PREDICTION OF CARDIOVASCULAR DISEASE USING SUPPORT VECTOR MACHINE ALGORITHM

[1]V.P Amudhini, [2]T. Santhini, [3] R.Poonguzhali
[1]Final Year Student, [2] Final Year Student [3]Assistant Professor (SS)
[1]Computer Science and Engineering,
[1]Periyar Maniammai Institute of Science and Technology, Tamilnadu,India

Abstract : A cardiovascular disease is one of the most threatening diseases today. It consists of coronary artery diseases (CAD) like angina and myocardial infarction (heart attack), stroke, hypertensive heart disease etc. To predict these diseases accurately and more effectively, Machine learning algorithms are used. The algorithms used for prediction are Support Vector Machine (SVM), Naïve Bayes, Decision tree, K-Nearest Neighbor, Neural Networks. Support Vector Machine (SVM) algorithm is best among other algorithms and its accuracy is not below 50% in any testing and training dataset. This algorithm has low generalization error and it is also computationally inexpensive. In this system, the attributes of dataset is been reduced to increase the accuracy of prediction. This selected attributes are more significant to predict the disease.

IndexTerms – Cardiovascular disease, Machine Learning, Support Vector Machine.

I. INTRODUCTION

Cardiovascular disease is the reason for enormous number of deaths all over the world. A person gets overwhelmed by cardiovascular disease from aspects such as genealogy, age, hypertension; activities such as smoking and alcohol etc., these diseases should be predicted and treated at the starting stage to avoid deaths. To recognize these diseases correctly an effective system is needed. At the right time curing and treating of these diseases will be difficult which will definitely lead to large number of deaths.

Machine learning is a contemporary system in which it deals with large amount of data and provides more accurate results compared to the traditional way of diagnosis. The main drawbacks of traditional diagnosis are a) very much expensive and b) the results are not accurate. The input data which is given is analyzed by machine learning algorithm and the accurate results is given as output.

II.LITERATURE SURVEY

There are many survey and studies done on prediction of cardiovascular diseases using support vector machine algorithm

Firstly, Shaikh Abdul Hannan et. al. have discussed the prediction of cardiovascular diseases using SVM and FeedForward BackPropogation algorithm techniques. SVM is based on Structural Risk Minimization technique. An input layer, an output layer and a thin hidden layer is been used in FeedForward BackPropogation system. The FFBP algorithm takes 45-50 mins though that for SVM algorithm is 5-10 sec. But for medical prescription these algorithms give less relevant results.

Rifki Wijaya et al. proposed a system which states the elimination of Heart Disease using Machine learning Artificial Neural Networks(ANN). Attributes such as Smart Mirror, Smart Mouse, Smart Phones and Smart Chair is been used. These attributes have in-built sensors that detect various factors like pulse rates, body temperature, etc and help to predict cardio diseases. The main aim of this system is to predict the diseases one year ahead in advance. This prediction made using the historical data analysis.The Neural network algorithm gives the accurate results.

Sanjay Kumar Sen proposed a system which states predicting heart diseases using machine learning algorithms. The algorithms used are Bayesian Networks, SVM, Decision tree and K-Nearest Neighbor (KNN). In Bayesian Networks, the predictive accuracy is reduced when attribute are redundant. SVM shows the minimization of number of errors. KNN method finds the closet training point.

Poornima Singh et al. proposed an efficient heart disease prediction system (EHDPS) from neural network for the prediction of the risk level of the heart disease. In artificial neural network, MLPNN is one of the most important models. It contains one input layer, one or more hidden layers and one output layer. In MLPNN, the values are passed to the first hidden layer in input nodes, and then nodes of first hidden layer passes the values to the second and so on until the output is been produced. The dataset contains 303 records which were divided into two sets, training set and testing set i.e., 40% and 60% respectively. The results show that there are zero False Negative or False Positive entries, suggesting the system which predicts heart disease with 100% accuracy after applying neural networks on training dataset

Tulay karayilan and Ozkan kilic developed a system to predict heart disease using Artificial Neural Network (ANN) Backpropagation algorithm. Backpropagation algorithm is commonly used Artificial Neural Network learning technique for developing the heart disease prediction system. This algorithm trains the dataset. The performance is predicted using Accuracy, Precision and Recall. This paper says that the performance can be increased by reducing the number of neurons of the input layer so that the size of the hidden layer is reduced in dimensions and also shows 95% accuracy.

Theresa Princy.R et al. proposed a system for predicting the heart diseases by using Data Mining techniques. It makes use of KNN data mining technique. The prediction is been done with the help of two modules; initial and prediction module. In classifier module, data is been tested using the KNN algorithm and also the prediction module is been tested and predicted by ID3 algorithm. The accuracy level of 40.3% and also it was determined by the basic attributes such as blood pressure, cholesterol, Pulse rate, age and gender. By adding attributes such as smoking and previous heart disease, the accuracy level was increased up to 80.6% for the prediction.

