Research Article

Automated Heart Disease Diagnostic System Based on SVM

Mr.Bramhadeo Bobade[1] and Dr. Anup Raut[2]

[1]PG Student, [2]Associate Professor, Department of Computer Engineering Imperial College of Engineering & Research Wagholi, Pune.

Abstract

In this paper the proposed framework predicts Heart Disease utilizing SVM calculation. Coronary illness is the most well-known reason for death universally. Coronary illness expectation utilizing information mining is one of the most intriguing and testing undertakings. The lack of authorities and high wrongly analyzed cases has required the need to build up a quick and proficient identification framework. As indicated by past framework the mix of clinical choice help with PC based patient record can lessen medicinal blunders, can be made progressively exact and thus improve understanding security. The framework helps in forecast of coronary illness by considering hazardous factor related with coronary illness. Parameters on which the disease is calculated or identified are based on Age, Gender, Chest Pain, Blood Pressure, ECG, Blood Sugar, and Heart Rate of the patient.

Keywords: SVM, Heart Disease, Data mining, support vector machine, prediction.

1. Introduction

Machine Learning Concept is one of the quickly developing perspective in current world. ML is useful in discovery and finding of different heart diseases. Support Vector Machine is an essential methodology of grouping in Machine Learning. The coronary illness comprises of set of range issue influencing the heart. It incorporates vein issues, for example, sporadic heart beat issues, powerless heart muscles, and inherent heart absconds, cardio vascular infection and coronary course infection. Coronary heart issue is a commonplace kind of coronary illness. It lessens the blood stream to the heart prompting a coronary failure. Life is totally reliant on productive working of the heart. The term Heart illness alludes to ailment of heart vein framework inside it. The heart is a significant organ of human body. On the off chance that the blood course to the body is deficient, the organs of the body that is cerebrum and heart quit working and passing happens in a couple of moment's minutes. Coronary illness is a main source of death worldwide from past 15 years. The normal hazard factors related are distinguished as age, family ancestry, Sex, Stress, elevated cholesterol, Heart rate, smoking, liquor consumption, overweight, physical inertia, chest torment type and terrible eating routine. Data got by analyzing the history record of the patient, it is conceivable to seclude the record and give report on HD on the off chance that it is certain or negative. Coronary illness is the most well- known reason for death all around. Numerous clinic data frameworks are intended to help persistent charging, stock administration and age of straightforward measurements. A few emergency clinics use choice emotionally supportive networks, yet they are to a great extent restricted. Mining is a strategy for investigating enormous arrangements of information to take out examples which are covered up and beforehand obscure connections and information location to help the better comprehension of restorative information to forestall coronary illness. Arrangement of coronary Heart Disease can be significant for the therapeutic experts if it is robotized with the ultimate objective of fast finding and accurate outcome. Nearness of heart malady accurately can save patients living days. The work consolidates the classes of Heart Disease using Support Vector Machine (SVM). In this a medicinal decision backing system for coronary sickness portrayal in a normal, reason, exact and quick way.

2. Literature Survey

This paper targets breaking down the different information mining strategies presented as of late for coronary illness expectation. Various information mining methods have been utilized in the analysis of CVD over various Heart sickness datasets. A few papers utilize just a single strategy for conclusion of coronary illness and different specialists utilize more than one information digging procedures for the analysis of coronary illness.

Nidhi Bhatla et al. [1] perceptions uncovered that the Neural Networks with 15 traits performed better in correlation with other information mining procedures

