International Journal of Applied Engineering Research ISSN 0973-4562 Volume 13, Number 21 (2018) pp. 15356-15361

Cardio Vascular Disease Diagnosis Using Data Mining Techniques and ANFIS Approach

B.Shuriya[a], A.Rajendran[b]

[a]*Department of Computer science and Engineering, Adithya Institute Technology, Coimbatore -641107, Tamilnadu, India.*

[b]*Department of Electronics and Communication Engineering, Karpagam College of Engineering , Coimbatore, Tamilnadu, India*.

Abstract

The data mining techniques plays a vital role in health care diagnosis system and decision making. Cardio vascular disease (CVD) prediction and the complexities are measured based on their risk level. The Association rule is framed using the fuzzy logic for predicting the disease and the complexities are measured by implementing the Adaptive Neuro fuzzy Inference System (ANFIS). The effects of different values for important parameters were investigated to obtain the highest performance. The accuracy of the cardio vascular disease is measured as 93.12%.

Keywords: CVD, ANFIS, Association rule, Accuracy, Cardio vascular.

INTRODUCTION

The aim of this paper is to study the different of Cardio vascular disease and to predict the disease using data mining category techniques. The Heart disease encompasses the diverse diseases that affect the heart. CVD is the major cause of casualties in the world. The cardiovascular disease includes a wide range of conditions that affect the heart and the blood vessels in which blood is pumped and circulated through the body. Some of burden of CVD can potentially be relieved by primary prevention. Detecting and treating those are at highest CVD risk level is regarded as an essential complement to a population-based approach [1].The rapid urbanization and industrialization have led to the major cause of human demise in India due to CVD[2]. The classification of data is obtained from the record of heart disease data set to increase the reliability [3]. In data mining, neural networks are competitive in classification due to their dimensionality and the low computational cost by using a large number of dimensionality and the huge volume of data [4]. The quality implies the treatment of the patient correctly [5].Discovering heart disease from several factors is a multilayered problem which may cause unpredictable effects. Healthcare industry today creates large amounts of complex data about patients, hospitals resources, disease diagnosis, patient records, medical devices etc. The huge amount of data is the key resource to be processed and analyzed for knowledge extraction that enables support for cost-savings and decision making. The back propagation technique is used in developing multilayer neural networks in a supervised manner. In the application related to the strategic management of business processes, the discipline of the data mining, which can be defined as the process of extracting useful and meaningful patterns from large volumes of data, emerges as a viable method [7]. The association rules have two parts antecedent and consequent which help uncover relationship between likely unrelated data in the relational database[8]. Data mining technique [31] is helpful for the decision making and diagnosis in the medical field such as text mining, classifications, etc.,. The remaining of this paper organized as follows: section2: Related work, Section 3: Neuro fuzzy classification, Section 4: Experimental results and section5: Conclusion

RELATED WORK

In this section we discuss the related work that used in the heart disease. The classifiers such as Rule set classifiers, decision trees, ANN, neuro fuzzy, Bayesian Network structure discovery is used in medical data set [9]. The decision tree classifiers are more sensitive. The ANN, genetic algorithm, association rule discovery techniques are implemented to diagnose the heart disease [10].Heart disease prediction system is implemented using artificial neural network technique by training data .The accuracy of the classification is near 80% , sensitivity is 85% [11]. The cross industry standard process for data mining (CRISP-DM) technique is implemented for classification process. The Naive Bayes, weighted association classifier, a priori algorithm are used for the decision making in heart disease dataset [12].Decision tree, navie bayes, ANN [13,14,15,17,21,33] technique is used to predict the heart disease based on classification accuracy. Associative classification, principle component analysis, lazy associative classification method is used in CVD[16]. Decision making in the heart disease is very important to predict the risk factor and precautions to be taken for the patient. Recent study of Registrar General of India (RGI) and Indian Council of Medical Research (ICMR) 25% of the deaths are between the age group 25-65 years because of heart disease. The CART, ID3, decision tree method are used for the heart disease diagnosis [18]. Neural networks, fuzzy logic [32] and genetic algorithm, supervised machine learning, genetic algorithm, IHDPS is used in predicting the heart disease [19]. Data analysis and encoding, neural network weight optimization by genetic algorithm, neural networks techniques are implemented in Data analysis and encoding, neural network weight optimization by genetic algorithm, neural networks are implemented in predicting the CVD [20].

