

Heart Disease Prediction using Artificial Neural Networks

Akash Mukherjee¹ Raj Manjrekar² Ashish Marde³ Prof. Rajesh Gaikwad⁴

^{1,2,3}Student ⁴Assistant Professor

^{1,2,3,4}Department of Computer Engineering

^{1,2,3,4}Atharva College of Engineering, Mumbai University, Mumbai, India

Abstract— Information mining systems have been generally utilized as a part of clinical choice emotionally supportive networks for forecast and determination of different illnesses with great exactness. These systems have been extremely powerful in outlining clinical emotionally supportive networks in light of their capacity to find concealed examples and connections in therapeutic information. A standout amongst the most imperative uses of such frameworks is in analysis of heart illnesses in light of the fact that it is one of the main sources of passings everywhere throughout the world. All frameworks that foresee heart maladies use clinical dataset having parameters what's more, inputs from complex tests directed in labs. None of the framework predicts heart ailments in view of danger variables, for example, age, family history, diabetes, hypertension, elevated cholesterol, tobacco smoking, liquor admission, heftiness or physical idleness, and so on. Heart ailment patients have part of these obvious danger elements in like manner which can be utilized adequately for determination. Framework in light of such hazard variables would help medicinal experts as well as it would give patients a notice about the plausible vicinity of coronary illness even before he visits a healing center or goes for excessive restorative checkups. Thus this paper exhibits a system for expectation of coronary illness utilizing significant danger elements. This method includes two best information mining instruments, neural systems and hereditary calculations. The half breed framework executed utilizations the worldwide improvement point of preference of hereditary calculation for instatement of neural system weights. The learning is quick, more steady and exact when contrasted with back engendering. The framework was actualized in Matlab and predicts the danger of coronary illness with a precision of 89%.

Key words: Data Mining, Heart Disease Risk Factors

I. INTRODUCTION

Heart diseases are on the first number that cause of death globally: Most of the people die annually from Heart diseases than from any other cause. An estimated 17.3 million people died from Heart diseases in 2008, representing 30% of all global deaths. Of these deaths, an estimated 7.3 million were due to coronary heart disease and 6.2 million were due to stroke [1]. Recent research in the field of medicine has been able to identify risk factors that do contribute to 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 established as the major risk factors of heart diseases. Life style risk factors which include eating habits, physical inactivity, smoking, alcohol intake, obesity are also associated with the major heart disease risk factors and heart disease[2]. There are studies showing that reducing these risk factors for heart disease can actually help in preventing heart ailments . There are numerous studies and inquires about on the counteractive action of coronary illness hazard. Information from investigations of populace has helped in forecast of heart illnesses, in view of circulatory strain, smoking propensity, cholesterol and pulse levels, and diabetes. Analysts have utilized these expectation calculations in adjusted type of streamlined score sheets that permit patients to compute the danger of heart sicknesses [3]. The Framingham Risk Score (FRS) is a mainstream hazard expectation basis which is utilized as a part of calculations for coronary illness expectation [4]. This study went for adding to a wise information mining framework in light of hereditary calculation enhanced neural systems for the expectation of coronary illness in light of danger elements' classes. The framework was actualized utilizing MATLAB R2012a.

II. DIFFERENT DATA MINING TECHNIQUES

Information mining procedures are utilized to investigate, examine and remove medicinal information utilizing complex calculations as a part of request to find obscure examples. Analysts are utilizing information mining procedures for the determination of numerous infections, for example, heart ailment [5], diabetes, stroke and growth and numerous information mining procedures have been utilized as a part of the determination of coronary illness with great exactness. Scientists have been applying diverse information mining systems, for example, gullible bayes, neural system, choice tree, stowing, part thickness, and bolster vector machine for forecast and determination of heart illnesses. One of the frameworks utilizes neural based learning classifier for characterizing information mining errands appeared that neural based learning classifier framework performs proportionally to oversee learning classifier. IEHPS canny and compelling heart assault forecast framework was assembled utilizing information mining and neural systems and it proposed separating huge examples for coronary illness expectation utilizing K-implies grouping and utilized MAFIA calculation to mine the incessant examples. Polatet al., created framework utilizing crossover fluffy and k-closest neighbour approach for the expectation of coronary illness, which had 87% precision in finding [6]. In another framework [7] neural system group was utilized as a part of the determination of coronary illness with a precision of 89.01%. Latha and Subramanian (2007), proposed a keen coronary

