ISSN: 2581-3404 (Online)
International Journal of Innovative Research in Technology and Management, Vol-4, Issue-6, 2020.

Prediction of Heart Diseases for Smart Health Care System Using Support Vector Machine

Fuzaila Siddiqui[1], Prof. Syed Imran Ahmed Qadri[2]

[1]M. Tech Scholar, Department of CSE, ASCT, Bhopal (India)

[2]Assistant Professor, Department of CSE, ASCT, Bhopal (India)

ABSTRACT

Medical diagnosis is considered an art regardless of all standardization efforts made, which is greatly due to the fact that medical diagnosis necessitates an expertise in coping with uncertainty simply not found in today's computing machinery. The researchers are encouraged by the advancement in computer technology and machine learning techniques to develop software to assist doctors in making decision without necessitating the direct consultation with the specialists. In this paper, application of data mining such as classification techniques in typical heart disease and diabetes diseases diagnosis has been investigated. The support vector machine with sequential minimal optimization algorithm is applied to India based patients' data set. Results obtained show that support vector machine can be successfully used for diagnosing heart disease and diabetes diseases. The role of effective diagnosis and the advantages of data training on machine learning based automatic medical diagnosis system are suggested by the outcomes.

Keywords: Data mining, Healthcare, Supervised techniques, support vector machines.

INTRODUCTION

The application of machine learning methods in medical field is the subject of considerable ongoing research, which mainly concentrates on modeling some of the human actions or thinking processes and recognizing diseases from a variety of input sources (e.g. patient's data cardiograms, CAT/MRI/Ultrasound scans, photomicrographs, etc.).

Other application areas are knowledge discovery [3] and biomedical systems, which include genetics and DNA analysis [1,4]. Heart diseases have emerged as the number one killer in both urban and rural areas in most of the countries. As of 2010, it is the leading cause of death in the U.S., England and Canada, accounting for 25.4% of the total deaths in the United States. Similar situation is found rest of the countries all over the world [7].

In case of heart disease time is very crucial to get correct diagnosis in early stage. Patient having chest pain complaint may undergo unnecessary treatment or admitted in the hospital. In most of the developing countries specialists are not widely available for the diagnosis. Hence, such automated system can help to medical community to assist doctor for the accurate diagnosis well in advance[5,15]. In a medical diagnosis problem, what is needed is a set of examples or attributes that are representative of all the variations of the disease. The examples need to be selected very carefully if the system is to perform reliably and efficiently. The fact that there is no need to provide a specific algorithm on how to identify the disease, presents a major advantage over the application of machine learning methods to this type of problems. However, development of artificial intelligence systems for medical decision making problems is not a trivial task. Difficulties include the acquisition, collection and organization of the data that will be used for training the system. This becomes a major problem especially when the system requires large data sets over long periods of time, which in most cases are not

available due to the lack of an efficient recording system.

The above mentioned problems or the existing procedures involved in the medical task may not be the only factors affecting the design of a Decision Support System(DSS). The design may also be influenced by the desired performance on one or more specific classes of the problem instead of the overall performance. This is usual in most medical tasks as a different degree of significance may be required for the system's performance on each class. For example, in a heart disease diagnosis task, it is necessary for the accuracy on healthy patients to be as high as possible, as a misclassification in this category may result in a healthy patient going under treatment for no reason. The balance of the system's performance between different classes could vary and is largely dependent on the medical problem itself and the collected data. In addition, in most of the countries, insufficient numbers of medical specialist have increased the mortality of patients suffering from various diseases. It is observed that, in many cases due to wrong diagnosis or trial/error procedure for diagnosis leads to patient health compromise. The computer programs or machine learning techniques can be used to reduce the mortality rate, improve the accuracy in disease diagnosis and mainly reduce the diagnosis time. The advancement in computer technology and communication encourages health-care providers to work using the Internet or Telemedicine technology [2.5].

