Network with Rough Set Theory (ANNRST), k-Nearest Neighbor (k-NN) and Concept Most Common Attribute Value Filling (CMCF) to University California Irvine (UCI) coronary heart disease data sets. The effect of missing attribute was investigated by comparing the three imputation methods with coronary heart disease data sets of University California Irvine. The rules were generated from the three data sets using the method called Rough Set Theory (RST). While filtering the generated rules the most complete data set of UCI coronary heart disease data is used as test data. Support filtering was applied on three sets of generated rules. In the case of UCI coronary heart disease data sets ANNRST could be considered as the best method.

Vahid Khatibi and Gholam Ali Montazer, [27] have proposed a hybrid engine to determine risk assessment problem in coronary heart disease which was designed using evidence and fuzzy set theories. Here unclear character and uncertainty modeling problems were interfered in two phase. The problems data were represented with fuzzy sets and fuzzy rules in first phase and completed by fuzzy interference rules. The systems problem opinions were calculated the belief and seeming reasonable function from previous stage result which provides belief interval as final output. It provides precise information on unclear character and uncertainty problems in CHD diagnosis risk assessment. Through information fusion it provided more accuracy result of 91.58% for CHD prediction in risk assessment.

S. Muthukaruppan and M.J. Er, [28] have proposed a system called Particle Swarm Optimization (PSO) for diagnosis of CHD. The proposed system uses Cleveland and Hungarian Heart Disease datasets. It includes four stages such as imputation of missing data, decision tree induction and rule extraction from imputed data set, using fuzzy membership functions, the crisp rules were transformed in to fuzzy rules and finally fuzzy membership functions were tuned by PSO. The generated Fuzzy Expert System (FES) based rules provides interpretation for the diagnosis of coronary heart disease. The approach has the ability to interpret the decisions made from the created FES. It provided 93.27% classification accuracy when compared to other approach.

R. Pfister *et al,* [36] have proposed an individual and cumulative effect of type 2 diabetes genetic vulnerability variants on risk of CHD. A primary risk factor for coronary heart disease is type 2 diabetes. They assumed that diabetes genetic vulnerability variants might be enhanced with CHD risk. They have analyzed the individual and cumulative effect of 38 widespread genetic variants reported earlier to be allied with type 2 diabetes on risk of incident CHD in 20,467 participants of the European Prospective Investigation into Cancer and Nutrition (EPIC). Their outcome showed that 2,190 of 20,467 entrants had an incident CHD event during a mean record of 10.7 years.

Giorgio Barbareschi *et al,* [39] have proposed socioeconomic status and the course of quality of life in matured patients with coronary heart disease. They have used two hundred and two coronary heart disease patients based on community based survey. Data on patients' quality of life were gathered prior to the diagnosis. Their outcome showed that high socioeconomic status patients notified better end result at the premorbid assessment with less depressive feelings and improved physical functioning. Eventually, they concluded that coronary heart disease modulates premorbid divergence in depressive feelings.

Keyue Ding *et al,* [40] have studied that 1243 participants had no known cardiovascular disease. Based on the Framingham risk score (FRS) they considered that cardiovascular disease to be at high, intermediate or low 10-year risk of CHD which includes age, sex, total and HDL cholesterol, blood pressure, diabetes and smoking status. A genetic risk score based on measured genotypes at 11 susceptibility Single Nucleotide Polymorphisms (SNPs), led to significant reclassification in the 10-year CHD risk categories. They calculated a multiplex genetic risk score for each patient based on the odds ratios of the susceptibility SNPs and incorporated this into the FRS. They investigated whether measured and imputed genotypes from a Genome-Wide Association (GWAS) dataset linked to the electronic medical record alter estimates of CHD risk. After incorporating the genetic risk score into the FRS, a total of 380 individuals (30.6%) were reclassified into higher (188) or lower risk groups (192). The mean standard deviation of the weighted genetic risk score was 12. 64 whose range is 5.75-18.20.

