International Journal of Research in Engineering, Science and Management
Volume-2, Issue-9, September-2019
www.ijresm.com | ISSN (Online): 2581-5792

Design and Develop Futuristic Prediction Regarding Details of Health System for Heart Diseases Using the Integration of Data Mining Techniques

S. M. Bhadkumbhe[1], Vaishnavi Metkar[2], Kshitija Shitole[3], Harshwardhan Suryawanshi[4], Ruchita Prabhune[5]

[1]*Assistant Professor, Department of Computer Engineering, PDEA's College of Engineering, Pune, India*
[2,3,4,5]*Student, Department of Computer Engineering, PDEA's College of Engineering, Pune, India*

***Abstract*: Tremendous amount of people are prone to heart deformity and in today's world lot of people are prone to these diseases that is CVD (Cardiovascular disease) and it is one of the major reasons for heart problem. Consequently, we examined and determine first stage it will increase death factor illness and causes death. There is no adequate research refereeing to the tools to discover relationships and trends in data especially in the medical sector. Complex clinical data is driven by technologies such as health care about patients and other hospital resources. Rich collection are used in order to Data mining techniques examine methodology in detail the different perspectives and deriving useful detail. Our project is perspective to design and develop futuristic prediction regarding details of health system for heart diseases based on predictive mining. There are various experiments have been conducted to a specific relationship for the performance of various predictive data mining techniques including Decision tree, Naïve Bayes and k-mean algorithms. In this proposed work, a 15 attribute structured clinical database from UCI Machine Learning Repository has been used as a source data. Naive Bayes have been prescribed and their schedule on diagnosis has been compared.**

***Keywords*: Data Mining, Naïve Bayes, k-mean, Decision tree, CVD.**

1. Introduction

Heart is the very important part of the body as it co-ordinates all the functioning of different organ through blood circulation. Heart leads to different type of diseases that affect the human blood vessels too. The symptoms of any disease vary from person to person. In the case of heart disease, the symptoms vary from person to person depending upon the type of heart disease. Coronary heart disease is the most common type of heart disease. It is occurred due to lack of blood supply to the heart. Coronary heart disease is the leading cause of death. There are 48% death due to coronary heart disease. Cardiovascular diseases (CVD) is one of the underlying cause of death. The cost for the detection of heart disease is more and it might be not affordable for all the peoples. Due to lack of income the proper diagnosis of heart disease not possible.

This paper identifies the risk factor for different types of heart disease. In today's world the hospitals manage the patient's data on the system. All of this data is stored in huge databases of electronic medical records systems. This system generates huge amount of data in different formats. But this data is not used for further processing of clinical decision. So the main aim is to utilize the clinical data for prediction of disease. Data mining for health care utilizes the medical records for the prediction of heart disease specially for CVD.

A data mining model was developed with a clinical laboratory database using a naïve Bayes, K-means classifier to detect cardiovascular risk, and it was tested for its accuracy in predicting three levels of risk.

2. Related work

A. Non-laboratory based risk factor for automated heart disease detection

In this paper Author examined the possibility of reducing the number of risk factors, especially those require costly and invasive laboratory-based results, to detect heart disease. We found that age, gender, rest blood pressure measurement, maximum heart rate, and the abnormality of ST segment in the rest ECG can be used to feed into a simple generalized linear model and achieved closely comparable accuracy as earlier works that utilized a more comprehensive input set. For example, authors of the well-known work that included laboratory based data only yielded approximately a 77% classification accuracy (using a logistic regression approach). Furthermore, research advances in processing massive datasets may provide a useful real-time tool and massive information learning platform that cardiologists can assess an individual patient's risk for heart disease more accurately with less laboratory cost and faster. It is worth noting that the above ECG data used in our method can be obtained easily given the recent advances in wearable sensor for automated ECG analysis.