The rest of this paper is organized as follows in the first section we describe an introduction of about the smart health care with data mining and classification techniques. In section II we discuss about the heart diseases diagnosis system, in section III we discuss about the proposed methodology. In section IV we discussed about the support vector machine based implementation and results, and finally in section V we conclude the our research works.

II HEART DISEASE DIAGNOSIS

In order to provide physicians with both structured questions and structured responses within medical domains of specialized knowledge or experience medical expert systems have been developed [21]. The advice of one or more medical experts, who also suggest the optimal questions to be considered, and provide the most accurate conclusions to be drawn from the answers the physician chooses, is used to embody the structure in the DSS. Table I show the Heart disease symptoms and tests which are usually observed during the diagnosis. Based on medical records 214 instances in total are selected for the automated diagnosis. Dataset consists of 19 different attributes with four class distribution: 0-Myalgia, 1-Myocardial Infarction (MI), 2-Ischemic Heart Disease (IHD), 3- Unstable Angina (UA).

TABLE I
HEART DISEASE ATTRIBUTES

Symptoms	Test
Chest pain types (Left or Right side), Arm pain, Backache, Sweating, Breathlessness, Addiction, Diabetic, MAP, Pulse rate.	ECG: ST elevation, ST depression, T elevation, T depression, Q waves, BSL, CK-MB test

For this study binary classification problem is considered for Heart disease diagnosis, in which 78 instances are belongs to 0 i.e. Myalgia (Normal), and 139 are considered as 1 i.e. Patient having Heart disease. A trained Radial Basis Function Neural Network (RBF) and Support Vector Machine (SVM) are used to assume the evolution of the biological indicators. Once the patients' personal data is presented along with the results of the tests taken at the onset of the treatment and the postulated code of reaction, the evolution in time of the illness can be specified by the expert system.

III PROPOSED METHODOLOGY

Data mining gives various types of classification algorithm, according to the diversity of data and variety of data. The variety of data induced the problem of classification issue and degreed the performance of classification algorithm. The classification technique gives different types of algorithm such as support vector machine, decision tree, KNN and ensemble based classifier. Now a day's ensemble based classifier used for the process of classification. The ensemble based classifier used three types of ensemble technique, bagging, boosting and random forest. The all three technique of ensemble proceed in different manners. Some authors also used clustering technique for the process of ensemble classifier. The feature attribute plays major role in classification technique, the diverse feature creates many problems related to the process of classification such as outlier, and boundary value and core point, without elimination of these problems the classification ratio can't improve.

Here we proposed a comparative classification approach for the healthcare data set, all the dataset values are taken from the UCI machine learning repository and simulated with the matlab software. The comparative classifications method are k nearest neighbor classification and the support vector machine, our proposed method gives better results than the existing techniques, The evaluated performance parameters are accuracy, precision and recall.