Jagdeep Singh et al. proposed a system for predicting heart diseases using Association and Classification techniques. There are 5 stages which is been involved in CAR (Classification Associative Rule). The stages are selection, pre-processing and transformation, selection of associative rules, performance evaluation and predict diseases. For selecting the best 10 rules, two associative algorithms such as Apriori and FP-Growth was implemented for the preparation of the training data set. The algorithms of classification like J48, ZeroR, Naive-Bayes, OneR and k-nearest neighbor was implemented on training dataset and the output for each algorithms was evaluated. the maximum accuracy of 99.19% has been produced by the IBK with Apriori associative algorithm.

Prashasti Kanikar et al. developed a system for prediction of cardiovascular diseases using SVM and Bayesian classification. SVM technique is easy for interpreting results and detects error easily. Naïve Bayes algorithm provides minimum error rate and it also works with small amount of data. It is been concluded that SVM algorithm is better than Naïve Bayes algorithm as SVM works better for larger datasets.

III. METHODOLOGY

In this system, the input variables are taken by the dataset and the output variables is presence or absence of the disease. Prediction of cardiovascular disease in done with the help of machine learning algorithms as both the input and output variables are given for training.

Machine Learning:

Machine learning helps the system to learn automatically and execute problems with correct output. The computer consists of two attributes X and Y i.e., (X, Y); where the input data is X and the target output is Y. The input data consists of the dataset which is provided to the computer. The dataset is divided into two parts; the training data and testing data. The system is trained with the training data, hence the system learns by itself. Once the system is trained, the testing dataset is given to the system and the expected output is achieved.

Support Vector Machine:

SVM is a supervised machine learning algorithm which can be used for both linear and non-linear data for classification from the field called computational learning theory. However, it is been widely used in classification problems. In this algorithm, every data item is plotted in a point as n-dimensional space (where n is the number of features) with the value of each feature is being the value of the particular coordinate. Then, classification is done by finding the hyper-plane which differentiates the two classes well.

There are 4 modules in our system. They are 1) Collection of data from dataset, 2) Load the training data, 3) Algorithm implementation, 4) Load the testing data

Module 1: Collection of Data from a dataset

Data set of 143 records and 10 attributes is collected from UCI. The dataset is divided into 80:20 ratios of training and testing data. The 10 attributes which is been collected are Age, Gender, Smoke, CP, Chol, Fbs, Restecg, Thalach, Oldpeak, Slope.

Module 2: Load the Training Data

After collecting the dataset from the above attributes, it can be classified into 80% of training data and 20% data of testing data. The data is gathered in an excel sheet and loaded. The training data contains the information's of patients for each attributes which is been collected. The below figure1 shows the dataset which is been trained.

	A	B	C	D	E	F	G	H	I	J	K
1	age	gender	chest pain	chol	smoke	rest fbs	rest ecg	thalach	oldpeak	slope	prediction
2	63	1	1	233	-9	1	2	150	2.3	3	0
3	67	1	4	286	-9	0	2	108	1.5	2	0
4	67	1	4	229	-9	0	2	129	2.6	2	1
5	37	1	3	250	-9	0	0	187	3.5	3	3
6	41	0	2	204	-9	0	2	172	1.4	1	4
7	56	1	2	236	-9	0	0	178	0.8	1	0
8	62	0	4	268	-9	0	2	160	3.6	3	0
9	57	0	4	354	-9	0	0	163	0.6	1	1
10	63	1	4	254	-9	0	2	147	1.4	2	2
11	53	1	4	203	-9	1	2	155	3.1	3	2
12	57	1	4	192	-9	0	0	148	0.4	2	0
13	56	0	2	294	-9	0	2	153	1.3	2	1
14	56	1	3	256	-9	1	2	142	0.6	2	0
15	44	1	2	263	-9	0	0	173	0	1	0
16	57	1	3	168	-9	0	0	174	1.6	1	1
17	48	1	2	229	-9	0	0	168	1	3	0
18	54	1	4	239	-9	0	0	160	1.2	1	1
19	48	0	3	275	-9	0	0	139	0.2	1	0
20	49	1	2	266	-9	0	0	171	0.6	1	2
21	64	1	1	211	-9	0	2	144	1.8	2	3
22	58	0	1	283	-9	1	2	162	1	1	4
23	58	1	2	284	-9	0	2	160	1.8	2	0
24	58	1	3	224	-9	0	2	173	3.2	1	1
25	60	1	4	206	-9	0	2	132	2.4	2	0

Figure 1: Training Data

Module 3: Algorithm Implementation

SVM algorithm is used for classification and regression. SVM is a supervised learning method. Any mathematical formula can be used in SVM The given data is segregated by a hyper-plane. The 80% of data or training process and the balance 20% is for testing process. SVM gives better and accurate results. In this algorithm the output is plotted in graph.

