IOSR Journal of Engineering (IOSRJEN) **www.iosrjen.org**
ISSN (e): 2250-3021, ISSN (p): 2278-8719
PP 55-59

Heart Disease Predication using Predicative Data Mining

Ekta Dhote[1], Pallavi Moon[1], Minal Kelwadkar[1], Alkatai Nakade[1], Vidhya bodhe[2]

[1]BE Students, Department of Computer Science Engineering, AbhaGaikwad-Patil College of Engineering, Nagpur, Maharashtra, India.

[2]Assistant Professor, Department of Computer Science Engineering, AbhaGaikwad-Patil College of Engineering, Nagpur, Maharashtra, India.

***Abstract**:-The successful application of data mining in highly visible fields for ex. e-business, marketing, applications in other industries and sectors. Among these sectors just discovering is healthcare. "The healthcare environment is information rich, but knowledge is poor". There is a wealth of some data are available within the healthcare systems. However, there is a lack of the effective analysis tools to discover the hidden relationships and trends in various data. The Number of experiment has been conducted to the compare of performance in predictive data mining technique on the same dataset. The outcome of reveals that the Decision Tree outperforms and sometime the Bayesian classification is having similar accuracy of decision tree but some other predictive methods like KNN, Neural Networks, Classification is based on clustering are not performing well. This to provide a survey of current techniques of knowledge discovery in databases using data mining techniques that are in use in today's medical research particularly in Heart Disease Prediction. For this purpose of seven different disorders have been specified such as sleep breathing disorder, rapid eye movement behavior disorder, periodic leg movement disorder, insomnia disorder, narcolepsy disorder, nocturnal frontal lobe epilepsy disorder, bruxism disorder and with them one healthy subject. Several ECG records have been collected for these sleep disorders and analyzed using transform.*

***keyword:**- Bayesian classifier, KNN algorithm,neural network, data mining, decision tree,*

I. Introducation

Medical data mining has the great potential for exploring the hidden patterns in the data sets of the medical domain. In this patterns can be utilized for clinical diagnosis. The available of raw medical data are widely distributed, heterogeneous in the nature and voluminous. These data are needed to be collected in organized form. These collected data can be the integrated to form of the hospital information system. The Data mining technology provides the user oriented approach to novel and hidden patterns in the data. The World Health care Organization has estimated that 12 million deaths occur worldwide, every year due to the Heart diseases. The Half deaths in the United States and other developed countries occur due to cardio vascular diseases. The term of Heart disease encompasses the various diseases that affect the heart. Heart disease was the major issues of casualties in the different countries including India. Heart disease are kills one person every 34 seconds in the United States. Coronary heart disease, Cardiomyopathy and Cardiovascular disease are some categories of heart disease. The term of "cardiovascular disease" include a wide range of condition that affects the heart and the blood vessels. In the manner which blood is pumped and circulated through the body. Cardiovascular disease (CVD) results in the several illness, disability and death. The diagnosis of diseases is essential and complex job in medicine. Medical diagnosis is regarded as an important task are complicated that are need to be executed accurately and efficiently. The automation of the system would be extremely advantageous. Unfortunately all the doctors do not possess expertise in every sub specialty and moreover there is a shortage of resource persons at certain places. Therefore, an automatic medical diagnosis system would be probably the extremely beneficial by bringing all of them are together. Appropriate computer-based information and/or decision support systems can aid in achieving the clinical tests at the reducing cost. Efficient and accurate implementations of automated system are needs comparative studies of various techniques are available. Analyse of the different predictive/ descriptive data mining techniques are proposed in recent years for the diagnosis of heart disease.

II. Literature Review

In the [1] research paper is major challenge facing of healthcare organizations (hospitals, medical centre) is the provision of quality services at inexpensive costs. Quality service implies diagnosing patients correctly and administering treatments are that effective. Poor clinical decisions can be lead to terrible consequences which are therefore unacceptable. Hospitals are must be also minimize the cost of clinical tests.