[1]. The examination study presumed that Decision Tree procedure demonstrated better execution with the assistance of hereditary calculations utilizing included subset determination. This examination work additionally proposed a model of Intelligent Heart Disease Prediction framework utilizing information mining methods specifically Decision Tree, Naïve Bayes and Neural Network. A sum of 909 records were gotten from the Cleveland Heart Disease database. The outcomes revealed in the exploration work advocated the better execution of Decision Tree methods with 99.6% exactness utilizing 15 characteristics. Be that as it may, Decision tree procedure in mix with hereditary calculation the exhibition detailed was 99.2% utilizing 06 characteristics. V.Manikandan et al. [2] suggested that affiliation rule mining is utilized to remove the thing set relations. The information order depended on MAFIA calculations which brought about better precision. The information was assessed utilizing entropy based cross approval and parcel strategies and the outcomes were analyzed. MAFIA (Maximal Frequent Itemset Algorithm) utilized a dataset with 19 characteristics and the objective of the exploration work was to have exceptionally exact review measurements with more significant levels of accuracy. Chaitrali S. Dangare and Sulabha S. Apte [3] demonstrated that Artificial Neural Network outflanks other information mining methods, for example, Decision Tree and Naïve Bayes. In this exploration work, Heart malady forecast framework was created utilizing 15 characteristics [3]. The exploration work included two additional characteristics weight and smoking for proficient determination of coronary illness in creating powerful coronary illness expectation framework.

In year 2014, Williamjeet Singh and Beant Kaur [5] distributed an examination paper in IJRITCC "Survey on Heart Disease utilizing Data Mining Techniques". The creator referenced crafted by huge number of specialists and thought about different information mining methods dependent on execution and precision. Jyoti Sonia, et.al. [6] in year 2011 exhibited three classifiers Decision Tree, Naïve Bayes and Classification by means of grouping to analyze the nearness of coronary illness in patients. Order by means of bunching: Clustering is the way toward gathering comparable components. This system might be utilized as a preprocessing step before sustaining the information to the characterizing model. Investigations were led with WEKA 3.6.0 apparatus. Informational index of 909 records with 13 distinct characteristics. All traits were made absolute and irregularities were settled for effortlessness. To improve the forecast of classifiers, hereditary hunt was consolidated. Perceptions display that the Decision Tree information mining method beats other two information mining systems in the wake of fusing highlight subset choice yet with high model development time.

M.V Shivsankar et al. [7] introduced utilization of Data Mining Technique in Healthcare and Prediction of Heart Attacks. The potential utilization of order based information mining procedures, for example, rule based Decision Tree, Naïve Bayes and Artificial Neural Network to the monstrous volume of human services information. Tanagra information digging device was utilized for exploratory information investigation, AI and measurable learning calculations. The preparation informational collection comprised of 3000 occurrences with 14 unique properties. The occurrences in the dataset were speaking to the aftereffects of various kinds of testing to anticipate the exactness of coronary illness. The exhibition of the classifiers was assessed and their outcomes were examined. The consequences of examination depended on 10 ten times cross-approvals. The examination made among these grouping calculations out of which the Naive Bayes calculation indicated better execution.

Vikas Chaurasia, et.al. [9] In their examination work utilized three well known information mining calculations CART (Classification and Regression Tree), ID3 (Iterative Dichotomized 3) and choice table (DT) separated from a choice tree or rule-based classifier to build up the forecast models utilizing a bigger dataset. Perception demonstrated that presentation of CART calculation was better when contrasted and other two order techniques. Abhishek Taneja [10] explore work was meant to structure a prescient model for coronary illness identification utilizing information mining strategies from Transthoracic Echocardiography Report dataset that is equipped for improving the unwavering quality of coronary illness determination utilizing echocardiography. The models were based on the preprocessed Transthoracic Echocardiography dataset with three diverse managed AI calculations J48 Classifier, Naïve Bayes and Multilayer Perception utilizing WEKA 3.6.4 AI programming. The presentation of the models was assessed utilizing the standard measurements of exactness, accuracy, review and F-measure. The best model to foresee patients with coronary illness seemed, by all accounts, to be a J48 classifier executed on chosen properties with a grouping precision of 95.56%. From a sum of 15 characteristics that were accessible, 8 properties that were profoundly pertinent in foreseeing coronary illness from Transthoracic Echocardiography dataset were chosen in the examination work.

Aditya Methaila et.al. [11] In their examination work concentrated on utilizing various calculations and blends of a few objective characteristics for powerful coronary episode forecast utilizing information mining. Choice Tree has beated with 99.62% precision by utilizing 15 properties. Additionally, the exactness of the Decision Tree and Bayesian Classification further improves in the wake of applying hereditary calculation to decrease the genuine information size to get the ideal subset of property adequate for coronary illness expectation.