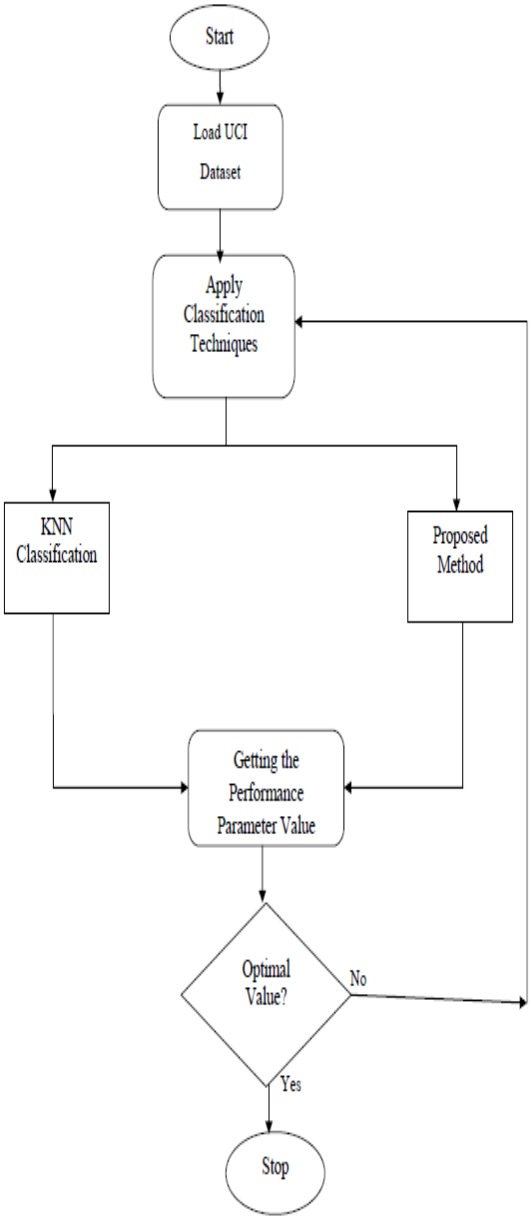

Fig 1: Proposed method flow graph.

IV SUPPORT VECTOR MACHINES

Classification algorithms are widely used in various medical applications. Classification aims to build an effective model for predicting class labels of unknown data. The model is built on the training data, which consists of data points chosen from input data space and their class labels. A Support Vector Machine (SVM) separates the data

ISSN: 2581-3404 (Online)
International Journal of Innovative Research in Technology and Management, Vol-4, Issue-6, 2020.

into two categories of performing classification and constructing an N-dimensional hyper plane. These models are closely related to classical multilayer perceptron neural networks.

A support vector machine constructs a hyper plane or set of hyper planes in a high- or infinite-dimensional space. A good separation is achieved by the hyper plane that has the largest distance to the nearest training data point of any class (so-called functional margin), since in general the larger the margin the lower the generalization error of the classifier.

There are an alternative training method for polynomial, radial basis function and multi-layer perceptron classifiers in which the weights of the network are found by solving a quadratic programming problem with linear constraints, rather than by solving a non-convex, the unconstrained minimization problem as in standard neural network training.

In the SVM literature, a predictor variable which is called an attribute and a transformed attribute that is used to define the hyper plane is called a feature [13]. Here, choosing the most suitable representation can be taken as feature selection. A set of features that describes one case is called a vector. The goal of this modeling is to find the optimal hyperplane which separates clusters of vector in such a way those cases with one category of the target Variable are on one side of the plane and cases with the other category are on the other side of the plane. The vectors near to the hyper plane are the support vectors as in below figure.

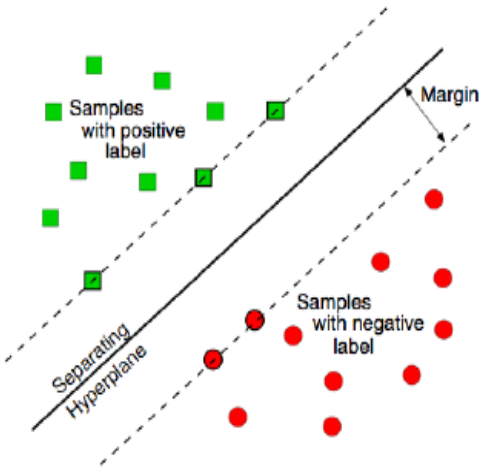

Fig 2: Maximum margin, the vectors on the dashed line are the support vectors.

