International Journal of Research in Engineering, Science and Management
Volume-2, Issue-11, November-2019
www.ijresm.com | ISSN (Online): 2581-5792

Development of Heart Disease Prediction System Using Optimization Algorithm

S. Muthuramya[1], J. Jeyaranjani[2]
[1]*Student, Department of Computer Science and Engineering, Kalasalingam Academy of Research and Education, Virudhunagar, India*
[2]*Assistant Professor, Department of Computer Science and Engineering, Kalasalingam Academy of Research and Education, Virudhunagar, India*

***Abstract*: Heart Attack is a term that assigns a large number of medical conditions related to heart. The key to Heart (Cardiovascular) diseases to evaluate large scores of data sets, compare information that can be used to predict, Prevent, Manage such as Heart attacks. Heart Disease is mainly because of stress, family backgrounds, High blood Pressure, etc… Data analytics is used to incorporate world for its valuable use to controlling, contrasting and Manage a large data sets. It can be applied with an much success to predict, prevent, Managing a Cardiovascular Diseases. To solve this, we aim to implement the Data Analytics based on SVM and Genetic Algorithm to diagnosis of heart diseases. This result reveals the Genetic Algorithm as best optimized Prediction Models.**

***Keywords*: Data Analytics, Heart Disease, Linear SVM, Genetic Algorithm.**

1. Introduction

Data Analytics is the process of finding previously unknown patterns and trends in databases and using as extracting useful knowledge from a large amount of data. Among many applications fields of Data Analytics, medical field is very important. In medical field, Data Analytics is mainly used for diagnosis a disease. There are many disease diagnosis systems today. Accuracy is a main factor in disease diagnosis.

A. Heart diseases

The heart is an efficient part of our body. Life's, every moment is based on heart of our body, so the heart is not works correctly affects all parts of the body. The Heart problem mainly affects the human parts such as brain, kidney, etc… Cardiovascular diseases mainly refers to conditions that involve narrowed (or) blocked blood vessels that can lead to a heart attack, chest pain, stroke. Prediction of cardiovascular diseases is regarded as one of the most important techniques in the clinical Data Analytics. Heart failure is also an outcome of heart disease, and breathlessness can occur when the heart becomes too weak to circulate blood. Nowadays, the healthcare centre has been generating large amounts of data's about patients and their disease diagnosis reports are being especially taken for the prediction of heart diseases.

2. Literature survey

Ajinikya Kunjir, Harshal Sawant (December – 2016) [1] In this paper highlights the applications of classifying and predicting specific diseases by implementing the operation on medical data generated in the field of medical and healthcare. This paper used Genetic algorithm and Naive Bayess algorithm are compared for others testing their accuracy and performance on the training medical datasets.

H.S.Niranjana Murthy, M.Meenakshi (November – 2014) [2] This paper presents the development of a Neuro genetic model for the prediction of coronary heart diseases. The novelty of this work is feature subset selection using multi objective genetic algorithm without sacrificing the accuracy of ANN based heart disease predictor. This study exhibits early detection of heart disease with high testing accuracy of 89.58% through minimized feature subset, thereby reducing the complexity.

Haiyan Yang, Qifang Luo (February – 2015) [3] In this paper, the application of support vector machine (SVM) approach based on the statistics-learning theory of structural risk minimization in heart disease diagnosis. In this paper and genetic characteristics of the subset of choices is applied to reduce large dimensions and improve greatly the accuracy of classification.

Kiran Jyoti (October – 2012) [4] This Paper Generally used for Predicting a Heart Diseases, it is developed by 15 attributes. Earlier numerous data processing techniques are analyzed on heart disease information. This Paper used Artificial Neural Network (ANN), it usually referred to as a "Neural Network" (NN). It is a machine model supported biological neural network.

M. Anbarasi, E. Anupriya (February – 2015) [5]. This paper concluded that the Descision Tree Data Mining Techniques shows more accuracy and Low Error Rate. In this Paper Genetic algorithm is used to determine the attributes. The Three Classification of algorithm, Naive Bayes, clustering and Decision Tree are used to predict the heart disease with the same accuracy.

3. Architecture diagram

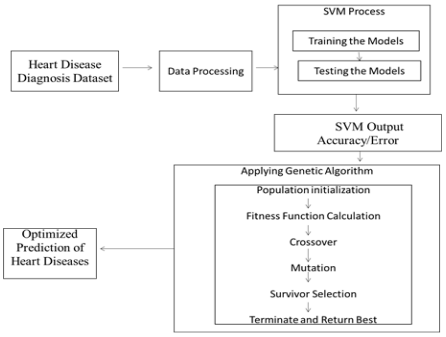

Fig. 1. The architecture of proposed work

In this prediction models it is analyzed with heart diseases data sets. The data processing techniques have been applied to correct the missing values present in the data set. The prediction system is optimized using SVM process used to applied produce balanced training and testing data sets. The SVM algorithm will proceed an exact output to show whether it is in accuracy rate or error rate. The Genetic Algorithm works with a set of an individual's, to represent possible solutions. The best-suited individuals create the next generation. Genetic algorithms are commonly used to generate high-quality solutions to optimization and search problems by relying on bio-inspired operators such as mutation, crossover and selection. The optimized prediction heart diseases are verified and validated with real time data sets successfully.