B. Review on fuzzy expert system and data mining techniques for the diagnosis of coronary artery disease

This paper reviews previous research using data mining, fuzzy, and combination techniques of data mining and fuzzy used to diagnose or predict heart disease or coronary artery disease. The dataset used varied widely into 2 namely the open source dataset derived from the UCI heart disease repository and data from hospitals and experts. The process of filling missing values in the dataset and selecting significant features to diagnose heart disease greatly affected the outcomes of the built system. In fuzzy systems, fuzzy rules and membership functions were important, the dependency of getting the rules of the experts and the results of the laboratory can be improved by the technique of data mining algorithm that can generate rules, as well as the fuzzy membership function can be optimized with Particle Swarm Optimization, Imperialist Competitive Algorithm, and Genetic Algorithm. The highest accuracy result for data mining technique usage was 99% using J48 algorithm, Naive Bayes, REPTREE, CART, and Bayes Net. For the results of the use of fuzzy was with the accuracy of 94% with the use of methods Mamdani inference system and triangle membership functions combined with trapezoidal. While the use of combination of data mining and fuzzy with 94.92% accuracy was by using a combination of decision tree as rule extraction, fuzzy, and ICA for the optimization of fuzzy membership function.

C. Prediction of cardiac disease based on patient's symptoms

The researchers that use pattern recognition of these data mining methods help in predicting models based on the cardiovascular diagnose domain The experiments that were carried out using these classifications based algorithm such as Naïve Bayes, Decision Tree, K-NN and Neural Network and these results have proven to be that of Naïve Bayes technique that have performed better than the others when utilized by thee techniques. The researchers uses K means clustering algorithm on that particular heart disease where the warehouse which relate to extract data relevance to the heart disease, and applies to type MAFIA (Maximal Frequent Item set Algorithm) algorithms to calculate weightage of the frequent patterns which are probably very significant to heart attack predictions.

D. Prediction of heart diseases using machine learning

The Heart Disease Prediction System using Machine learning algorithm, viz. MLP provides its users with a prediction result that gives the state of a user leading to CAD. Due to the recent advancements in technology, the machine learning algorithms are evolved a lot and hence we use Multi Layered Perceptron (MLP) in the proposed system because of its efficiency and accuracy. Also, the algorithm gives the nearby reliable output based on the input provided by the users. If the number of people using the system increases, then the awareness about their current heart status will be known and the rate of people dying due to heart diseases will reduce eventually.

E. Empirical study on classification of heart disease dataset-its prediction and mining

Application of data mining techniques in HDD is an emerging trend in the world. It has attracted the attention of medical practitioners and academics. This paper has identified ten articles related to application of data mining techniques in HDD, and published between 2006 and 2016. It endeavors to provide a research review on the application of data mining in the HDD domain and methods which are most often used. Even though this review cannot claim to be through, it does give reasonable imminent and shows the frequency of research on this subject. The results presented in our paper have a number of important implications they are,

1. Research on the application of data mining in HDD will boost significantly in the future, based on past publication rates and the increasing awareness in the area.
2. With respect to the research conclusion, we advise more research can be accomplished in the Heart Disease domain.

3. Algorithm

A. Naïve Bayes

Naïve Bayes is a probabilistic machine learning algorithm based on Bayes Theorem, used in a wide variety of classification tasks. Naive Bayes classifiers are a collection of classification algorithms based on Bayes Theorem. Bayes theorem plays a critical role in probabilistic learning and classification. Using Naive Bayes classifier, the system can conceal knowledge associated with diseases from historical records of the patients having heart disease. The probability of heart disease can be calculating by analyzing the historical reports dataset and predict the result of heart disease.

Working:

- Convert a dataset into a frequency table.
- Create Likelihood table by finding the probabilities like age probability = 54
- Now, Use Naïve Bayesian equation to calculate the posterior probability for each class. The class with the highest posterior probability is the outcome of prediction.

B. K-means

k-means algorithm is an iterative algorithm that tries to partition the dataset into K pre-defined distinct non-overlapping subgroups (Clusters) where each data point belongs to only one group. It tries to make the inter-cluster data points as similar as possible while also keeping the clusters as different as possible. It assigns data points to a cluster such that the sum of the squared distance between the data points and the clusters centroid is at minimum. The less variation we have within clusters, the more homogeneous the data points are within the same cluster.

Working:

- Specify number of clusters K.

- Initialize centroids by first shuffling the dataset and then randomly selecting K data points for the centroids without replacement.
- Keep iterating until there is no change to the centroid i.e. assignment of data points to clusters isn't changing.
- Compute the sum of the squared distance between the data points and all the centroids.
- Assign each data point to the closest cluster.
- Compute the centroid for cluster by taking average of all the data points that belongs to that cluster.