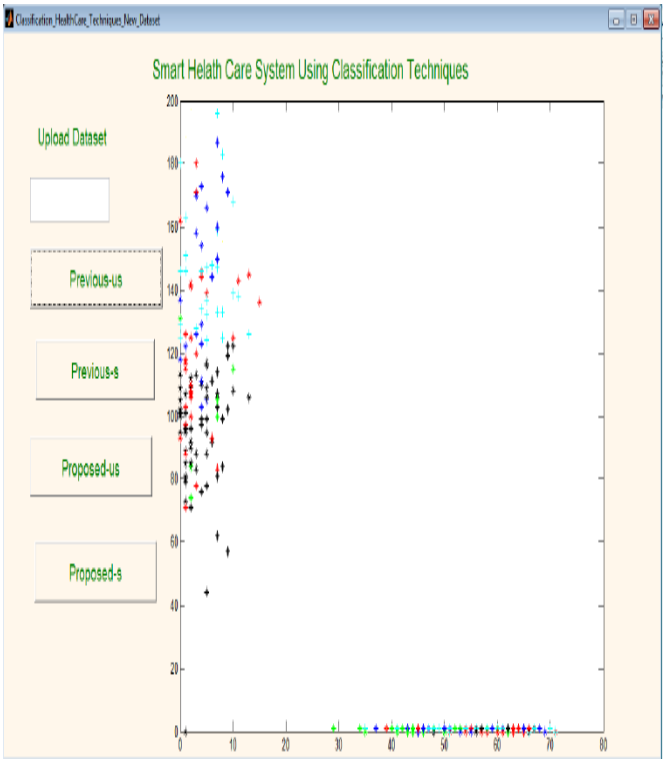

Fig 3: The above figure shows that the experimental classification works with using the previous unsupervised techniques.

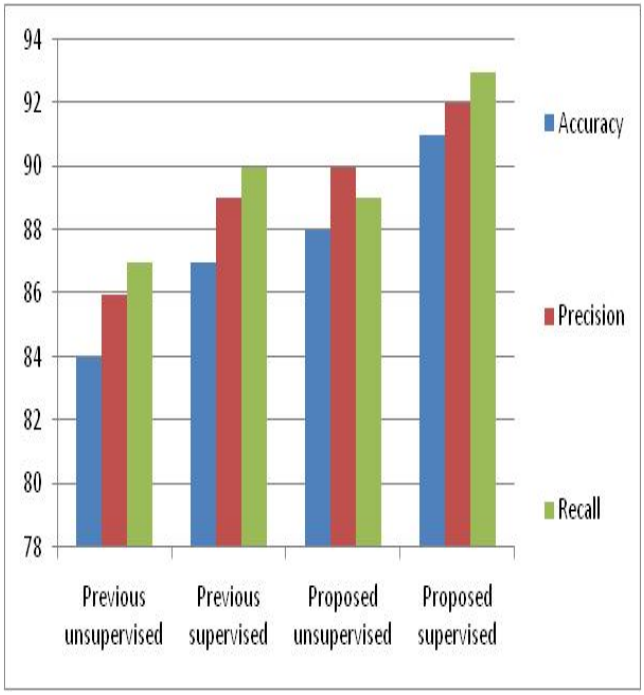

Fig 4: The above figure show that the comparative result analysis for the heart disease dataset.

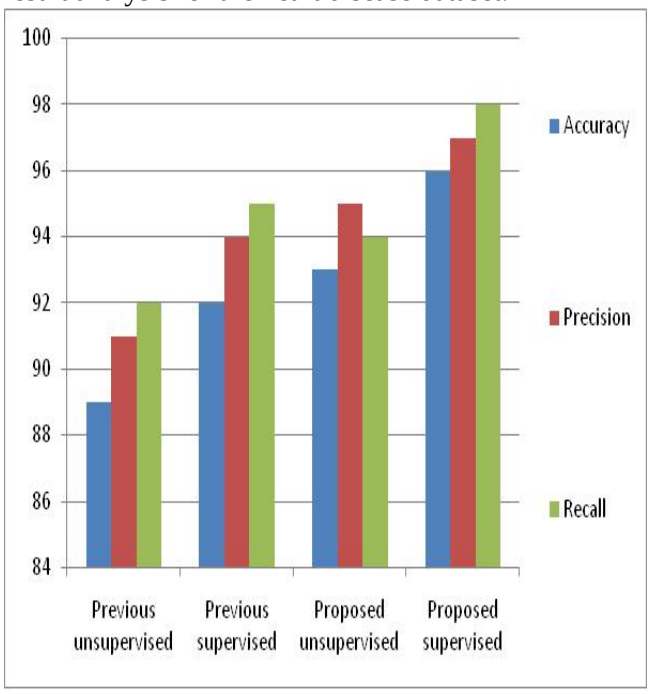

Fig 5: The above figure show that the comparative result analysis for the diabetes disease dataset.