4. Dataset description

The Dataset used Cleaveland heart diseases datset obtained from UCI Repository. It includes 14 attributes and 298 instances of records with no missing values. The attribute is an integer data type that represents the age of the patient and ranges from 29 to 77 years. The Cp is an attribute of character data type for determining the chest pain type, represent the range 1 to 4. The tresbps is an attribute of integer data type of a resting blood pressure that lies from 94 to 200; the fbs is an attribute of logical data type lies from 1 or 0 which representing a true or false. The restecg is an attribute of character data type has the representative of resting electro cardio graphic results as three cases from 0 to 2. The thalach is an attribute of integer data type, the maximum heart rate achieved in beats per minute ranging from 71 to 202. The exang is an attribute character data type has the values from 1 or o which represents yes or no. The last attribute Target displays whether the patient is suffering from Heart Diseases. Similarly, all the attributes and their values are represented in following Table.

5. Algorithms used for implementation

A. Genetic algorithm

Genetic Algorithms (GAs) are search based algorithms based on the concepts of natural selection and genetics. GAs are a subset of a much larger branch of computation known as Evolutionary Computation. GA provides to solve a both constrained and unconstrained optimization problems based on a natural selection process. GA generally produce an optimization process using an initial process. The algorithm terminates when either a maximum number of generations has been produced.

Table 1
Attribute and Description of the Dataset

S. No.	Attribute Name	Type	Description	Range
1	Age	int	Age in years	29 – 77
2	Sex	char	Sex in num format	1- male 0- female
3	Cp	char	Chest pain type	1- typical angina 2-atypical angina 3-non angina pain 4-asymptotic
4	Trestbps	int	resting blood pressure	94 – 200
5	Chol	Int	serum cholesterol in mg/dl	131 – 417
6	Fbs	logical	fasting blood sugar	True – 1 False – 0
7	Restecg	char	resting electro cardio graphic results	Value 0 : normal Value 1: having ST-T wave abnormality Value 2: left ventricular hyperthrophy
8	thalach	int	maximum heart rate achieved in beats per minute	71 – 202
9	exang	char	exercise induced angina	Yes – 1 No – 0
10	Old peak	numeric	ST depression induced by exercise relative to rest	0 – 4.2
11	Peak slope	char	slope of the peak exercise ST segment	1 – 3
12	Ca	int	number of major vessels	0 – 3
13	Thal	char	displays the thalassemia	3 - normal 6 - fixed defect 7 - reversible defect
14	Target	int	Patient is suffering from Heart Diseases.	Yes – 1 No – 0

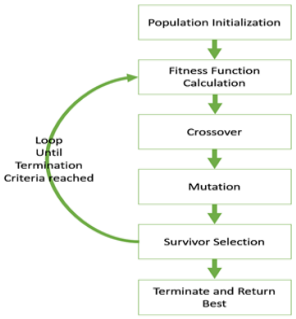

Fig. 2. Structure of genetic algorithm

GA population size depends on a great extent based on the problem domain. This population usually combined numerous possible solutions which are generated randomly forming the search space. Independent solutions are selected through a fitness - based process. GA is used for optimization problems because no previous knowledge is required about the domain problem. GA starts with N number of the accidental selected classifiers output. Genetic Algorithms belong to an evolutionary algorithms (EA), which generate solutions to optimization problems using techniques inspired by natural evolution, such as inheritance, mutation, selection, and crossover.

Algorithm:

1. To generate a random population of n chromosomes.
2. Evaluate the fitness f(x) of each chromosomes x in the population.
3. Create a new population to replicate following steps until the new population is reached.
4. Select two parent chromosomes from a population as maintained by their fitness.
5. With a crossover to crossover their parents to form a child. If no crossover was performed, children is an exact copy of parents.
6. With a mutation process, to mutate new children at each locus.
7. To accept new children in a new population.
8. To replace, new generated population for a future run of algorithm.
9. To test the end condition is whether satisfied, stop and return the best solutions in a current population.
10. Conclude the loop, go to step 2.

B. SVM algorithm

Support Vector Machine is a machine learning algorithm that analyzes data for classification and regression analysis. An SVM outputs a map of the sorted data with the margins between the two as far apart as possible. SVM algorithm is a kind of non-binary linear classifier. SVM is an algorithm that takes the data as an input and output of a line that separates those classes if it is possible.

SVM algorithm classifies both linear and non-linear method. The SVM process to find this hyper plane using support vector margins. The kernel trick is helpful for doing this which makes a linear classification in the feature space equivalent to nonlinear classification in the original space. SVM algorithm trained the models in different methods, and also it testing this process to analyze the data classification and regression analysis.

6. Result analysis and discussions

The trained SVM graph has gave an accurate error rate of RMSE (Root Mean Square Error) RMSE is the standard deviation of the prediction errors.