 R. Chitra et.al. [12] Researchers in year 2013 introduced Hybrid Intelligent Techniques for the expectation of coronary illness. Some Heart Disease grouping framework was checked on in this investigation and closed with avocation significance of information mining in coronary illness conclusion and characterization. Neural Network with disconnected preparing is useful for malady expectation in beginning time and the great execution of the framework can be acquired by preprocessed and standardized dataset. The arrangement precision can be improved by decrease in highlights.
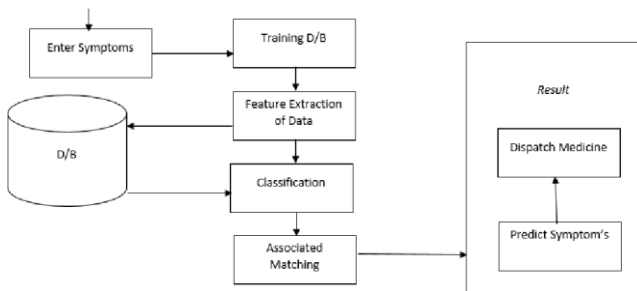
3. Methodology

Fig. 1. Proposed System Architecture.

Coronary illness is considered as one of the significant reasons for death all through the world. It can't be effectively anticipated by the medicinal specialists as it is a troublesome errand which requests mastery and higher information for forecast .System will anticipate heart infection relying upon the patient's verifiable dataset and SVM classifier. The framework likewise performs investigation of the coronary illness in light of age, sexual orientation and area

Classification is one of the well-known data mining issues. Data / objects should be divided into different classes or categories. For example, data can be broken down by file type, average file size, gigabytes, and megabytes of topical content. Classification is the learning process of a method that can be used to assign data objects to a subset of a class set. Some types of classification goals, first finding a good general that can predict with high accuracy the class of but far unknown data objects. Second, to consider each other's compact and easy to understand class pattern.

Cluster is a set of items. For example, data elements in different similarity groups between the data set in cluster classes in a single group cluster partitions. Every object nearby is a part of the neighborhood. There are two cluster targets. First, an intra-class is an inter-class second. Intercluster implies increasing the distance from the cluster. Intracluster implies reducing cluster distances.

In this process of selecting a subset of appropriate features for use in model design, feature selection also known as variable selection attribute selection and variable subset selection.

4. Algorithm

Support Vector Machine

Support vector machine (SVM) is a machine learning algorithm which is used for solving classification and regression problems. Vladimir Naumovich Vapnik and his colleagues in 1992 introduced SVM. SVM regression is taken into account a nonparametric technique because it relies on kernel functions. It uses the maximum margin algorithm: a non-linear function is leaned by linear learning machine mapping into high dimensional kernel induced feature space.
1. Prepare the pattern matrix
2. Select the kernel function to use
3. Select the kernel function parameter and the value of C
Use the values suggested by the SVM software, or you can set apart a validation set to determine the values of the parameter.
4. Execute the training algorithm and obtain the $\propto i$
5. Unseen data can be classified using the $\propto i$ and the support vectors.

Table 1: - Features of Heart Perdition

Sr.No	Information of Features
1	Age
2	Gender
3	Chest Pain
4	Blood Pressure
5	ECG
6	Blood Sugar
7	Thal
8	Heart Rate

5. Experimental and Result

The dataset consists of total 1300 records in Heart disease database. The total records are divided into two data sets one is used for training consists of 900 records & another for testing consists of 400 records. Some of the field where been missing which were replaced by finding mean mode method, such method is known as Data-pro-processing. Once Data preprocessing is performed data mining techniques using SVM has be applied. We have achieved 89.5% of accuracy by using SVM technique.

Below is the table where we had a survey and have mentioned the different techniques used by various authors and their achieved accuracy.

