CDSS for Heart Disease Prediction Using Risk Factors

SEMINAR REPORT

Submitted by

HARIPRIYA A

IDK16CS029

to

the APJ Abdul Kalam Technological University
in partial fulfillment of the requirements for the award of the degree

of

BACHELOR OF TECHNOLOGY

in

Computer Science and Engineering

DEPARTMENT OF COMPUTER SCIENCE AND ENGINEERING

GOVERNMENT ENGINEERING COLLEGE IDUKKI

PAINAVU-685603

November 2019

GOVERNMENT ENGINEERING COLLEGE
PAINAVU IDUKKI-685 603

CERTIFICATE

This is to certify that the seminar report entitled **"CDSS for Heart Disease Prediction Using Risk Factors"** *has been submitted by* **HARIPRIYA A (IDK16CS029)** *in the partial fulfillment for the award of B.Tech Degree in COMPUTER SCIENCE AND ENGINEERING for the academic session 2019-2020 under the supervision and guidance of Prof.Deepa S S, Department of Computer Science and Engineering,Govt. Engineering College, Idukki.*

Seminar Guide
Prof. Deepa S S.
Associate Professor,
Department of CSE,
GEC Idukki.

Seminar Co-ordinator
Prof. Deepa S S.
Associate Professor,
Department of CSE,
GEC Idukki.

Head of the Department
Dr.Madhu K P.
Department of CSE,
GEC Idukki.

ACKNOWLEDGEMENT

I give all honour and praise to the **GOD** who gave me wisdom and enabled me to complete my seminar on " Learning Regular Sets from Queries and Counterexamples " successfully.

I express my sincere thanks to **Dr. Satheesh Kumar, Principal, Government Engineering College, Idukki,** for providing the right ambiance to work on the seminar.

I would like to extend my sincere gratitude to **Dr. Madhu K P , Head of Department, Computer Science and Engineering** for permitting me to work on the seminar and for her guidance, encouragement, support and care throughout the entire period of my course of study.

I deeply indebted to my Seminar Coordinator **Prof. Deepa S S,Associate Professor, Department of Computer Science and Engineering** for her continued support throughout our seminar.

It is with great pleasure that I express my deep sense of gratitude to my seminar guide **Dr. Prof. Deepa S S, Associate Professor, Department of Computer Science and Engineering** for her guidance, supervision, encouragement and valuable advice in each and every phase of my seminar.

I would like to thank all other faculty members and the fellow students of Government Engineering College, Idukki, for their warm friendship, support and help.

Also, I express my hearty thanks to my beloved parents for their love, encouragement and dedication in shaping my career.

HARIPRIYA A

ABSTRACT

Clinical Decision Support System(CDSS) is an efficient tool used in medical field.It helps medical practitioners to make better decisions.This work proposes new approaches for prediction of heart diseases based on some risk factors such age,ECG,diabetes,slope,hypertension,high cholesterol or physical inactivity etc.Heart patients have these kind of risk factors which can be used for the prediction of the disease.

The purpose of this work is to predict heart disease based on the above mentioned risk factors by using a new paradigm called Neuro-Fuzzy model(NFS).Neuro-Fuzzy model is combines with the adaptive capabilities of the neural network and the reasoning approach of the fuzzy logic.The proposed system will provide an intelligent system for predicting heart disease.

Contents

List of Figures

List of Tables

Chapter 1

INTRODUCTION

Heart disease is known widely all over the world and it is also considered as one among the major disease.The disease includes a variety of problems including high blood pressure,hardening of the arteries,chest pain,heart attack and so on. Diagnosis of the heart disease is a complex task which requires much experience and knowledge.all most all doctors are predicting the disease by their previous experience and learning.Prediction of disease like heart disease is a complex and tedious task.Therefore there is a chance of creating false pre-assumption and incorrect results.