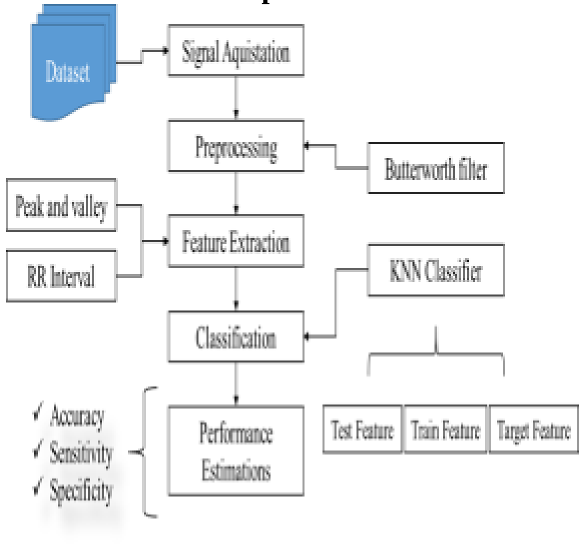
They can achieve these results by employing the appropriate computer-based information and/or decision support systems. Health care data is massive. It includes the patient centric data, resource management data and transformed data. Health care organizations must have ability to analyzethe data. The Treatment records of the millions patients can be stored and computerized and data mining techniques may be help in answering of several important and critical questions are related to the formal definition of KDD is given as follows: “Data mining is the non trivial extraction of implicit previously unknown and potentially useful information about the data”. Data mining technology provides a user-oriented approach to the novel and hidden patterns of the data. The discovered knowledge can be used by the healthcare administrator to improve the quality of health care services.

In the[2] research paper, they can be achieved these results by employing appropriate computer-based information. Discovery of hidden patterns are the decision support systems. The healthcare industry collects the huge amounts of data. Unfortunately, are not “mined” to discover hidden information for the effective decision making and relationships often goes to unexploited. Advanced data mining techniques can be help to remedy in this situation. The research paper has developed a prototype Heart Disease Prediction System (HDPS) using data mining techniques, Decision Trees, Naïve Bay’s and Neural Network. Results show that each technique has its unique strength in realizing the objectives of the mining goals. Using the medical profiles such as age, sex, blood pressure and blood sugar it can be predict the likelihood of patients getting a heart disease. It enables significant knowledge such as patterns, relationships between medical factors related to heart disease to be established. HDPS are Web-based, user-friendly, scalable, reliable and expandable.

In the[3] research paper Data Mining is the nontrivial process of identifying the valid, novel, potentially useful and ultimately understandable pattern in data with the wide use of databases and explosive the growth in their sizes. Data mining refers to the extracting or “mining” knowledge from large amounts of data. Data mining is the search for the relationships and global patterns that exist in the large databases but are hidden data among large amounts of data. The essential process of Knowledge Discovery databases is the conversion of the data into knowledge in the order to aid in decision making, referred to as the data mining. The Knowledge Discovery process are consists of an iterative sequence of data cleaning, data integration, data selection, data mining pattern recognition and the knowledge presentation. Data mining is the search for the relationships and global patterns that exist in large databases but are hidden data among large amounts of data.

In the [4] research paper,a major challenge are faced by the health care organizations, such as hospitals and medical centers. The provision of quality services at the affordable cost. The quality of service implies diagnosing patients properly and administering effective treatments. The available heart disease database consists of the both numerical and categorical of the data. Before further the processing, cleaning and filtering are applied on these records in order to filter to the irrelevant data from the database. The proposed system can determine an exact the hidden knowledge, i.e. patterns and relationships associated with heart disease from the historical heart disease database. It can be answer the complex queries for diagnosing heart disease.Therefore, it can be helpful to health care practitioners to make the intelligent clinical decisions. Results showed that the proposed system has its unique potency in the realizing objectives of the defined mining goals.