Shou-En Lu *et al,* [42] have proposed an estimation of risk equations for forecast of short term coronary heart disease events in patients with long standing type 2 diabetes. They computed the U.K. Prospective Diabetes Study (UKPDS) and Framingham risk equations for discriminating power and calibration. Their study showed that UKPDS and Framingham CHD risk equations may have restricted utility to guess CHD risk for adults with long-standing type 2 diabetes in U.S population. They highlighted the need for new refined CHD risks equations to reassess the CHD event risk and understand factors that affect CHD event risk in adults with common diabetes in modern U.S. cohort.

V. Sree Hari Rao *et al,* [45] have proposed a technique for predicting risk factors of atherosclerosis. They have considered the clinical observations and practice of individuals for forecasting the risk factors of CHD. Their technique used an in-built imputation algorithm and PSO. They have compared the performance of their technique with other machine learning techniques using STULONG dataset that was derived from longitudinal of middle aged individuals enduring for 20 years. Their technique functioned by PSO search has discovered the physical inactivity as a risk factor for the onset of atherosclerosis besides to other known factors. The decision rules derived by their technique predicted the risk factor with higher accuracy compared to the other techniques.

B. Coronary Artery Disease (CAD)

In this section, 5 research articles related to coronary artery disease have been reviewed.

Imran Kurt and Mevlut Ture, *et al,* [12] have compared performances of Logistic Regression (LR), Classification and Regression Tree (CART), Multi-Layer Perceptron (MLP), Radial Basis Function (RBF) and Self-Organizing Feature Maps (SOFM). Here the performances of classification techniques were compared in order to predict the presence of CAD. It was compared using ROC curve, Hierarchical Cluster Analysis (HCA) and Multi-Dimensional Scaling (MDS). The area under the ROC curve where 0.783, 0.753, 0.745, 0.721 and 0.675 respectively for MLP, LR, CART and RBF. To predict the presence or absence of a characteristic based on values of set of independent variables which are continuous, categorical, or both. The LR was useful and an analysis was performed in 1245 subjects which includes the presence and absence of CAD. According to HCA and MDS the performance of MLP, CART, LR and RBF were better than SOFM in predicting coronary artery disease.

Ismail Babaoglu and Omer Kaan Baykan *et al,* [13] have said that based on Exercise Stress Testing (EST), the coronary artery disease existence and localization of lesion was determined using Artificial Neural Networks (ANN). The exercise testing data was performed on 330 patients. Then 10-fold cross-validation methods were involved to select training and test data and multi neural network was employed for classifications. The diagnostic accuracy for the Left Main Coronary (LMCA) left anterior descending and left circumflex coronary arteries by the interpretation of EST using ANN were 91%, 73% and 65%. And 69% for the right coronary artery was also predicted. The elimination of LMCA lesions was by 94% negative predictive value. The diagnosing of coronary artery disease was obtained by the knowledge through the assessment of EST through ANN.

Debabrata Pal and K.M. Mandana, *et al,* [19] have proposed a method to detect coronary artery disease at early stage by designing expert system. Because CAD affects millions of people and early detection of this disease is still a challenge for prevention. The knowledge acquisition and knowledge representation techniques were the two methods used to avoid uncertainty present in medical domain. It was prevented by creating rules from the doctors and fuzzy expert system. The implementation of the system was done using object oriented analysis and design. The rules provided by medical expert predicted the patient's risk status of CAD. Organization of rules were focused using the concept of modules, meta-rule base, rule address storage in tree representation and rule consistency checking for efficient search of large number of rules focused on the organization of created rules. In CAD risk computation it leads to 95.85% sensitivity and 83.33% specificity.

Ismail Babaoglu and Ogˇuz Findik *et al,* [20] have compared Binary Particle Swarm Optimization (BPSO) and GA techniques as feature selection model to determine coronary artery disease based on EST data. In patient EST and coronary angiography were performed and 23 features were obtained from a dataset. In both BPSO and GA the Support vector machine with k-fold cross-validation method was used in CAD existence. The result of the feature selection method using BPSO and GA was compared each other and also compared the result with simple SVM model. The compared result showed that BPSO was better than GA on coronary artery disease determination in patient. The data sets composed by BPSO were succeeded in coronary artery disease existence because of more little complexity of classifier system and time compared with SVM features.