There are many studies and researches have been proposed for the prediction of heart disease.Among these methods,Neuro-Fuzzy based method is more better than other conventional methods.Because it has high rate of accuracy than other methods.The proposed system provides a decision support platform with the combination of Genetic Algorithm(GA) and Neuro-Fuzzy system(NFS).It consists of Fuzzy Logic(FL) component,Neural Network(NN) and a GA. The proposed system will helpful to both medical professionals as well as the patients.The system will be done using MATLAB.

Chapter 2

RELATED WORKS

For providing clinical decision support systems,literature presents a number of researches that have made use of artificial intelligence and data mining techniques.Among those techniques,to support decision makers in the risk prediction of heart disease,a lot of researches have been presented.A few of the literature are given below.

The first concept uses neural based learning classifier for classifying data mining tasks which shows that neural based learning classifier system performs equivalently to supervise learning classifier.Here Unified Computing System (UCS) is one of the supervised learning classifier system used as a tool for classifying the data mining tasks.This paper proposes a novel way to incorporate neural networks into UCS.The approach provides compactness,expressiveness,accuracy.By using a simple artificial neural network as the classifier's action,a more compact population size, better generalization, and the same or better accuracy while maintaining a reasonable level of expressiveness can be obtained.A negative correlation learning (NCL) is also applied during the training of the resultant NN ensemble. NCL is shown to improve the generalization of the ensemble

Second method is the intelligent and effective heart attack prediction system

(IEHPS) was built using data mining and neural networks.This work provides a proficient methodology for the extraction of significant pattern from the heart disease warehouses for the heart attack prediction.Warehouse is clustered with the aid of K-means cluster algorithm and the patterns are mined with the aid of MAFIA algorithm.Patterns are selected on the basis of computed significant weightage.And the neural network is trained with these selected patterns.Multi-Layer Perceptron Neural Network with Back Propagation is used as training algorithm.

And another system is the Intelligent Heart Disease Prediction System using CANFIS and Genetic Algorithm. In this paper, a new paradigm based on coactive neuro-fuzzy inference system (CANFIS) was proposed for prediction of heart disease. CANFIS model combines the neural network adaptive capabilities and the fuzzy logic qualitative approach which is then integrated with genetic algorithm to diagnose the heart disease.The performances of the CANFIS model were evaluated in terms of training performances and classification accuracies.The results showed that the proposed CANFIS model has great potential in predicting the heart disease.

The heart disease is the leading cause of high rate of morality.Therefore decision support system is needed for effective prediction of the existence or absence of heart disease.In correct diagnosis or poor clinical decisions leads to mortality.Traditional medical domain applications predict heart disease using computer aided diagnosis methods,where the data are obtained from some other sources and are evaluated based on computer aided applications. A medical practitioner uses several sources of data and tests to make a diagnostic impression but all tests are not necessary and useful for the diagnosis of a heart disease.This process is time consuming, costly, requires lots of tests and really depends on medical expert's opinions. All medical experts are not equally good in predicting the heart disease

in which diagnosis plays a very important role in the case of heart disease.Proper diagnosis at the right time saves life of many patients. To avoid this problem, machine learning techniques have been developed to gain knowledge automatically from raw data which saves time,money and accurate prediction of the heart disease.

Chapter 3

GENETIC ALGORITHM AND THEIR INTEGRATION WITH NFS

3.1 GENETIC ALGORITHM

A genetic algorithm is a search strategy that is inspired by Charles Darwin's theory of natural evolution. This algorithm reflects the process of natural selection where the fittest individuals are selected for reproduction in order to produce offspring of the next generation.

The process of natural selection starts with the selection of fittest individuals from a population. The offspring inherits the characteristics of their parents and will be added to the next generation. Offspring with better fitness have a better chance at surviving. Genetic algorithm iterate till the end, a generation with the fittest individuals will be found.

five phases in genetic algorithm. 1.Initial population 2.Fitness function 3.Selection 4.Crossover 5.Mutation

3.1.1 Initial population

The process starts with a set of individuals which is called a Population. Each individual is considered to be a solution to the problem. An individual is characterized by a set of variables known as Genes. Genes are joined together to form a Chromosome (solution).In a genetic algorithm, the set of genes of an individual is represented using a string,known as chromosomes.Binary values are used to represent chromosome.