Table 2:- Comparison of Accuracy

Author	Technique	Accuracy
Jyoti Sonia, et.al.	Naïve Bayes, Decision Tree, KNN	77%

K.Srinivas et.al.	Naïve Bayes, knn and D.L.	81.11% 81.48% 81.11% 80.96%
Nidhi Bhatla et.al.	Naïve Bayes,	84.5%
	Decision Tree, Neural Network	
Chaitrali S. dangare & Sulabha S. Apte, Abhishek Taneja	Naïve Bayes, Decision Tree, Neural Network	87.0%
R. Chitra et. al.	Naïve Bayes,J48 unpruned tree, Neural Network	87.0%
Vikas Chaurasia, et.al.	Hybrid Intelligent Techniques	78.9% 81.41%
Proposed System	**SVM**	**89.5%**

Conclusion

The heart is a significant organ of the human body. On the off chance that the blood course to the body is insufficient, the organs of the body that are mind and heart quit working and passing happens in a couple of moment's minutes. Coronary illness is a main source of death worldwide from recent years. So it is essential to anticipate Heart ailment at a beginning period to stay away from human passing. The significance of information mining in therapeutic area is acknowledged and steps are taken to apply important procedures in the Disease Prediction. The parameter on which coronary illness is for the most part subordinate is incredibly defenseless and variation. So in the wake of getting authentic data about the patient, coronary illness can be anticipated. Here, the proposed framework predicts the coronary illness dependent on the verifiable clinical information of patient utilizing SVM (Support Vector Machine) calculation. SVM classifier is utilized, as the characterization exactness, affectability and particularity of the SVM calculation is seen as high, consequently making it a predominant option for conclusion. Information Analysis is likewise done dependent on age, and gender.

Acknowledgment

This paper would not be possible without the contributions of numerous researchers in this ever growing field. Thus we would like to thank all of them.

References

[1]. Nidhi Bhatla, Kiran Jyoti, "An Analysis of Disease Prediction using Different Data Mining Techniques" International Journal of Engineering and Technology Vol.1 issue 8 2012.

[2]. V. Manikandan and S. Latha, "Predicting the Analysis of Disease Symptoms Using Medical Data Mining Methods "International Journal of Advanced Computer Theory and Engineering", Vol. 2, Issue. 2, 2013.

[3]. Chaitrali S. Dangare, Sulabha S. Apte, "Improved Study of Heart Disease Prediction System using Data Mining Classification Techniques." International Journal of Computer Applications (0975 – 888) Volume 47– No.10, June 2012.

[4]. Shadab Adam Pattekari and Asma Parveen, "Prediction system for heart disease using naïve bayes", International Journal of Advanced Computer and Mathematical Sciences, 2012.

[5]. Beant Kaur and Williamjeet Singh., "Review on Heart Disease Prediction System using Data Mining Techniques", IJRITCC ,October 2014.

[6]. Jyoti Soni et.al. "Predictive Data Mining for Medical Diagnosis: An Overview of Heart Disease Prediction", International Journal of Computer Applications (0975 – 8887) Volume 17– No.8, March 2011.

[7]. B.Venkatalakshmi, M.V Shivsankar, "Heart Disease Diagnosis Using Predictive Data mining", International Journal of Innovative Research in Science, Engineering and Technology Volume 3, Special Issue 3, March 2014.

[8]. Hlaudi Daniel Masethe, Mosima Anna Masethe, "Prediction of Heart Disease using Classification Algorithms", Proceedings of the World Congress on Engineering and Computer Science 2014.

[9]. Vikas Chaurasia, et al, "Early Prediction of Heart Diseases Using Data Mining Techniques", Caribbean Journal of Science and Technology ISSN 0799-3757, Vol.1,208-217, 2013.

[10]. Abhishek Taneja, "Heart Disease Prediction System Using Data Mining Techniques" Oriental Journal of computer science & Technology ISSN: 0974-6471December2013.

[11]. Aditya Methaila , Prince Kansal , Himanshu Arya , Pankaj Kumar , "Early heart disease prediction using data mining techniques" , Computer Science & Information Technology (CS & IT), 2014.

[12]. R. Chitra , V.Seenivasagam, "Heart Disease Prediction System" ,ResearchGate , Conference: ICPERES 2014