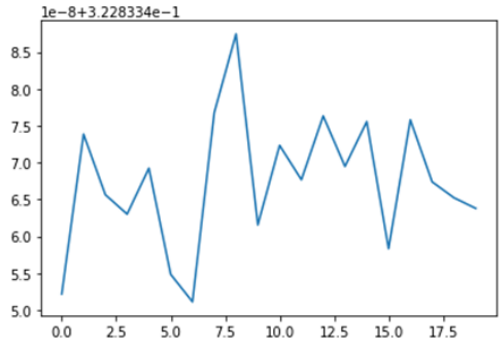

Fig. 3. RMSE graph between algorithm

Predictions rate are to measure of how far from the regression line points. When standardized observations are used as RSME inputs, there is a direct relationship with the correlation coefficient.

RSME formula for the sum of all values by the number of observation through trained SVM.

$$RMSE = \sqrt{\frac{\sum_{i=1}^{n} (P_i - O_i)^2}{n}}$$

Finally, it's found that the proposed SVM based prediction system offers valuable improvement in prediction of heart diseases. The results show that an improvement of 5% accuracy has been achieved by SVM classifier. It can be used more efficient way to predict the heart diseases.

7. Conclusion

In this paper, the overall objective of the work is to find the risk of heart diseases of a patient using the data's collected from the patients. The heart disease Prediction system using SVM and GA with a basic functions are analyzed. The results of classifications are performed over data sets obtain from 290 patients; has achieved 95% accuracy rate in SVM classifier. There are many interesting aspects for future work. In future the performance can be improved by subset selection process Genetic Algorithms can be used to select the input features.

References

[1] Ajinikya Kunjir, Harshal Sawant, "Data mining and visualization for prediction of multiple Diseases in health care", 2017 International

Conference on Big Data Analytics and Computational Intelligence (ICBDACI) IEEE 2017.
[2] H. S. Niranjana Murthy, M. Meenakshi "Dimensionality Reduction Using Neuro-Genetic Approach for Early Prediction of Coronary Heart Disease". International Conference on Circuits, Communication, Control and Computing (I4C 2014).
[3] Haiyan Yang, Qifang Luo, "Using Chaotic Adaptive PSO-SVM for Heart Disease Diagnosis" 2015 International Conference on Intelligent Computing, Communication and devices (ICCD), IEEE 2015.
[4] Kiran Jyoti "An Analysis of Heart Disease Prediction using Different Data Mining Techniques" International Journal of Engineering Research & Technology (IJERT 2012).
[5] M. Anbarasi, E. Anupriya, N. Ch. S. N. Iyengar, ―Enhanced Prediction of Heart Disease with Feature Subset Selection using Genetic Algorithm‖; International Journal of Engineering Science and Technology, Vol. 2(10), 2010.
[6] Chaitrali S. Dangare, Sulabha S. Apte, Improved Study of Heart Disease Prediction Systemusing Data Mining Classification Techniques; International Journal of Computer Applications, Volume 47, No. 10, June 2012.
[7] Asha Rajkumar, G. Sophia Reena, Diagnosis of Heart Disease Using Datamining Algorithm; Global Journal of Computer Science and Technology, Page 38, Vol. 10 Issue 10, September, 2010.
[8] E. P. Ephzibah, V. Sundarapandian, ―Framing Fuzzy Rules using Support Sets for Effective Heart Disease Diagnosis‖; International Journal of Fuzzy Logic Systems (IJFLS) Vol.2, No.1, February 2012.
[9] Niti Guru, Anil Dahiya and Navin Rajpal, "Decision Support System for Heart Disease Diagnosis using Neural Network", Delhi Business Review, Vol. 8, No. 1, pp. 1-6, 2007.
[10] Sellappan Palaniappan and Rafiah Awang, "Intelligent Heart Disease Prediction System using Data Mining Techniques", International Journal of Computer Science and Network Security, Vol. 8, No. 8, pp. 1-6, 2008.
[11] Humar Kahramanli, et al. "Design of a Hybrid System for The Diabetes and Heart Diseases" Elsevier, Expert Systems with Applications, Vol-35, pg. no.82–89, 2008.
[12] K. Usha Rani, "Analysis of heart diseases dataset using neural network approach", International Journal of Data Mining & Knowledge Management Process, Vol.-1, Issue-5, pg. no.1-8, 2011.
[13] Abhishek Taneja, "Heart Disease Prediction System using Data Mining Techniques", Oriental Scientific Publishing Corporation India, Vol.-6, Issue-4, Pg. no. 457-466, 2013.
[14] Chaitrali S. Dangare, et. al., "A Data Mining Approach for Prediction of Heart Disease using Neural Networks", International Journal of Computer Engineering and Technology, Vol-3, Issue-3, pg. no. 30-40, 2012.
[15] Resul Das, Ibrahim Turkoglu, Abdulkadir Sengur Diagnosis of valvular heart disease through neural networks ensembles, Elsevier 2009.