3.1.2 Fitness Function

The fitness function determines how fit an individual is,that is the ability of an individual to compete with other individuals.Each of the individual will be provided with a fitness score.Individuals with higher fitness value have the probable chances of selecting to produce the offspring.

3.1.3 Selection

Fittest individuals and let them pass their genes to the next generation.Parents are selected based on their fitness scores. Individuals with high fitness have more chance to be selected for reproduction processes.

3.1.4 Cross Over

Crossover is the most one of the significant phase in a genetic algorithm.A crossover point is chosen at random mannerfrom within the genes.

3.1.5 Mutation

Some of their genes can be face mutation with a low random probability. This implies that some portions of the bits in the bit string can be changed.

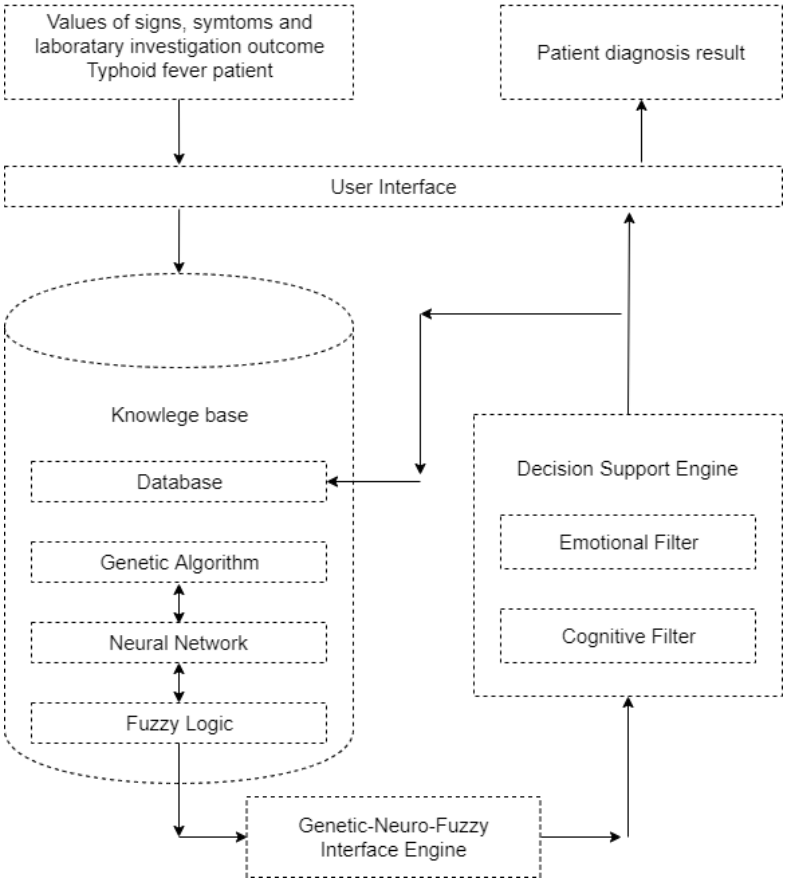

Figure 3.1: Architecutre of the proposed genetic adaptive neuro-fuzzy interference

The genetic algorithm applies the mutation, crossover and selection operations to the individuals in the population to detremine all promising regions in the solution space till it achieves the desired solution.This advantage can be applied to neural networks to calculate the weight parameters.For a genetic algorithm-based algorithm employed in fuzzy neural networks, the first network structure is usually generated or given firstly. The genetic algorithm is then used to optimize the

topology of a network in terms of membership functions and/or fuzzy rules.The GA is used to determine all the parameters of the membership functions of the fuzzy controller.