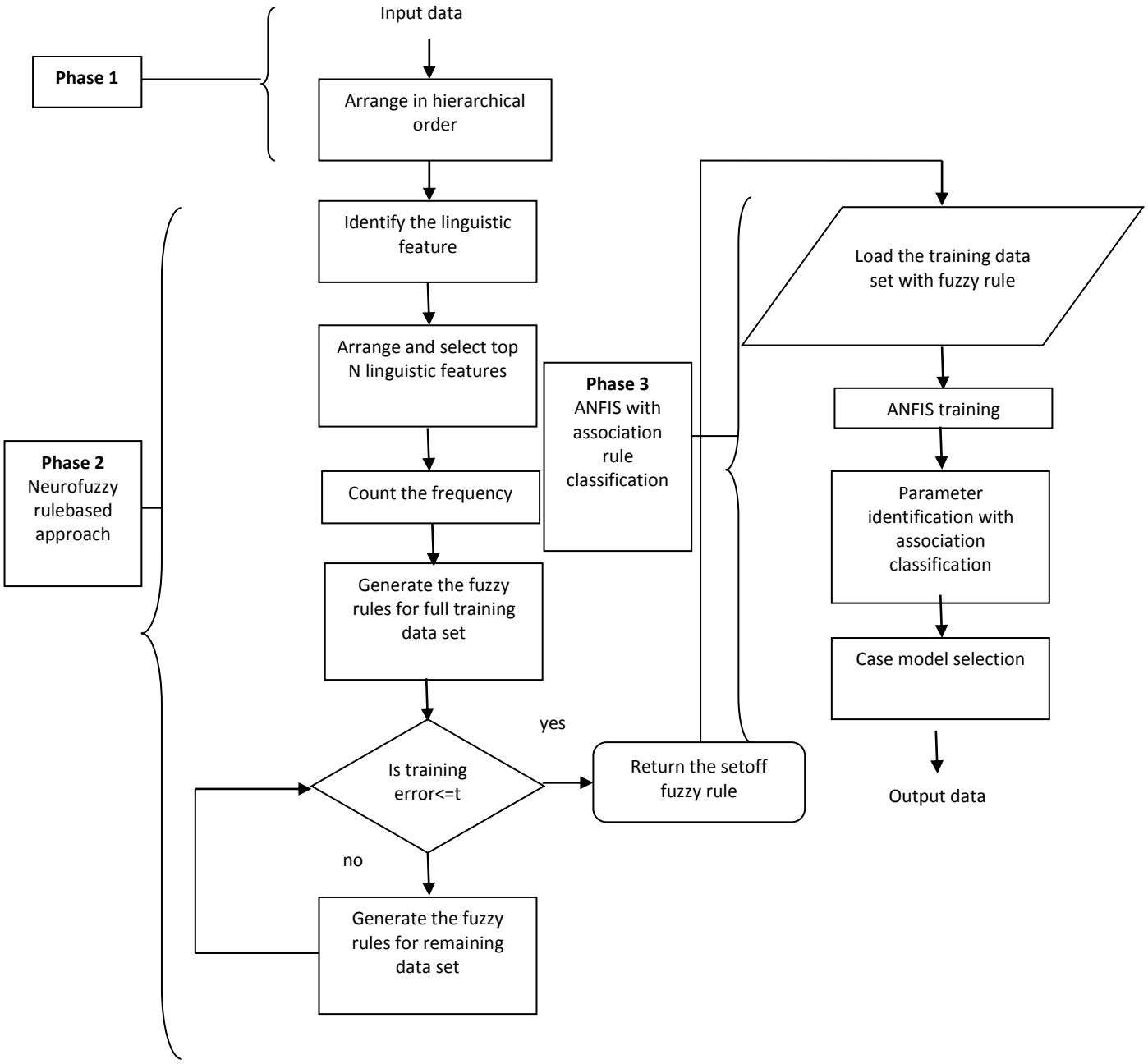

Figure 1. Proposed method

NEURO FUZZY CLASSIFICATION

The proposed method based on ANFIS and rule extraction algorithm for diagnosing the cardio vascular disease. It consists of three phases. In first phase the data set is organized in hierarchical structure to have the relationship between the attributes. In second phase based on the linguistic features the neuro fuzzy rule based extraction is performed to remove the redundancy of the attributes. In the third phase the ANFIS approach with the association rule classifies the data. The flow chart of the proposed method is shown in figure 1.