Markos G. Tsipouras and Themis P. Exarchos, *et al* [30] have proposed a novel Decision Support System (DSS) for the diagnosis of CAD. The CAD was the leading cause of death in western countries. The DSS was automatically generated using a data driven four stages innovative methodology .The methodology starts from the initial annotated dataset generating a DSS that was normally a fuzzy model with its parameters optimized subject to a specific dataset. The data set were demographic and history data along with some basic laboratory examinations and indices of arterial stiffness, thus being very easily obtained. In this method the

results of crisp rule based classifier were significantly improved when it is transformed to a fuzzy model and its parameters were optimized. This paper produced the fuzzy modeling which was able to deal with the fuzziness, which was inherent in biomedical problems. And the performance showed better results than the ANFIS. So, the methodology provided CAD diagnosis based on easily and noninvasively acquired features.

C. Heart Failure

In this section, 3 research articles related to Heart failure have been reviewed.

Peter C. Austin and Jack V. Tu, *et al*, [6] have described a heart disease prediction method by using data mining and machine learning approach. Patient with or without a specific disease was classified using classification trees. Because of limited accuracy in classification trees, data mining and machine learning have developed an alternative classification which comprise of boosted trees, bagged trees and random forests. The performance was compare with classification of different types of heart failures such as preserved ejection fraction and reduced ejection fraction. Reduced ejection fraction was based on a large number of clinical trials and preserved ejection was much smaller and the overall prognosis was similar. The two different types of heart failures were diagnosed from the results of echocardiography which was commonly done in heart patient. In high resource region and treatment decision need before echocardiogram data were available and the test was not always performed. Based on this comparison substantial improvement in prediction and classification of heart failure was provided by flexible tree based method of data mining when compared with conventional classification and regression trees. And it provided high quality performance over conventional classification.

Chang-Sik Son and Yoon-Nyun Kim, *et al*, [22] have proposed a decision making model which provides critical factors and knowledge associated with Congestive Heart Failure (CHF). The accurate diagnosis of heart disease characteristics was quite difficult in emergency room patient. The accurate diagnosis of heart disease made use of Rough Sets (RS) and decision trees. RS-based model and Logistic Regression (LR) were two subset necessary factors to differentiate CHF patients with risk factor were founded among 72 laboratories. 10-fold cross-validation was conducted by RS and LR-based decision model and showed the usefulness of proposed system. The result of RS-based model was consistently better than LR-based model after the comparison of accuracy, sensitivity, specificity, positive predictive value and negative predictive value in both models.

Daisy JA Janseen *et al*, [41] have recommended a research protocol for self perceived indications and care needs of patients with severe to very severe Chronic Obstructive Pulmonary Disease (COPD), Congestive Heart Failure (CHF) or Chronic Renal Failure (CRF) and its consequences for their closest relatives. They have configured a cross sectional comparative and prospective longitudinal study in patients with end stage COPD, CHF or CRF. The obtainable study enhances the knowledge about the indications, care needs, care-giver burden, end-of-life care treatment preferences and communication needs from the views of patients, their families and their treating physician. This knowledge is indispensable to optimize painkilling care for patients with COPD, CHF or CRF. They described the study protocol and a preliminary examination of the possible strengths, weakness and clinical consequences is outlined.

D. Ischemic Heart Disease (IHD)

In this section, 2 research articles related to Ischemic heart disease have been reviewed.

K.Rajeswari and V.Vaithiyanathan, *et al*, [14] have said about the feature selection in data mining to reduce the number of inputs under evaluation. The proposed method was to select a system with efficiency in terms of improved cost, time and accuracy. It used an Artificial Neural Network to select important features from the input layer of the network. And to select important features from an Ischemic Heart Disease (IHD) data base with 712 patients, the Multi Layer Perception Neural Network was used. The initial number of attribute and feature selection of the attributes were 17 and 12. The reduced 12 feature attributes were attempted as input for neural network. The predicted accuracy in training was 89.4% high and in testing 82.2% high. And reduced 12 were considered as interesting feature selection for Ischemic Heart Disease data set.