Chapter 4

PROPOSED SYSTEM

The proposed system uses a new computational paradigm called an neuro fuzzy system (NFS). NFS is based on fuzzy system which is trained by a learning algorithm derived from neural network.NFS combines the adaptive capabilities of the neural network and the logical approach of the fuzzy logic.The system use ANN's theory in order to determine their properties like fuzzy sets and fuzzy rules by processing data samples. The major objective of this system is to develop an Intelligent Heart Disease Prediction system.

The process of Neuro-Fuzzy with GA:

1. Initialize the process of predicting Cardiovascular Disease.
2. Extract the patient's details from dataset.
3. Assign the input to NFS.
4. Selection process starts by assigning weights to each attributes randomly.
5. Training the network using Back-propagation algorithm.
6. Compute output values.
7. Compute fitness using below equation

$$Mean\ Square\ Error(MSE) - \frac{\sum(Output\ Targets)^2}{Number\ of\ Samples}$$

8. If MSE is less than error then go to step 10, otherwise, go to step 9.
9. Select the parents and apply crossover and mutation.
10. Train NFS with selected connection weights.
11. Study the performance of test data

The process initializes with the prediction of Heart Diseases.Then extracts the details of the patient.The system then gives training to the Neuro-Fuzzy system. Training is generally consists of four sub processes, which are:1. giving the inputs and outputs parameters to the system.2. the selection of sample 3.training using the back- propagation algorithm 4. The output results are calculated and Mean Square Error (MSE) is computed. If the MSE is less than error then training to the Neuro-Fuzzy System is completed else the two chromosomes or samples are selected for further process. In this further process genetic algorithm is applied, in which crossover and mutation is applied to the samples. New weights are taken for the samples and again training process is carried out until the MSE is less than error.

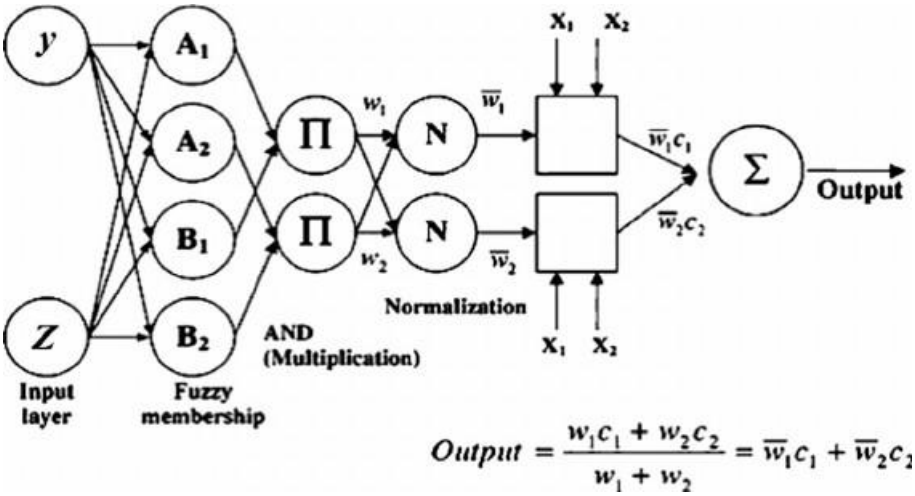

Figure 4.1: A prototype NFS network and output calculation

4.0.1 Heart Disease Data Set

Modern research in the field of medicine has been able to identify risk factors that may contribute toward the development of heart disease but more research is needed to use this knowledge in reducing the occurrence of heart diseases. Diabetes, hypertension, and high blood cholesterol have been treated as the major risk factors of heart diseases. Life style risk factors like eating habits, physical inactivity, smoking, alcohol intake, obesity are also associated with the major heart disease risk factors and heart disease. Here the dataset related to the heart disease is provided to the neuro-fuzzy system. The dataset consist of the patients symptoms of heart diseases. It consists of heart disease patients' information. The system uses Cleveland databases which is usually available.This dataset includes the information of the patients like Age in years, Sex, Blood Pressure, Blood Cholesterol, Diabetes, Electrocardiographic results, Heart rate, Physical activity, Slop of the peak, Number of major vessels colored by fluoroscopy, Thalassemia(Defect type) which is considered as a most commonly risk factors of the cardiovascular disease. There are samples for 12 attributes like sex, age, blood cholesterol, blood pressure, chest pain, electrocardiographic results, heart rate,physical activity, diabetes, diet number of vessels, Thalassemia to predict whether one can have a cardiovascular disease or not.