Neural Network Architecture

Neural Network is a computational model. The functions and methods of the neural network are based on the structure of the brain. There are different types of architecture such as feed forward network, Radial basis function (RBF) Network, self organizing network, recurrent neural network, Modular neural network and so on.

The neuro fuzzy is the fuzzy system that is derived from the neural network theory which is trained by learning algorithm. The neuro fuzzy system is the 3 layered feed forward network. The first layer is the input variable, second is the hidden layer and the third is the output layer. Fuzzy sets are encoded as connection weights.

Rule Extraction Algorithm

Neuro fuzzy systems can always be interpreted as a system of fuzzy rules[22].There is no difference between the neuro fuzzy model [23] and proposed neuro fuzzy model except rule extraction.

Proposed rule extraction algorithm is given below:

Step 1: Arrange the linguistic features in descending order with their weights.

Step 2: From that order select top N linguistic features for each class until each linguistic feature gets selected at least once.

Step 3: Count the frequency of linguistic features of each input feature.

Step 4: Keep the lower linguistic features separately and count the frequency in input feature.

Step 5: The input feature which has higher frequency is considered as more important than others.

(If two input features are having same frequency then their top most linguistic feature weights determine the important one)

Step 6: Layered OR rules for classification are generated based on important input features, order of classes and top linguistic features in the following way

For all input features from most important to least important

If f1 is lf1_1 then Class1_1

Else if f1 is lf1_2 then Class1_2 …….. Else if f1 is lf1_N then Class1_N

…. ….

Else if f2 is lf2_1 then Class2_1 …. …. Else if f2 is lf2_N then Class2_N

…. ….

Else if fn is lfn_1 then Classn_1 …. …. Else if fn is lfn_N then Classn_N

where fn represents nth input feature, lfn_N represents the Nth important linguistic feature of fn feature and Classn_N represents the class of Nth important linguistic feature of fn feature

Step 7: For all the input features

If all the rules of an input feature are found redundant in classification, the input feature is removed else step 10

Step 8: Layered OR rules for classification are generated order of classes and lowest linguistic features.

Step 9: if redundant, then input feature is removed else step 10

Step 10: Rules of the remaining input features are the final rules.

ANFIS Architecture

The fuzzy modeling and membership function reduces the human error to the greater extent [24,25].The ANFIS Architecture for two input first order fuzzy model with two rules is shown in figure 2.

Layer 1: Fuzzification layer

Assume two inputs X and Y and one output Z.

Rule 1: if X is A1 and Y is B1, then f1= p1X+q1Y+r1

Rule 2: if X is A2 and Y is B2, then f2= p2X+q2Y+r2

Here every node i in this layer is an adaptive node with a node function,

$O_{i,1}=\mu_{Ai}(x)$ for i=1,2 … (1)

$O_{i,1}=\mu_{Bi-2}(x)$ for i=3, 4 … (2)

Layer 2: Rule antecedent layer

This layer includes the nodes which represents the antecedent part of association rule. The output of each node is given as

$O_i=\mu_{Ai}(x)$, $\mu_{Bi}(x)$ for i=1,2 … (3)

Layer 3: Strengthen normalization layer

The weight obtained from the layer 2 is normalized by fixed nodes present in this layer. The output of the normalization layer is given as,

$O_i= W_{-1}=\frac{wi}{w1+w2}$ for i=1,2 … (4)

Layer 4: Consequent layer

The adaptive nodes included in this layer are the products of normalization strength in polynomial order.

$O_i= W_{-1}f_i= W_{-1}(r_i(x) +s_i(x)+t_i)$ for i=1,2 …(5)

Layer 5: Inference Layer

The overall output is given as,

$O =\sum_{i=1}^{i=2} 1\ W_{-1}f_i=\frac{\sum_{i=1}^{i=2} Wifi}{\sum_{i=1}^{i=2} Wi}$ for i=1,2 … (6)

In layer 1 the eight modifiable parameters c_i and σ_i where i=1,2,3,4. In layer 4 six modifiable parameters r_i,s_i, and t_i for i=1,2.