C. *Decision Tree*

Decision tree algorithm belongs to the family of supervised learning algorithms. It can be used for solving regression and classification problems. The general motive of using decision tree is to create a training model which can use to predict class or value of target variables by learning decision rules inferred from prior data.

Working:

- Place the best attribute of the dataset at the root of the tree.
- Split the training set into subsets. Subsets should be mad in such a way that each subset contains data with the same value for an attribute.
- Repeat step 1 and step 2 on each subset until you find leaf nodes in all the branches of the tree.
- In decision tree, for predicting a class label for a record we start from the root of the tree. We compare the values of the root attribute with records attribute. On the basis of comparison, we follow the branch corresponding to that value and jump to the next node. We continue comparing our records attribute values with other internal nodes of the tree until we reach a leaf node with predicted class value.

4. System Architecture

A. *User module*

User module is own by patient. New patient registers themselves by using register procedure and get their id and password. User login using id and password. Patient entered the data like name, age, gender, email, mobile number at the time of registration.

B. *Doctor module*

Doctor module also have login and registration same as in user module. At the time of new doctor registration doctor must fill details like name, id, email, contact no., speciality in etc.

C. *Checking*

Checking modules consist of 15 type of different fields such as age, sex, chest pain type, BP, cholesterol, blood sugar level, heart rate, resting ECG, induce angina, old peak, slope, major vessel, smoke, alcohol.

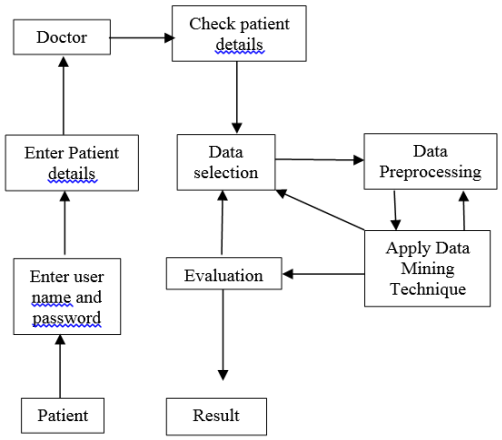

Fig. 1. System Architecture

5. Conclusion

To diagnose the disease at early stage it uses k-means, Naïve Bayes, Decision tree. Early detection of disease and use of different attributes to detect disease such as medical as well as non-medical terms. Due to the recent advancements in technology, the data mining algorithms are evolved a lot and hence use of k-means, Naïve Bayes in the proposed system because of its efficiency and accuracy. Also, the algorithm gives the nearby reliable output based on the input provided by the users. If the number of people using the system increases, then the awareness about their current heart status will be known and the rate of people dying due to heart diseases will reduce eventually.

References

[1] P. Sudeshna, S. Bhanumathi and M. R. A. Hamlin, "Identifying symptoms and treatment for heart disease from biomedical literature using text data mining," *2017 International Conference on Computation of Power, Energy Information and Commuincation (ICCPEIC)*, Melmaruvathur, 2017, pp. 170-174.

[2] K. M. M. H. Sonet, M. M. Rahman, P. Mazumder, A. Reza and R. M. Rahman, "Analyzing patterns of numerously occurring heart diseases using association rule mining," *2017 Twelfth International Conference on Digital Information Management (ICDIM)*, Fukuoka, 2017, pp. 38-45.

[3] C. Sowmiya and P. Sumitra, "Analytical study of heart disease diagnosis using classification techniques," *2017 IEEE International Conference on Intelligent Techniques in Control, Optimization and Signal Processing (INCOS)*, Srivilliputhur, 2017, pp. 1-5.

[4] S. Babu *et al.*, "Heart disease diagnosis using data mining technique," *2017 International conference of Electronics, Communication and Aerospace Technology (ICECA)*, Coimbatore, 2017, pp. 750-753.

[5] Shadab Adam Pattekari and Asma Parveen, "Prediction system for heart disease using naive bayes," in International Journal of Advanced Computer and Mathematical Sciences, vol. 3, no. 3, pp. 290-294, 2012.