The below table represents some of the important risk factors and the corresponding values and their encoded values in brackets, which were used as input to the system.

	Risk factors	**Description with encoded values**
1	Age	20-34 (-2), 35-50 (-1), 51-60 (0),61-79 (1) , >79 (2)
2	Blood pressure	Below 120 mm Hg- Low (-1) 120 to 139 mm Hg- Normal (0) Above 139 mm Hg- High (-1)
3	Blood cholesterol	Below 200 mg/dL - Low (-1) 200-239 mg/dL - Normal (0) 240 mg/dL and above - High(1)
4	Diabetes	Yes (1) or No (0)
5	Physical Activity	Yes (1) or No (0)
6	Slope	The slope of the peak exercise ST segment Value 1: up sloping Value 2: flat Value 3:down sloping
7	Chest Pain	Yes (1) or No (0)
8	ECG	The slope of the peak exercise ST segment Normal(0) Abnormal(1) Hyper(2)
9	Heart Rate	Below 100- Low (-1) 100 to 150- Normal (0) otherwise- High (1)
10	No. of major vessels	Number of major vessels Colored by fluoroscopy(0-3)
11	Thal	Normal-3(0) fixed defect-6(1) reversible defect-7 (2)
12	Sex	Male(1) or Female (0)
o/p	Heart Disease	Yes (1) or No (0)

Table 4.1: Description Of Cleaveland Heart Disease Database

Chapter 5

SIMULATION RESULTS

The medical records of 100 Heart Disease patients were collected from the UCI Machine learning repository.This collected data was analyzed and pre-processed to the required format. Matrix Laboratory (MATLAB) was used to implement the proposed system. In this heart disease prediction system one can know whether the person having the disease or not with higher accuracy than other prediction systems. Accuracy is measured on the basis of Mean Square Error (MSE) means that if the mean square error is low then accuracy is high and if mean square is high then accuracy is low. Accuracy is determined using the following formula: Accuracy= (TP+TN)/ (TP+FP+TN+FN) Where, TP-True positive, TN- True Negative, FP-False positive and FN-False negative.

Technique	**Neuro-Fuzzy System (without GA)**	**Neuro-Fuzzy System (with GA)**
True Positive	38	40
True Negative	44	50
False Positive	5	0
False Negative	12	10
Accuracy	82%	90%

Table 5.1: Comparison of NFS with GA and without GA approach for heart disease prediction by using 100 records

As shown in Table, for cardiovascular disease prediction 100 records is used for testing the performance of the system on the basis of TP (True Positive), TN (True Negative), FP (False Positive) and FN (False Negative) which calculate the accuracy of the system. This that shows that the neuro-fuzzy system with genetic algorithm is applied then the result or the prediction of the cardiovascular disease is more accurate than when it is applied without genetic algorithm.