R. Goekmen Turan, *et al*, [37] have suggested an enhanced functional activity of bone marrow derived circulating progenitor cells after intra coronary freshly secluded bone marrow cells transplantation in patients with IHD. There has a growing testimony that the postinfarction remodeling was influenced by intracoronary autologous Bone Marrow Cells Transplantation (BMCs-Tx) in patients with chronic myocardial infarction. They have examined the influence of intracoronary autologous recently secluded bone marrow cells transplantation from the randomized controlled study by use of point of care system on cardiac function and on the functional activity of Bone Marrow-Circulating Progenitor Cells (BM-CPCs) in patients with IHD. They have taken 56 patients with IHD that are randomized to either received recently secluded BMC-Tx or a control group that did not receive cell therapy. Their outcome showed that the intracoronary transplantation of autologous recently secluded BMCs led to a significant reduction of infarct size and an augment of global ejection fraction as well as infracts wall movement velocity after 3 and 12 months record compared to control group. Finally they concluded that intracoronary transplantation of autologous recently secluded BMCs by use of point of care

system may lead to enhancement of BM-CPCs functional activity in peripheral blood which might enhance the regenerative effectiveness in patients with IHD.

E. Cardiovascular Disease

In this section, 4 research articles related to cardiovascular disease have been reviewed.

Jae-Hong Eom *et al,* [32] have introduced a system named AptaCDSS-E, a classifier ensemble-based CDSS for cardiovascular disease level prediction that can be used to identify the cardiovascular disease. The cardiovascular disease is one of the dangerous diseases that can cause death. We can save the lives of the people who were affected by the cardiovascular disease by a meticulous diagnosis technique for cardiovascular disease. To identify the cardiovascular disease, their system integrated four different classifiers with ensembles. They used support vector machines and neural networks as base classifiers and they used decision trees and bayesian networks to augment the system. They used four aptamer based biochip datasets including cardiovascular disease data that contains sixty six samples to train and test the system. They have also used three extra supplementary datasets to ease the data insufficiency. They investigated the effectiveness of the ensemble based system with diverse aggregation techniques by contrasting the result with single classifier based techniques. Their experimental outcome showed high diagnosis accuracy that proved its usefulness in the clinical decision process of disease identification.

S. Paredes *et al,* [33] have addressed two primary disadvantages of the current cardiovascular risk score system which are minimized number of risk factors considered by each individual tool and the inability of these tools to manage incomplete information. They have used two phase strategy to attain this motto. They have used a common representation procedure in the first phase which was based on the technique of Naïve-Bayes classifier and applied to a set of current risk assessment tools. Using frequency estimation technique, the classifiers' individual parameters and conditional probabilities were computed. They have proposed a combination scheme in the second phase using specific features of Bayes probabilistic reasoning and conditional probabilities optimization derived from genetic algorithm technique.

Swati Shilaskar *et al,* [34] have proposed a technique to predict the presence of cardiovascular disease accurately with reduced number of attributes. They investigated the intelligent system to construct feature subset with enhancement in diagnostic performance. They proposed a hybrid forward selection model to diagnose cardiovascular disease. Their experiment demonstrated that their technique found smaller subsets and enhanced the accuracy of diagnosis contrast to forward inclusion and back elimination models.

Li-Na Pu *et al,* [35] have suggested an investigation on cardiovascular risk prediction by exploiting genetic information. Cardiovascular disease has become a major destroyer and it is expected to cause more death in future. More prediction techniques have been developed by many groups for asymptomatic cardiovascular disease by categorizing its risk derived from established risk factors. An eligible genome-wide association studies for cardiovascular disease outcomes is overviewed by Li-Na Pu *et al.* From overall aspects, clinical trials on cardiovascular disease estimate using genetic data were summarized.

F. Congenital Heart Diasease

In this section, 2 research articles related to Congenital Heart Diasease have been reviewed.