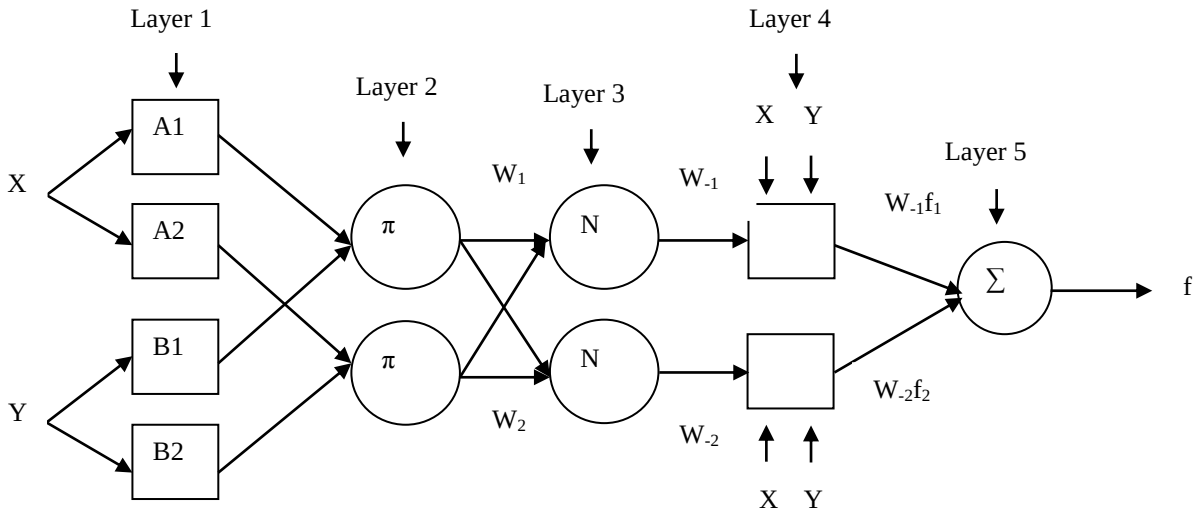

Figure 2. ANFIS Architecture

EXPERIMENTAL RESULTS AND DISCUSSION:

The Cleveland data set for heart disease is used for the evaluation. It consists of 303 instances with 14 medical attributes such as age: age in years ;sex: sex (1 = male; 0 = female) ; cp: chest pain type , Value 1: typical angina , Value 2: atypical angina ,Value 3: non-anginal pain ,Value 4: asymptomatic ;trestbps: resting blood pressure (in mm Hg on admission to the hospital) ;chol: serum cholestoral in mg/dl ;fbs: (fasting blood sugar > 120 mg/dl) (1 = true; 0 = false) ; restecg: resting electrocardiographic results , Value 0: normal ,Value 1: having ST-T wave abnormality (T wave inversions and/or ST elevation or depression of > 0.05 mV) ,Value 2: showing probable or definite left ventricular hypertrophy by Estes' criteria ; thalach: maximum heart rate achieved ;exang: exercise induced angina (1 = yes; 0 = no) ; oldpeak = ST depression induced by exercise relative to rest ;slope: the slope of the peak exercise ST segment , Value 1: upsloping ,Value 2: flat, Value 3: downsloping ;ca: number of major vessels (0-3) colored by flourosopy ;thal: 3 = normal; 6 = fixed defect; 7 = reversable defect ;num: diagnosis of heart disease (angiographic disease status) ,Value 0: < 50% diameter narrowing ,Value 1: > 50% diameter narrowing .The programming language used for diagnosis heart disease is MATLAB which is an efficient tool for data analysis and visualization.

Two parameters are considered in training ANFIS system such as maximum number of epochs and number of hidden neurons. The number of epochs varies from 1000 to 5000 and hidden neuron ranges from 7 to 21. The results are shown in table 1. The system also has an additional module known as case-based module, where the user has to input values for 14 required attributes as specified by the Cleveland data set, in order to test the status of the patient whether heart disease is present or absent from that particular patient.

However the combination of 4000 epochs and 15 hidden neurons with an increment of 7 achieved the highest accuracy results as 93.12%. The system performs 5000 epochs with 21 hidden layers achieves 84.47%.