III. Proposed Work



IV. Proposed Work

Modules:

- Signal Aquistation
- Preprocessing
- Feature Extraction
- Classification
- Performance Estimations

The modules are as follows:- 1) signal aquisation:- Electroencephalography (ECG) is an electrophysiological monitoring method to the record of electrical activity of the brain. ECG measures voltage fluctuations resulting from ionic current within the neurons system of the brain. In clinical concept ECG refers to the recording of brains spontaneous to the electrical activity over a period time, as recorded from multiple electrodes placed on the scalp. The Diagnostic applications are generally focus on the spectral content of ECG.The type of neural oscillations (“brain waves") that can be observed in ECG signals.ECG is most often used to diagnose epilepsy, which causes the abnormalities in ECG readings. ECG used to be a first-line method of diagnosis for tumors, stroke and other focal brain disorders, but this use has decreased with the advent of high-resolution anatomical imaging techniques such as the magnetic resonance imaging (MRI) and the computed tomography (CT). ECG continues to be a valuable tool for research and diagnosis, especially when the millisecond-range of temporal resolution (not possible with CT or MRI) is required.

Derivatives of the ECG technique include Evoked potentials (EP), which involves averaging the ECG activity of the time-locked to the presentation of a stimulus of some sort. Event-related potentials (ERPs) refer to the averaged ECG responses that are time-locked to the more complex processing of stimulation of this technique is used in cognitive science, cognitive psychology and psycho physiological research.

Preprocessing:

ECG is the main tool used by the physicians for identifying the interpretation of Heart condition. The ECG should be free from the noise of good quality for the correct diagnosis. In the real time to the situations ECG are corrupted by the many types of artifacts. The high frequency noise is one of ECG. The present paper deals with the removing of noise from ECG of the high frequency contents with help of the Low pass digital filter of the cutoff frequency 100Hz. The sampling period used is to .001sec. The filter is designed with Butterworth Approximations. The results of before filtration and after filtration are depicted in the paper. Paper contains the detail design of the digital Butterworth filter and its realization. The experimentation is performed by on the database generated in the Laboratory. The Butterworth filter is the type of signal processing filter is designed to have a frequency response flat as the possible in the pass band. It is also referred to as a maximally flat magnitude the filter. It was first described in 1930 by the British engineer and physicist Stephen Butterworth in his paper entitled "On the Theory of the Filter Amplifiers. Butterworth filter had a reputation for the solving "impossible" in the mathematical problems. At the time of filter design required to a considerable amount of designer experience due to limitations of the theory in use.

Feature Extraction

Energy

The feature extraction using these methods is based on the energy, frequency, and the length of the principal track. The ECG signal is firstly divided into segments; then, the construction of a three-dimensional feature vector for each segment will be take place in data.

Variance & Standard Deviation

The variance and the closely-related to the standard deviation are measures of the how to spread out of the distribution. In other words, these are measures of variability. The variance is computed to the average squared deviation of the each number from its mean. For example, for the numbers 1, 2, and 3, the mean is 2 and the variance is:

$$\sigma^2 = \frac{(1-2)^2 + (2-2)^2 + (3-2)^2}{3} = 0.667$$

$$\sigma^2 = \frac{\sum (X-\mu)^2}{N}$$

Where μ is mean and N is number of scores.

When the variance is computed to in a sample, the statistic

$$S^2 = \frac{\sum (X - M)^2}{N}$$

This gives an unbiased estimate of σ². Since the samples are usually used to estimate the parameters s² is the most commonly used to measure of the variance. Calculating of the variance is the most important part of the many statistical applications and analyses. In the first step in calculating of the standard deviation.

Classification

In the pattern recognition, the k-nearest neighbor's algorithm (k-NN) is a non-parametric method used for the classification and regression method. In both the cases input consists of the k closest in the training examples of the feature space. The output depends on whether k-NN is used for the classification and regression. In the k-NN classification of the output is the class membership.

In k-NN regression method, the output is the property value for the object. In This value is the average of the values of it's the KNN.