Petra A. Karsdorp *et al,* [38] have suggested false heart rate feedback and perception of heart indications in patients with congenital heart disease and anxiety. A few things were known about the mechanisms that details an enhanced perception of heart indications in congenital heart disease. In the obtainable study it was recommended that an assortment of high trait anxiety and disease history enhances the perception of heart symptoms. They have tested whether fake heart cues will result in an enhanced perception of heart indications in patients with congenital heart disease and anxiety. They have performed two exercise tasks using 36 patients with congenital heart disease and 44 healthy controls. Their outcome suggests that congenital heart disease in assortment with high trait anxiety results in a vulnerability to over perceive heart symptoms.

Lucile Houyel *et al,* [43] have proposed a population based computation of a suggested anatomic and clinical sorting of congenital heart defects derived from the international paediatric and congenital cardiac code. They coded each individual deformity using six digit codes from the long list of IPCCC. They regrouped all lesions into 10 sorts and 23 sub-sorts derived from multi dimensional technique encompassing anatomic, diagnostic and therapeutic criteria. Their classification is configured to use the code numbers of the long list of IPCCC but can lodge ICD-10 codes. The exhaustiveness, effortlessness and anatomic basis make their recommended classification valuable for clinical and epidemiologic studies.

G. Valvular Heart Disease

In this section, 1 research article related to Valvular Heart Disease has been reviewed.

Resul Das and Ibrahim Turkoglu, *et al,* [11] have developed a method to improve the ability of physician in diagnosis of valvular heart disease. A method was introduced which uses SAS base software 9.1.3 for diagnosing of the valvular heart disease. In the centre of the proposed system a neural network ensemble

method was present which creates new models by combining the posterior probabilities or the predicted values from multiple predecessor models. The experiment was made on data set containing 215 samples and obtained 97.45% of classification accuracy. And also 100% and 96% sensitivity and specificity values were obtained in valvular heart disease diagnosis.

H. Hypoplastic Left Heart Syndrome

In this section, 1 research article related to Hypoplastic Left Heart Syndrome has been reviewed.

Andrew Kusiak and Christopher A. Caldarone *et al,* [2] have used a data mining approach with Combined Classification Quality (CCQ) measure to efficiently analyze the data. It provides a set of rules that were easily interpretable and highly accurate. The rules predicted the health status of a new born child, interventions and other user defined outcomes. The system designed to collect data on numerous patients with 30 second intervals and wellness score was computed for each data record. It leads to high accuracy in predicating the "wellness score" which indicates patient health. The prediction accuracy was important because it improves the overall understanding of postoperative management. It was developed for the assessment of data utility. A data acquisition system was developed for collection of 73 physiologic, laboratories and nurse-assessed parameters. This measure assessed the impact of a feature on classification accuracy without performing computationally expensive cross-validation. CCQ have provided a low computational complexity and proved as highly effective method to improve the care of neonates with hypoplastic left syndrome. It provided high classification accuracy.

IV. FUTURE RESEARCH DIRECTIONS

After analyzing the literature survey, the following are some of the issues identified that can be taken further to do the research. Even though various techniques have been used in the literature, still there is a need of good techniques to solve the following research issues.

- Data cleaning is an important problem for heart disease database since some of the attributes values cannot be obtained usually. So handling of missing values for diagnosis problem is a challenging task.
- The dimensionality of the heart database is high generally, so identification of significant attributes for better diagnosis of heart disease is very challenging task.
- Complexity of handling the large dataset for diagnosis is recent issue since most of classification algorithms are not suitable to handle it.
- Selection of most suitable sample of data for classification instead of the whole data is another risk for getting better diagnosis.
- Weighting the attributes is difficult task but it shows the significant research direction to obtain best diagnosis report.
- Selecting the suitable classification techniques without much computation complexity is another positive direction but the effectiveness should not be affected.
- Improvement on effectiveness is important research direction because the heart disease database is very sensitive so that much consideration and attention is need in accuracy of diagnosis.
- Handling of multiple class labels for prediction can be another positive direction of research since it can affect the performance of the medical diagnosis significantly.