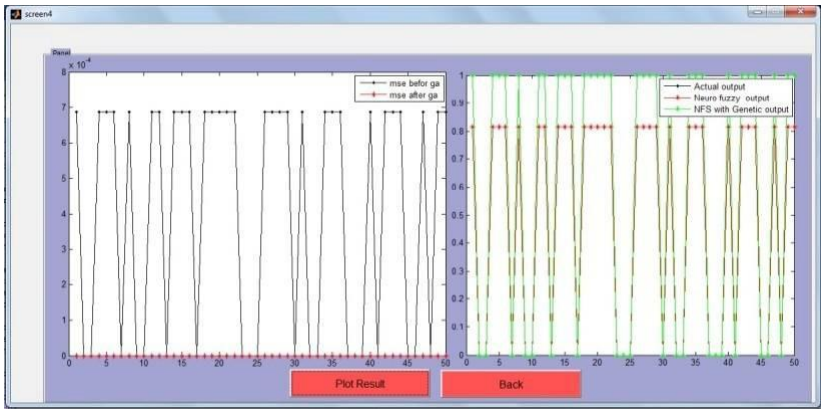

Figure 5.1: Comparison Graph

The above figure shows various comparisons. In the first part the value of Mean Square Error before application of Genetic algorithm (GA) are compared with the values of MSE after the application of genetic algorithm. From this graph, it can be easily deduce that the value of MSE after the application of GA (red) is far less than that of before application of GA (black). The substantial reduction in MSE will lead to increase the accuracy of the system and helps in realistic prediction of the heart disease.

In the second part, simply using neuro-fuzzy system (without application of GA) the assigned values of outputs (which isinput to the NFS) i.e. ‘0’, ‘1’ are not recognized perfectly (red). On the other hand the assigned values to the outputs (which is input to the NFS) like ‘0’, ‘1’ etc. are clearly recognized using

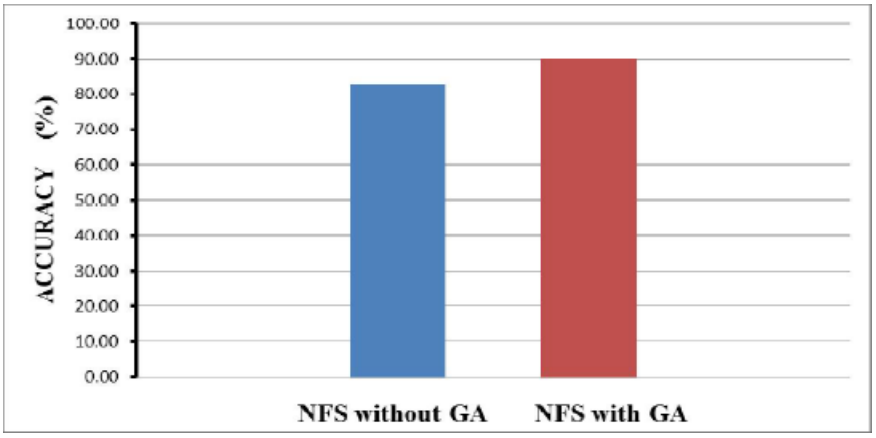

Figure 5.2: Accuracy of the NFS with and without GA

neuro-fuzzy system in connection with genetic algorithm (green). This comparison graph proves that neuro-fuzzy system with genetic algorithm can produce more realistic and accurate results.

It can be seen from Figure 4 that the proposed technique (NFS with GA) had an overall average diagnosis accuracy of 90 percentage as against that of the (NFS without GA) method which was 82 percentage and this seem promising and could help increase the overall accuracy in the CDSS of heart disease and other diseases in the world.

Chapter 6

CONCLUSION

A Multi-technic decision support system powered by genetic algorithm, neural network, and fuzzy logic concepts for the diagnosis of heart disease has been investigated in this study.An improved genetic algorithm concept was used to automatically supply the optimal set of weights needed to effectively train the neural network module. Usually, the membership function parameters of FIS are manually set thereby making it difficult for the FIS to provide accurate diagnosis results when confronted with new cases. To address this problem, the trained, validated, and tested neural network module was configured to automate the provision of membership function parameters for the fuzzy inference system, that is, building some form of learning and tuning capability into the fuzzy inference system. With this development, the fuzzy inference system was able to provide timely and reliable diagnosis outcome for new cases. The outcome of the evaluation process conducted in this research shows that the proposed system (NFS with GA) had it attained a diagnosis accuracy of 90 percentage as compared to 82 percentage of the (NFS without GA) method. Also, in terms of timetaken to diagnose a patient, the proposed system also performed better than the conventional (NFS without GA).Therefore, the proposed technique (NFS with GA) has the capability to alle-

viate the key problems associated with Neuro-Fuzzy Based diagnostic methods if fully embraced and as well it could be adopted to solve challenging problems in several other domains.