K-NN is the type of the instance-based learning or lazy learning. Where the function is only approximated locally and the all computation is deferred to until classification. The k-NN algorithm is among the simplest of the all machine learning algorithms.

Both for the classification and regression method are useful technique can be assigned weight to the contributions of the neighbors so that the nearest neighbors contribute to the more average than the more distant ones. For ex. a common weighting scheme consists in giving the each neighbor a weight of 1/d, where d is the distance to the neighbor.

Performance Estimation

The performance of the process is measured in the terms of performance the metrics like Accuracy, Sensitivity, Specificity and time consumption.

TP - the total number of the correctly classified foreground (true positives).

TN - the total number of the wrongly classified foreground (true negatives).

FN - the total number of false negatives values, which accounts for the incorrect number of foreground pixels classified as the background (false negatives).

FP - the total number of false positives values, which means the pixels are incorrectly are classified as the foreground (false positives). The performance values are calculated for the each frames of the input based.

V. Conclusion

The problems of constraining and summarizing the different algorithms of data mining are used in the field of medical prediction are discussed. In The focus is on using a different algorithm and the combinations of several target attributes for the intelligent and the effective heart attack of the prediction using data mining techniques. For the predicting heart attack are significantly 10 attributes are listed and with basic of data mining technique are used other approaches e.g. time series, ANN, soft computing approach Clustering and Association Rules, etc. can be also incorporated. The outcome of the predictive data mining technique on the same dataset reveals that of the Decision Tree outperforms and sometime of the Bayesian classification is having similar accuracy as of the decision tree but other predictive methods like as KNN, Neural Networks, Classification of the based on clustering are not performing well.

The conclusion is that the accuracy of the Decision Tree and the Bayesian Classification further improving that after applying the genetic algorithm to reduce that the actual data size to get the optimal subset of the attribute sufficient of the heart disease prediction. The proposed work can be further enhanced and expanded to the automation of Heart disease prediction. Real data from the Health care organizations and agencies are needs to be collected and all the available techniques will be compared to the optimum accuracy.

Reference

[1]. "Applications of Data Mining Techniques" K.Srinivas, B.Kavihta Rani Dr. A.GovrdhaninHealt.

[2]. "PREDICTION SYSTEM FOR HEART DISEASE USING NAIVE BAYES"Shadab Adam Pattekari and AssamPraveen.

[3]. "Decision Support in Heart Disease Prediction System using Naive Bayes" Mrs. G.Subbalakshmi, Mr. K. Ramesh, Mr. M. ChinnaRao.

[4]. "Effective heart disease prediction system using data mining techniques"Poornima Singh, Sanjay Singh and Gayatri S Pandi-Jain.

[5]. Miller, A., B. Blott and T. Hames, Review of neural network applications in medical imaging and signal processing. Med. Biol. Engg. Pedro Domingo comp.

[6]. "Hospitalization for Heart Attack, Stroke, or Congestive Heart Failure a among Persons with Diabetes", Special report: 2001 – 2003, New Mexico.

[7]. Rumelhart, D.E., McClelland, J.L., and the PDF Research Group (1986), Parallel Distributed Processing, MA: MIT Press, Cambridge. 1994.

[8]. Heckerman, D., A Tutorial on Learning with Bayesian Networks. 1995, Microsoft Research.

[9]. Neapolitan, R., Learning Bayesian Networks. 2004, London: Pearson Printice Hall.

[10]. Krishnapuram, B., et al., A Bayesian approach to joint feature selection and classifier design. Pattern Analysis and Machine Intelligence, IEEE Transactions on, 2004. 6(9): p. 1105-1111.

[11]. ShantakumarB.Patil, Y.S.Kumaraswamy, Intelligent and Effective Heart Attack Prediction System Using Data Mining and Artificial Neural Network, European Journal of Scientific Research ISSN 1450- 216X Vol.31 No.4 (2009), pp.642-656 © EuroJournals Publishing, Inc. 2009.