V. CONCLUSION

A detailed survey of various papers related to heart diasease diagnosis published in the standard journals of IEEE, Elsevier, springer and inderscience is presented. Here, 47 articles were identified from 2005 to 2013 related to heart disease diagnosis and various categories of heart diasease. From the review, the identification was that neural network based techniques contribute more effectiveness. Also from the accuracy perspectives, seven techniques provide more than 90% accuracy as compared with the other techniques presented in the literature. Finally, some of the research issue is also addressed to precede the further research in the same direction.

REFERENCES

[1] Carlos Ordonez," Association Rule Discover with the Train and Test Approach for the Heart Disease Prediction", IEEE Transactions on Information Technology in Biomedicine, Vol. 10, No. 2, PP. 334-343, April 2006

[2] Andrew Kusiak and Christopher A. Caldarone,*et al* "Hypo plastic left heart syndrome knowledge discovery with a data mining approach" Journal of Computers in Biology and Medicine, Volume 36,Issue 1,PP 21-40, January 2006

[3] Mu-Jung Huang and Mu-Yen Chen *et al,* "Integrating data mining with case-based reasoning for chronic diseases prognosis and diagnosis" Journal of Expert Systems with Applications, Vol. 32, PP.856–867, 2007

[4] K.C. Tan and E.J. Teoh *et al,* "A hybrid evolutionary algorithm for attribute selection in data mining", Journal of Expert system with applications, Vol.36, PP.8616-8630, 2009

[5] Jesmin Nahar and Tasadduq Imam *et al,*" Association rule mining to detect factors which contribute to heart disease in males and females", Journal of Expert Systems with Applications Vol.40, PP.1086–1093, 2013

[6] Peter C. Austin and Jack V. Tu, *et al,*" Using methods from the data-mining and machine-learning literature for disease classification and prediction: a case study examining classification of heart failure subtypes" Journal of Clinical Epidemiology, Vol.66, PP. 398-407, 2013

[7] Kemal Polat and Salih Gunes," A new feature selection method on classification of medical datasets: Kernel F-score feature selection", Journal of Expert Systems with Applications, Vol. 36, PP.10367–10373, 2009

[8] N.A. Setiawan, *et al*," A Comparative Study of Imputation Methods to Predict Missing Attribute Values in Coronary Heart Disease Data Set**",** Journal in Department of Electrical and Electronic Engineering,Vol.21, PP. 266–269, 2008

[9] Pasi Luukka and Jouni Lampinen," A Classification Method Based on Principal Component Analysis and Differential Evolution Algorithm Applied for Prediction Diagnosis from Clinical EMR Heart Data Sets" , Journal of Computer Intelligence in Optimization Adaption, Learning and Optimization, Volume 7, pp 263-283, 2010

[10] Resul Das and Ibrahim Turkoglu, *et al,* "Effective diagnosis of heart disease through neural networks ensembles", Journal of expert system with applications, Vol.36, PP. 7675–7680, 2009

[11] Resul Das and Ibrahim Turkoglu, *et al,* "Diagnosis of valvular heart disease through neural networks ensembles", Journal of Computer Methods and Programs in Biomedicine, Vol.93, PP. 185-191, 2009

[12] Imran Kurt and Mevlut Ture, *et al, "*Comparing performances of logistic regression, classification and regression tree, and neural networks for predicting coronary artery disease", journal of Expert Systems with Applications, Vol. 34, PP. 366–374, 2008

[13] Ismail Babaoglu and Omer Kaan Baykan *et al,* "Assessment of exercise stress testing with artificial neural network in determining coronary artery disease and predicting lesion localization", Journal of Expert System With Applications, Vol.36, PP. 2562–2566, 2009

[14] K.Rajeswari and V.Vaithiyanathan, *et al,* "Feature Selection in Ischemic Heart Disease Identification using Feed Forward Neural Networks ", International Symposium on Robotics and Intelligent Sensors, Vol.41, PP. 1818–1823, 2012.