illness expectation framework utilizing CANFIS what's more, hereditary calculation which had a low mean square mistake [8].Analyzing the distinctive systems talked about, this paper proposes a novel framework utilizing hereditary calculation and neural Data mining strategies are used to research, look at and evacuate therapeutic data using complex figuring as a piece of solicitation to find dark illustrations. Examiners are using data mining methods for the determination of various contaminations, for instance, heart affliction [5], diabetes , stroke and development and various data mining methods have been used as a piece of the determination of coronary sickness with incredible precision. Researchers have been applying differing data mining frameworks, for instance, naïve bayes, neural framework, decision tree, stowing, part thickness, and support vector machine for conjecture and determination of heart sicknesses. One of the structures uses neural based learning classifier for portraying data mining errands showed up that neural based learning classifier structure performs relatively to manage learning classifier. IEHPS watchful and convincing heart attack conjecture system was gathered using data mining and neural frameworks and it proposed isolating gigantic cases for coronary sickness desire using K-infers gathering and used MAFIA figuring to mine the relentless illustrations. Polatet al., made system using hybrid feathery and k-nearest neighbour approach for the desire of coronary ailment, which had 87% exactness in discovering [6]. In another structure [7] neural framework gathering was used as a piece of the determination of coronary ailment with an accuracy of 89.01%. Latha and Subramanian (2007), proposed a sharp coronary ailment desire structure using CANFIS besides, computation which had a low mean square misstep [8].Analyzing the unmistakable frameworks discussed, this paper proposes a novel system using genetic count and neural

III. DIFFERENT METHODS

The issue with danger variables identified with coronary illness is that there are numerous danger variables included like age, use of cigarette, blood cholesterol, individual's wellness, circulatory strain, stress and so forth and understanding and classifying every one as indicated by its significance is a troublesome errand. Likewise a heart ailment is frequently identified when a patient compasses progressed phase of the sickness [9]. Subsequently the danger elements were broke down from different sources [10]-[11]. The dataset was made out of 12 essential danger elements which were sex, age, family history circulatory strain, Smoking Habit, liquor utilization, physical dormancy, diabetes, blood cholesterol, less than stellar eating routine, corpulence .The framework demonstrated whether the patient had danger of coronary illness on the other hand not. The information for 50 individuals was gathered from reviews done by the American Heart Association [11]. A large portion of the coronary illness patients had numerous likenesses in the danger variables [12]. The TABLE I underneath demonstrates the distinguished critical danger variables and the relating values and their encoded values in sections, which were utilized as data to the framework.

	Risk Factors	Values
1	Sex	Male (1), Female (0)
2	Age (years)	20-34 (-2), 35-50 (-1), 51-60 (0), 61-79 (1) , >79 (2)
3	Blood Cholesterol	Below 200 mg/dL - Low (-1) 200-239 mg/dL - Normal (0) 240 mg/dL and above - High (1)
4	Blood Pressure	Below 120 mm Hg- Low (-1) 120 to 139 mm Hg- Normal (0) Above 139 mm Hg- High (-1)
5	Hereditary	Family Member diagnosed with HD -Yes (1) Otherwise -No (0)
6	Smoking	Yes (1) or No (0)
7	Alcohol	Yes (1) or No (0)
8	Physical Activites	Low (-1) , Normal (0) or High (-1)
9	Diabetics	Yes (1) or No (0)
10	Diet	Poor (-1), Normal (0) or Good (1)
11	Obesity	Yes (1) or No (0)
12	Stress	Yes (1) or No (0)
Output	Heart Disease	Yes (1) or No (0)

Table 1: Different Methods

Information investigation has been completed keeping in mind the end goal to change information into helpful structure, for this the qualities were encoded for the most part between an extent [-1, 1]. Information examination likewise evacuated the irregularity and peculiarities in the information. This was required. Information examination was required for right information preprocessing. The evacuation of missing and wrong inputs will help the neural system to sum up well.