V CONCLUSION

The main purpose of pattern recognition in the field of health care diseases diagnosis is to solve the pattern classification dilemma where a pre-described set of input features is used to determine if a patient has a particular disorder or not. Support vector machine has become an increasingly popular tool for machine learning tasks involving classification, regression or novelty detection. Training a support vector machine requires the solution of a very large quadratic programming problem. Traditional optimization methods cannot be directly applied due to memory restrictions. Up to now, several approaches exist for circumventing the above shortcomings and work well. In this paper we present the support vector machine based model for the diseases diagnosis system and our model show better results than the previous techniques.

REFERENCES:-

[1] Subasish Mohapatra, Prashanta Kumar Patra, Subhadarshini Mohanty, Bhagyashree Pati, "Smart Health Care System using Data Mining", International Conference on Information Technology (ICIT), IEEE 2018, pp 44-49.

[2] Ramin Ghorbani, Rouzbeh Ghousi, "Predictive data mining approaches in medical diagnosis: A review of some diseases prediction", International Journal of Data and Network Science, 2019, pp 44-70.

[3] Jaswinder Singh, Sandeep Sharma, " Prediction of Cervical Cancer Using Machine Learning Techniques", International Journal of Applied Engineering Research, 2019, pp. 2570-2577.

[4] Hina Firdaus, Syed Imtiyaz Hassan, Harleen Kaur, "A Comparative Survey of Machine Learning and Meta-Heuristic Optimization Algorithms for Sustainable and Smart Healthcare", Vol. 11, No. 4, December 2018, pp. 1 – 17.

[5] Beant Kaur, Williamjeet Singh, "Review on Heart Disease Prediction System using Data

Mining Techniques", International Journal on Recent and Innovation Trends in Computing and Communication, 2014, pp 3003-3008.

[6] K. Lakshmi, D. Iyajaz Ahmed, G. Siva Kumar, " A Smart Clinical Decision Support System to Predict diabetes Disease Using Classification Techniques", IJSRSET, vol-4, 2018, pp 1520-1522.

[7] Shubham Banka, Isha Madan, S.S. Saranya, " Smart Healthcare Monitoring using IoT", International Journal of Applied Engineering Research, 2018, pp. 11984-11989.

[8] Dmitry Korzun, Alexander Meigal, "Multi-Source Data Sensing in Mobile Personalized Healthcare Systems: Semantic Linking and Data Mining", IEEE, 2018, pp1-6.

[9] Mehmet Akif CIFCI, Sadiq Hussain, "Data Mining Usage and Applications in Health Services", International Journal on Informatics Visualization, 2018, pp 225-231.

[10] Noble Mary Juliet Augusstine, Suba Rani Narayana Samy, "Smart healthcare monitoring system using support vector machine", Australian Journal of Science and Technology, 2018, pp 108-111.

[11] T. Senthilkumar, B. Manikandan, M. Ramya Devi, S. Lokesh, "Technologies Enduring in Internet of Medical Things (IoMT) for Smart Healthcare System", International Journal of Scientific Research in Computer Science, Engineering and Information Technology, 2018, pp 566-572.

[12] Saima Anwar Lashari, Rosziati Ibrahim, Norhalina Senan, N. S. A. M. Taujuddin, " Application of Data Mining Techniques for Medical Data Classification: A Review MATEC Web of Conferences, 2018, pp 1-6.

[13] Rahul Joshi, Minyechil Alehegn " Analysis and prediction of diabetes diseases using machine learning algorithm: Ensemble approach" International Research Journal of Engineering and Technologyol-4, 2017. Pp 426-435.