Module 4: Load the testing data

In the previous modules the data set is collected, trained and then the algorithm is implemented. In this module test data is loaded and it tests the classifier using the test dataset where we can predict the class label for test data. The output is given by comparing the given training and testing data

	A	B	C	D	E	F	G	H	I	J	K
1	age	gender	chest pain	chol	smoke	rest fbs	rest ecg	thalach	oldpeak	slope	prediction
2	57	0	4	303	-9	0	2	159	0	1	2
3	71	0	3	265	-9	1	2	130	0	1	3
4	49	1	3	188	-9	0	0	139	2	2	4
5	54	1	2	309	-9	0	0	156	0	1	0
6	59	1	4	177	-9	0	0	162	0	1	1
7	57	1	3	229	-9	0	2	150	0.4	2	0
8	61	1	4	260	-9	0	0	140	3.6	2	0
9	39	1	4	219	-9	0	0	140	1.2	2	1
10	61	0	4	307	-9	0	2	146	1	2	3
11	56	1	4	249	-9	1	2	144	1.2	2	4
12	52	1	1	186	-9	0	2	190	0	2	0
13	43	0	4	341	-9	1	2	136	3	2	0
14	62	0	3	263	-9	0	0	97	1.2	2	1
15	41	1	2	203	-9	0	0	132	0	2	2
16	58	1	3	211	-9	1	2	165	0	1	2
17	35	0	4	183	-9	0	0	182	1.4	1	0
18	63	1	4	330	-9	1	2	132	1.8	1	1
19	65	1	4	254	-9	0	2	127	2.8	2	0
20	48	1	4	256	-9	1	2	150	0	1	0
21	63	0	4	407	-9	0	2	154	4	2	1
22	51	1	3	222	-9	0	0	143	1.2	2	0
23	55	1	4	217	-9	0	0	111	5.6	3	1
24	65	1	1	282	-9	1	2	174	1.4	2	0
25	45	0	2	234	-9	0	2	175	0.6	2	2

Figure 2: Testing Data

IV. RESULTS

The result of this paper states that the prediction of cardiovascular diseases using SVM was accurate. It processes the dataset given in Module 1. The attributes in Module 1 are taken into consideration and processed. The output of the processed data is shown in Module 3 where the last column predicts whether the person will be affected with diseases or not. They are ranged from 0 to 4 where 0 means no disease, 1 is likely to have disease, 2 means the slight presence of disease, 3 is presence of disease and 4 is highly diseased.

V. CONCLUSION

In this modern era, heart disease is very decisive health problem in recent times in our society. This paper has given advanced techniques for predication of this heart disease. Machine learning is a developing area of artificial intelligence which has showed good result in all fields of medical diagnosis with more accuracy. We have worked by reducing the attributes of dataset to increase the accuracy of prediction.

VI. ACKNOWLEDGMENT

We acknowledge faculty members of our department from Periyar Maniammai Institute of Science and Technology who greatly assisted our work

VII. REFERENCES

[1] V.P Amudhini, T. Santhini, Pavitra Kailash, D.Nivetha, R.Poonguzhali, "Survey on machine learning algorithms for prediction of cardiovascular disease, International Journal of Research and Analytical Reviews (IJRAR), 2018 IJRAR January 2019, Volume 6, Issue 1

[2] Shaikh Abdul Hannan, V. D. Bhagile, R. R. Manza and R. J. Ramteke, "Diagnosis and Medical Prescription of Heart Disease Using Support Vector Machine and Feedforward Backpropagation Technique", International Journal on Computer Science and Engineering, 2018.

[3] Rifki Wijaya, Ary Seijadi Prihatmanto and Kuspriyanto, "Preliminary Design of Estimation Heart Disease by using Machine learning ANN within one year" Joint International conference on Rural Information & Communication Technology and Electric- Vehicle Technology (rICT & IC2V-T), 2013.

[4] Sanjay Kumar Sen, "Predicting And Diagnosing Of Heart Disease Using Machine Learning Algorithms", International Journal Of Engineering And Computer Science,2017.

[5] Prashasti Kanikar and Disha Rajeshkumar Shah, "Prediction of cardiovascular diseases using SVM and Bayesian Classification", International Journal of Computer Applications (0975 – 8887) Volume 156 – No 2, December 2016. [6] Tülay Karayilan and Özkan Killiç, "Prediction of Heart Disease Using Neural Network" 2 nd International Conference on Computer Science and Engineering. 2017.

[6] Theresa Princy R and J Thomas, "Prediction of heart diseases using Data Mining techniques", International Conference on Circuit, Power and Computing Technologies [ICCPCT], 2016.

[7] Jagdeep Singh, Amit Karma and Harbagh Singh, "prediction heart diseases using Associative Classification", 5th International Conference on Wireless Networks and Embedded Systems (WECON), 2016.

[8] Poornima Singh, Sanjay Singh, Gayatri S Pandi-Jain, "Effective heart disease prediction system using data mining techniques", International Journal of Nano Medicine,2018.