This system uses backpropagation algorithm for learning and training the neural network, but there are two major disadvantages with backpropagation algorithm. First is that the initialization of the NN weights is a blind process hence it is not possible to find out globally optimized initial weights and there is a danger that the network output would run towards local optima hence the overall tendency of the network to find out a global solution is greatly affected. The second problem is that back propagation algorithm is very slow in convergence and there is a possibility that network never converges . This problem of local optimum solution can be solved by optimizing the initial weights of neural network. For this we use a

genetic algorithm which is specialized for global searching. For this we first determine the number of inputs, layers and hidden neurons of the neural network and then we would use the backpropagation algorithm to train the networks using the weights optimized by GA.

IV. NEURAL NETWORK ARCHITECTURE

A multilayered forward system is utilized having 12 data hubs 10 shrouded hubs and 2 yield hubs. The quantity of inputs depends on the last arrangement of danger components for every patient which is given in TABLE I. number of concealed hubs must be chosen for which the preparation is quick and the system gives the best yield. The initial step is to instate the weights of neural system utilizing the "design" capacity accessible in MATLAB. At that point these designed weights are gone to the hereditary calculation for streamlining as indicated by the wellness capacity. Once the weights are improved, the Levenberg-Marquardt backpropagation calculation is utilized for preparing and learning also, "trainlm" is a system preparing capacity that redesigns weight and inclination values as indicated by Levenberg-Marquardt streamlining. The "trainlm" is regularly the quickest backpropagation calculation in the tool kit, and is exceptionally prescribed as a first-decision directed calculation, despite the fact that it requires more memory than different calculations. Most extreme number of ages to prepare is set to default esteem 100. The learning stops at a predefined least mistake in the wake of changing system weights and conforming them to an ideal amount at which the characterization is exact. The anticipated yield would be vicinity or nonattendance of a coronary illness.

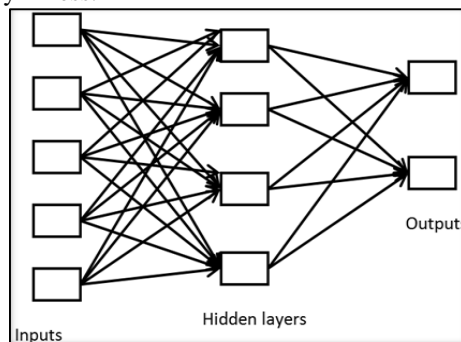


Fig. 1: ANN Architecture

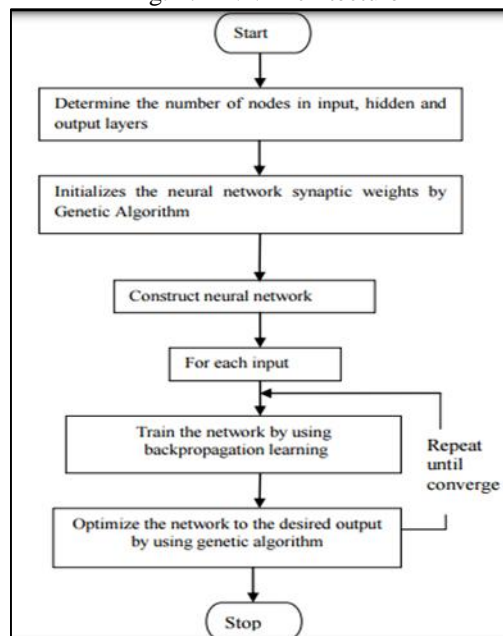


Fig. 2: System Flowchart

V. PARAMETER SETTINGS

The framework was produced utilizing MATLAB R2012a. Worldwide Streamlining Toolbox and the Neural Network Toolbox were utilized for executing the calculation. The information for danger elements identified with heart sicknesses gathered from 50 individuals is given in TABLE II. ANN is instated with the "arrange" capacity, with each weigh being between - 1.0 to +1.0. These weights are then gone to the hereditary calculation which employments the mean square mistake as the wellness capacity. The interconnecting weights and edges of the prepared neural system are gone to the hereditary calculation. The quantity of neurons in the three layer neural systems is 12, 10, and 2 separately in info, shrouded and yield layer. Thus there are $(12 \times 10 + 10) + (10 \times 2 + 2) = 152$ aggregate weights and inclinations. The weights in the ANN are encoded in such a way every weight is being between - 1.0 to +1.0. After that weights are allocated to every connection. Weights

alteration utilizing GA is finished with 'populace size =20'.In this application, every string or Chromosome in the populace speaks to the weight and inclination estimations of the system. Wellness capacity is ascertained for each chromosome taking into account mean square blunder. The wellness capacity utilized is mean square blunder (mse) which is ascertained as underneath:

$$mse = \frac{1}{n} \sum_{k=1}^m (O_k - T_k)^2$$

After choice, hybrid and transformation in GA, the chromosomes with lower adjustment are supplanted with better ones, and the better and fitter chromosomes (advanced arrangements) that compare to the interconnecting weights and limits of neural system are produced. A little esteem, closes to zero, demonstrates that the system has summed up well