Table 1. ANN results

No.of epochs	No. of hidden neurons	Training data	Testing data	Case based classification
1000	7	77.11	76.54	Absent
1000	14	83.71	78.12	Absent
1000	21	84.45	87.56	Absent
2000	7	86.23	77.12	Absent
2000	14	85.89	86.54	Absent
2000	21	87.54	88.78	Absent
3000	7	84.52	81.64	Absent
3000	14	87.54	83.12	Absent
3000	21	87.47	88.81	Absent
4000	7	86.54	79.37	Absent
4000	14	84.32	**93.12**	Absent
4000	21	90.89	85.21	Absent
5000	7	87.12	86.54	Absent
5000	14	86.14	84.33	Absent
5000	21	83.58	84.47	Absent
Average results		85.5366	84.08	100% Absent

So the system performs with large combinations. The system classifies the data that are "Absent" of heart disease, therefore the system is achieved 100% accuracy. In most systems, the testing data is very important and systems are evaluated on how best they can perform when receiving data from of someone who has not been trained earlier. Therefore, if considering this aspect, the combination of 5000 epochs and 15 hidden neurons is selected since it performs the highest using the testing data set.

Table 2. Accuracy

Authors	**Method**	**Accuracy**	**Year**
Das et al[29]	Neural network ensemble	89.01%	2009
Anuj et al [30]	Weighted fuzzy rules	57.85%	2012
Samuel et al[26]	ANN and Fuzzy AHP	90.1%	2017
Bala et al[28]	Hadoop	86.2%	2016
Proposed method	ANFIS+ASSOCIATION RULE	**93.12%**	-

In this model Receiver Operating Characteristic Curve is under the threshold 0.8 and error rate is 0.05. The sensitivity is measured. The decision support system [26] ANN and fuzzy AHP achieved the accuracy of 90.1% by two steps. At first, the problem is represented in hierarchical structure. Second, decomposed into goal, attributes and alternatives. Then the fuzzy triangular membership function prior to the pair wise comparison is formed. The ANN and ANFIS approach of decision making [27] achieved the accuracy of 90.74% by implementing three modules such as training, testing and case based module.

CONCLUSION

In this paper the ANFIS with association rule approach in order to develop a Cardio Vascular disease diagnosis system. The table 2 clearly shows that the 93.2% accuracy is achieved using the training dataset. The ANFIS with association rule yields better results than the other method such as neural network ensemble, weighted fuzzy rules , ANN with Fuzzy AHP, Hadoop in cloud computing. This system can be used at hospital level by doctors and physicians to classify the patient's heart disease. Despite the promising results recorded by the proposed method, the optimal sets of hidden layer nodes and their respective weighted links still are a challenge. And as a result, the ANN classifiers in the study were trained and re-trained until the performance of the networks were relatively stable. The future work can be through applying Various ANFIS architecture with KNN.

REFERENCES

[1] Jackson R, Lynch J, Harper S.(2006) Preventing coronary heart disease. *British Medical Journal*.

[2] World Health Organization. http://www.who.int/whr/2006/whr06_en.pdf. *The world health report 2006–working together for health. Geneva*.

[3] Kahramanli H, Allahverdi N.(2008) Design of a hybrid system for the diabetes and heart diseases. *Expert Systems with Applications.*

[4] Wang S.(2005) Classification with incomplete survey data: a Hopfield neural network approach. *Computers & operations research*.

[5] Mann DL, Zipes DP, Libby P, Bonow RO.(2014) *Braunwald's heart disease: a textbook of cardiovascular medicine*. Elsevier Health Sciences.

[6] Principe JC, Euliano NR, Lefebvre WC. (1999)*Neural and adaptive systems: fundamentals through simulations with CD-ROM*.John Wiley & Sons, Inc.

[7] Onan A, Bal V, Yanar Bayam B.(2016) The Use of Data Mining for Strategic Management: A Case Study on Mining Association Rules in Student Information System. *Hrvatski časopis za odgoj i obrazovanje*.

[8] Li W, Han J, Pei J. (2001) CMAR: Accurate and efficient classification based on multiple class-association rules. In: *Data Mining.ICDM 2001. Proceedings IEEE International Conference.*

[9] Bhatla N, Jyoti K. (2012)An analysis of heart disease prediction using different data mining techniques. *International Journal of Engineering*.