REFERENCES

[1] Koehn P. Combining Genetic Algorithms and Neural Networks: The Encoding Problem, A thesis presented for the Master of Science Degree, the University of Tennessee, Knoxville,1994

[2] Zadeh LA The calculus of fuzzy if/then rules. AI Expert 7: 27-27,1992

[3] Ebene S. M. Metev and V. Ebenezer O. O, Oyebade K. O.," Heart Diseases Diagnosis Using Neural Networks Arbitration", I.J. Intelligent Systems and Applications, vol. 12,pp. 75-82,2015

[4] Hai H.Dam, Hussain A.Abbass and Xin Yao, "Neural – Based Learning Classifier Systems", IEEE Transactions on Knowledge and Data Engineering, Vol.20, No.1, pp.26-39, 2008

[5] Shantakumar B.Patil and Y.S.Kumaraswamy, "Intelligent and Effective Heart Attack Prediction System Using Data Mining and Artificial Neural Network", European Journal of Scientific Research, Vol.31, No.4, pp.642-656, 2009

[6] Polat , K., S. Sahan, and S. Gunes, Automatic detection of heart disease using an artificial immune recognition system (AIRS) with fuzzy resource allocation mechanism and k-nn (nearest neighbour based weighting preprocessing. Expert Systems with Applications 32 p. 625– 631,20077. Latha

Parthiban and R. Subramanian, "Intelligent Heart Disease Prediction System using CANFIS and Genetic Algorithm", International Journal of Biological and Life Science, Vol. 15, pp. 157 - 160, 2007

[7] G.Camps-Valls, L.Gomez-Chova, J.Calpe-Maravilla, J.D.Martin-Guerrero, E.Soria-Olivas, L.Alonso-Chorda, J.Moreno, "Robust support vector method for hyperspectral data classification and knowledge discovery." vol.42, no.7, pp.1530–1542, July.2004

[8] N. Al-Milli, "Backpropagation Neural Network for Prediction of Heart Disease", Journal of Theoretical and Applied Information Technology, vol. 56, 2013.

[9] Sellappan Palaniappan, Rafiah Awang, Intelligent Heart Disease Prediction System Using Data Mining Technique, 978-1-4244-1968- 5/08/25.00 2008 IEEE.

[10] P.K. Anooj,Clinical decision support system: Risk level prediction of heart disease using weighted fuzzy rules", Journal of King Saud University – Computer and Information Sciences 2012

[11] "Coronary heart disease statistics fact sheets 2008/2009," British Heart Foundation, London 2008

[12] L. Zaret, M. Moser, and L. S. Cohen, Yale university school of medicine heart book. New York: Hearst Books, 1992.

[13] B. Phibbs, The human heart: a basic guide to heart disease. Philadelphia: Lippincott Williams and Wilkins, 2007.

[14] A. Selzer, Understanding heart disease. Berkeley: University of California Press, 1992

[15] O. S. Randall and D. S. Romaine, The encyclopaedia of the heart and heart disease. New York, NY: Facts on File, 2005.

[16] Wood D, De Backer, Prevention of coronary heart disease in clinical practice: recommendations of the Second Joint Task Force of European and other Societies on Coronary Prevention. Atherosclerosis 140: pp.199–270, 1998

[17] K. Priya, T. Manju and R. Chitra, "Predictive Model of Stroke Disease Using Hybrid Neuro-Genetic Approch", International Journal of Engineering and Computer Science, vol.2, no.3, pp 781-788, Mar 2013

[18] Syed Umar Amin, Kavita Agarwal and Dr. Rizwan Beg, "Genetic Neural Network Based Data Mining in Prediction of Heart Disease Using Risk Factor", Proceeding of IEEE Conference on Information and Communication Technologies(ICT), pp. 1227-1231, April 2013