[15] Hongmei Yan and Jun Zheng, *et al,* "Selecting critical clinical features for heart diseases diagnosis with a real-coded genetic algorithm", Journal of Applied Soft Computing, Vol.8, PP.1105-1111, 2008

[16] Akin Ozcift and Arif Gulten, "Classifier ensemble construction with rotation forest to improve medical diagnosis performance of machine learning algorithms", Journal of Computer Methods and Programs in Biomedicine, Vol.104, PP.443-451, 2011.

[17] Chih-Lin Chi and W. Nick Street, *et al,* "A decision support system for cost-effective diagnosis", Journal of Artificial Intelligence in Medicine, Vol.50, PP. 149-161, 2010.

[18] Yoon-Joo Park and Se-Hak Chun, *et al,* "Cost-sensitive case-based reasoning using a genetic algorithm: Application to medical diagnosis", Journal of Artificial Intelligence in Medicine, Vol.51, PP.133-145, 2011

[19] Debabrata Pal and K.M. Mandana, *et al,* "Fuzzy expert system approach for coronary artery disease screening using clinical parameters", journal of knowledge based system, Vol.36, PP.162-174, 2012

[20] Ismail Babaoglu and Og˘uz Findik *et al,* "A comparison of feature selection models utilizing binary particle swarm optimization and genetic algorithm in determining coronary artery disease using support vector machine", Journal of Expert System With Applications, Vol.37, PP.3177-3183, 2010

[21] Jesmin Nahar and Tasadduq Imam, *et al,* "Computational intelligence for heart disease diagnosis: A medical knowledge driven approach", Journal of Expert System with Application, Vol.40, PP.96-104, 2013

[22] Chang-Sik Son and Yoon-Nyun Kim, *et al,* "Decision-making model for early diagnosis of congestive heart failure using rough set and decision tree approaches", Journal of Biomedical Informatics, Vol.45, PP. 999–1008,2012

[23] Laercio Brito Gonçalves and Marley Maria Bernardes Rebuzzi Vellasco,*et al,*" Inverted Hierarchical Neuro-Fuzzy BSP System: A Novel Neuro-Fuzzy Model for Pattern Classification and Rule Extraction in Databases, " Journal of IEEE Transactions On Systems, Man, And Cybernetics, Vol. 36, No. 2, March 2006

[24] Kemal Polat and Salih Gu nes, "A hybrid approach to medical decision support systems: Combining feature selection, fuzzy weighted pre-processing and AIRS", Journal of Computer Methods and Programs in Biomedicine, Vol.88, PP.164-174, 2007

[25] Kemal Polat and Seral S ahan *et al,* "Automatic detection of heart disease using an artificial immune recognition system (AIRS) with fuzzy resource allocation mechanism and k-nn (nearest neighbor) based weighting preprocessing", Journal of expert system with applications, Vol.32, PP.625-631, 2007

[26] Humar Kahramanli and Novruz Allahverdi, "Design of a hybrid system for the diabetes and heart diseases", Journal of Expert Systems with Applications, Vol. 35, PP. 82–89, 2008

[27] Vahid Khatibi and Gholam Ali Montazer, "A fuzzy-evidential hybrid inference engine for coronary heart disease risk assessment", Journal of Expert Systems with Applications, Vol. 37, PP. 8536–8542, 2010

[28] S. Muthukaruppan and M.J. Er, "A hybrid particle swarm optimization based fuzzy expert system for the diagnosis of coronary artery disease", Journal of Expert Systems with Applications, Vol.39, PP. 11657–11665, 2012

[29] P.K. Anooj, "Clinical decision support system: Risk level prediction of heart disease using weighted fuzzy rules", Journal of Computer and Information Sciences, Vol.24, PP. 27–40, 2012

[30] Markos G. Tsipouras and Themis P. Exarchos, *et al,* "Automated Diagnosis of Coronary Artery Disease Based on Data Mining and Fuzzy Modeling", Journal of IEEE Transactions On Information Technology In Biomedicine, Vol. 12, No. 4, July 2008

[31] Nazri Mohd Nawi and Rozaida Ghazali *et al,* "The Development of Improved Back-Propagation Neural Networks Algorithm for Predicting Patients with Heart Disease", In proceedings of the first international conference ICICA, Vol.6377, PP.317-324,2010