[10] Soni J, Ansari U, Sharma D, Soni S. (2011)Predictive data mining for medical diagnosis: An overview of heart disease prediction. *International Journal of Computer Applications.*

[11] Chen AH, Huang SY, Hong PS, Cheng CH, Lin EJ.(2011) HDPS: Heart disease prediction system. In: *2011 Computing in CardiologyIEEE*.

[12] Sundar NA, Latha PP, Chandra MR.(2012) Performance analysis of classification data mining techniques over heart disease database. International Journal of engineering science & advanced technology.

[13] Dangare CS, Apte SS.(2012) Improved study of heart disease prediction system using data mining classification techniques. International Journal of Computer Applications.

[14] Taneja A. (2013)Heart disease prediction system using data mining techniques. *Oriental Journal of Computer science and technology*.

[15] SA S. (2013) Intelligent heart disease prediction system using data mining techniques. International Journal of Healthcare & Biomedical Research.

[16] Jabbar MA, Deekshatulu BL, Chandra P.(2013) Heart disease prediction using lazy associative classification. In: Automation, Computing, Communication, Control and Compressed Sensing (iMac4s). IEEE International Multi-Conference.

[17] Sudhakar K, Manimekalai DM.(2014) Study of heart disease prediction using data mining. International

Journal of Advanced Research in Computer Science and Software Engineering.

[18] Chaurasia V, Pal S.(2013) Early prediction of heart diseases using data mining techniques. Carib. j. SciTech, 2013.

[19] Bhatla N, Jyoti K. (2012)An analysis of heart disease prediction using different data mining techniques. International Journal of Engineering. 2012.

[20] Amin SU, Agarwal K, Beg R.(2013) Genetic neural network based data mining in prediction of heart disease using risk factors. In: Information & Communication Technologies (ICT). IEEE Conference on IEEE.

[21] Masethe HD, Masethe MA.(2014) Prediction of heart disease using classification algorithms. In Proceedings of the world congress on engineering and computer science.

[22] Nauck D, Kruse R. (1997)A neuro-fuzzy method to learn fuzzy classification rules from data. Fuzzy sets and Systems.

[23] Eiamkanitchat N, Umpon T.(2010) Colon tumor microarray classification using neural network with feature selection and rule based classification. Advances in Neural Network Research and Applications.

[24] Al-Sakran HO.(2015) Framework architecture for improving healthcare information systems using agent technology. Int. J. Managing Inf. Technol.

[25] Devi CS, Ramani GG, Pandian JA.(2014) Intelligent E-healthcare management system in medicinal science. Int. J. PharmTech Res.

[26] Samuel OW, Asogbon GM, Sangaiah AK, Fang P, Li G. (2017)An integrated decision support system based on ANN and Fuzzy_AHP for heart failure risk prediction. Expert Systems with Applications.

[27] Abushariah MA, Alqudah AA, Adwan OY, Yousef RM. (2014)Automatic Heart Disease Diagnosis System Based on Artificial Neural Network (ANN) and Adaptive Neuro-Fuzzy Inference Systems (ANFIS) Approaches. Journal of Software Engineering and Applications.

[28] Bala A, Malhotra S, Gupta N, Ahuja N.(2016) Emerging Green ICT: Heart Disease Prediction Model in Cloud Environment. In: Proceedings of International Conference on ICT for Sustainable Development. Springer Singapore.

[29] Das R, Turkoglu I, Sengur A.(2009) Effective diagnosis of heart disease through neural networks ensembles. Expert Systems with Applications.

[30] Anooj PK.(2012) Clinical decision support system: Risk level prediction of heart disease using weighted fuzzy rules.uhkJournal of Computer and Information Sciences.

[31] Su Gon Cho, Seoung Bum Kim.(2017) A Data-Driven Text Similarity Measure Based On Classification Algorithms, International Journal Of Industrial Engineering.

[32] Coskun Ozkana , Gulsen Aydin Keskinb, Sevinc Ilhan Omurcac.(2014)A Variant Perspective To Performance Appraisal System: Fuzzy C – Means Algorithm, International Journal Of Industrial Engineering.

[33] Gyu M Lee(2013) System Engineering Approach To Build An Information System For Emergency Cesarean Deliveries In Small Hospitals International Journal Of Industrial Engineering.