[14] S.Poongothai, C.Dharuman and P.Venkatesan "A Comparison of Fuzzy Genetic and Neuro Genetic Hybrid Algorithm for the Classification of Diabetes Disease" International Journal of Pure and Applied Mathematics Volume 113, 2017, Pp 208-216.

[15] M.Akhil jabbar , Priti Chandra and B.L Deekshatulu "Prediction of Risk Score for Heart Disease using Associative Classification and Hybrid Feature Subset Selection", IEEE, 2012, Pp 628-634.

[16] Fabio Mendoza Palechor, Alexis De la Hoz Manotas, Paola Ariza Colpas , Jorge Sepulveda Ojeda, Roberto Morales Ortega, Marlon Piñeres Melo " Cardiovascular Disease Analysis Using Supervised and Unsupervised Data Mining Techniques" Volume 12, Number 2, February 2017. Pp 81-90.

[17] Syed Umar Amin, Kavita Agarwal, Rizwan Beg "Genetic Neural Network Based Data Mining in Prediction of Heart Disease Using Risk Factors", IEEE, 2013, Pp 1227-1231.

[18] B V Sumana, T. Santhanam "Prediction of diseases by Cascading Clustering and Classification", ICAECC, 2014, Pp 1-8.

[19] V Krishnaiah, M Srinivas, Dr.G Narsimha, Dr.N Subhash Chandra "Diagnosis of Heart Disease Patients Using Fuzzy Classification Technique", IEEE, 2010, Pp 1-7.

[20] Sellappan Palaniappan ,Rafiah Awang "Intelligent Heart Disease Prediction System Using Data Mining Techniques", IEEE, 2008, Pp 108-115.

[21] Zhao, S., Heittola, T., & Virtanen, T. (2017). Active Learning for Sound Event Classification by Clustering Unlabeled Data. In 2017 IEEE

International Conference on Acoustics, Speech and Signal Processing, 2017.

[22] R. Elankavi, R. Kalaiprasath, R. Udayakumar, "A FAST CLUSTERING ALGORITHM FOR HIGH-DIMENSIONAL DATA", International Journal of Civil Engineering and Technology, 2017, pp 1220–1227.

[23] Liling Li, Sharad Shrestha, Gongzhu Hu, "Analysis of Road Traffic Fatal Accidents Using Data Mining Techniques", IEEE, 2017. pp 363-369.

[24] ASHISH DUTT, MAIZATUL AKMAR ISMAIL, AND TUTUT HERAWAN, "A Systematic Review on Educational Data Mining", IEEE, 2017. pp 15991-1604.

[25] Mohammed G. Ahamad, Mohammed F. Ahmed and Mohammed Y. Uddin, "Clustering as Data Mining Technique in Risk factors Analysis of Diabetes, Hypertension and Obesity", European Journal of Engineering Research and Science, 2016. pp 88-94.

[26] Dong Li, Cheng Tang, Chunlei Xia, Hua Zhang, "Acoustic mapping and classification of benthic habitat using unsupervised learning in artificial reef water", Estuarine, Coastal and Shelf Science, 2017. pp 11-21.

[27] YANFANG YE, TAO LI, DONALD ADJEROH, S. SITHARAMA IYENGAR, "A Survey on Malware Detection Using Data Mining Techniques", ACM Computing Surveys, 2017. pp 41-80.

[28] Sunith Bandaru, Amos H. C. Ng, Kalyanmoy Deb, "Data Mining Methods for Knowledge Discovery in Multi-Objective Optimization", 2017. pp 1-40.

[29] R. Sukanya, K. Prabha, " Comparative Analysis for Prediction of Rainfall using Data Mining Techniques with Artificial Neural Network", International Journal of Computer Sciences and Engineering, 2017. pp 288-293.

[30] Naresh Kumar Nagwani, Aakanksha Sharaff, "SMS spam filtering and thread identification using bi-level text classification and clustering techniques", Journal of Information Science 2017, pp 75–87.