[32] Jae-Hong Eom and Sung-Chun Kim, *et al,* "AptaCDSS-E: A classifier ensemble-based clinical decision support system for cardiovascular disease level prediction", Journal of Expert Systems with Applications, Vol. 34 2465, PP.2479, 2008

[33] S. Paredesa *et al,* "Long term cardiovascular risk models' combination", Journal of computer methods and programs in biomedicine, Vol.101, PP 231-242, 2011

[34] Swati Shilaskar *et al,* Feature selection for medical diagnosis: Evaluation for cardiovascular diseases", Journal of Expert System with Application, Vol.40, PP.4146-4153, 2013

[35] Li-Na Pu and Ze Zhao, *et al,* "Investigation on Cardiovascular Risk Prediction Using Genetic Information", Journal of IEEE Transactions On Information Technology In Biomedicine, Vol. 16, No. 5, September 2012

[36] R. Pfister and D. Barnes *et al,* "Individual and cumulative effect of type 2 diabetes genetic susceptibility variants on risk of coronary heart disease", Journal of Diabetologia, Vol.54, PP.2283-2287, 2011

[37] R. Goekmen Turan and I. Bozdag *et al,* "Improved Functional Activity of Bone Marrow Derived Circulating Progenitor Cells After Intra Coronary Freshly Isolated Bone Marrow Cells Transplantation in Patients with Ischemic Heart Disease", Journal of stem cell review and report, Vol.7, PP.646-656, 2011

[38] Petra A. Karsdorp and Merel Kindt *et al,* "False Heart Rate Feedback and the Perception of Heart Symptoms in Patients with Congenital Heart Disease and Anxiety", International Journal of behavioral Medicine, Vol.16, PP.81-88, 2009

[39] Giorgio Barbareschi and Robbert Sanderman *et al,* "Socioeconomic Status and the Course of Quality of Life in Older Patients with Coronary Heart Disease", International Journal of behavioral Medicine, Vol.16, PP.197-204, 2009

[40] Keyue Ding and Kent R Bailey *et al, "* Genotype-informed estimation of risk of coronary heart disease based on genome-wide association data linked to the electronic medical record", International journal of BMC cardiovascular Disorders, Vol.11,2011
[41] Daisy JA Janssen and Emiel FM Wouters *et al,* "Self-perceived symptoms and care needs of patients with severe to very severe chronic obstructive pulmonary disease, congestive heart failure or chronic renal failure and its consequences for their closest relatives: the research protocol", Journal of BMC palliative care, Vol.7, 2008
[42] Shou-En Lu and Gloria L Beckles *et al*, "Evaluation of risk equations for prediction of short-term coronary heart disease events in patients with long-standing type 2 diabetes: the Translating Research into Action for Diabetes", International Journal of BMC Endocrine Disorders, Vol.12, 2012.
[43] Lucile Houyel and Babak Khoshnood *et al*, "Population-based evaluation of a suggested anatomic and clinical classification of congenital heart defects based on the International Paediatric and Congenital Cardiac Code", International Journal of rare diseases, Vol.6, 2011.
[44] Evanthia E. Tripoliti and Dimitrios I. Fotiadis *et al,* "Automated Diagnosis of Diseases Based on Classification: Dynamic Determination of the Number of Trees in Random Forests Algorithm", Journal of IEEE Transactions On Information Technology In Biomedicine, Vol. 16, No. 4, July 2012
[45] V. Sree Hari Rao and M. Naresh Kumar, "Novel Approaches for Predicting Risk Factors of Atherosclerosis" Journal of IEEE Journal Of Biomedical And Health Informatics, Vol. 17, No. 1, January 2013.
[46] Zhihua Cui *et al,* "Training artificial neural networks using APPM", International Journal of wireless and mobile computing, Vol.5, PP.168-174, 2012
[47] P.K. Anooj, "Clinical decision support system: Risk level prediction of heart disease using weighted fuzzy rules and decision tree rules", Journal of Computer Sciences, Vol.24, PP. 27–40, 2012.