What's more, is prepared for the grouping issue. In this system GA seeks among a few arrangements of weight vectors all the while. The beginning populace is haphazardly produced. By selecting suitable parameters, similar to choice criteria, likelihood of traverse, likelihood of change, starting populace, and so forth., to the GA, high productivity and execution can be accomplish

VI. FORMULAE AND CALCULATIONS

The information comprised of danger elements gathered from 50 individuals through contextual analyses gave at the site of the American Heart Association [11]. The information was encoded as appeared in TABLE II. 70% of the information was utilized for preparing and 15% each to test and acceptance. A perplexity framework is delivered utilizing Matlab and precision is resolved (appeared in TABLE IV) as Accuracy = (TP + TN)/(TP + FP + TN + FN); where TP, TN, FP and FN means genuine positives, genuine negatives, false positives and false negatives, separately. The exactness of expectation of coronary illness on the preparation information was ascertained as 89% and precision on approval information was 96.2%. The minimum mean square mistake (MSE) accomplished was 0.034683 after 12 ages, as appeared in Figure 1. Results appear hereditary calculation and neural system approach gives better normal forecast exactness than the conventional ANN.

REFERENCES

- [1] "Global atlas on cardiovascular disease prevention and control", WHO, 2011.
- [2] Mozaffarian D, Wilson PW, Kannel WB, beyond established and novel risk factors: lifestyle risk factors for cardiovascular disease. *Circulation* 117: 3031–3038, 2008.
- [3] Anderson KM, Odell PM, Wilson PWF, Kannel WB. Cardiovascular disease risk profiles. *Am Heart J.*, 121: 293–298, 1991.
- [4] Kannel WB, An investigation of coronary heart disease in families. The Framingham offspring study. *Am J Epidemiol* 110:281–290, 1979.
- [5] R. Das, I. Turkoglu, and A. Sengur, Effective diagnosis of heart disease through neural networks ensembles, *Expert Systems with Applications*, Elsevier, pp. 7675–7680, 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* 2007. 32p.625–631.
- [7] Das, R., I. Turkoglu, and A. Sengur, Effective diagnosis of heart disease through neural networks ensembles. *Expert Systems with Applications*, Elsevier, 2009. 36 (2009): p.7675–7680.
- [8] LathaParthiban and R.Subramanian, "Intelligent Heart Disease Prediction System using CANFIS and Genetic Algorithm", *International Journal of Biological and Life Sciences*, Vol.3, No.3, pp.157-160, 2007.
- [9] N. Elfadil and A. Hossen, "Identification of Patients with Congestive Heart Failure Using Different Neural Networks Approaches", *Journal Technology and Health Core*, vol.17Issue4, December2009.
- [10] Centre for Disease Control and Prevention, http://www.cdc.gov/heartdisease/risk_factors.html.
- [11] American Heart Association <http://www.heart.org/HEARTORG/Conditions>
- [12] D. Isern, D. Sanchez, and A. Moreno, "Agents Applied in Health Care:A Review", *International Journal of Medical Informatics*, 79(3), pp.146-166, doi:10.1016/j.ijmedinf2010.01.003,2010.
- [13] Adebayo Peter Idowu, *Data Mining Techniques for Predicting Immunize-able Diseases: Nigeria as a Case Study*, *International Journal of Applied Information Systems*, 2013.
- [14] Monali Dey, Siddharth Swarup Rautaray, *Study and Analysis of Data mining Algorithms for Healthcare Decision Support System*, *International Journal of Computer Science and Information Technologies*(2014).
- [15] Jesmin Nahar, Tasadduq Imama, Kevin S. Tickle, Yi-Ping Phoebe Chen, *Association rule mining to detect factors which contribute to heart disease in males and females*, Elsevier, 2013.