Health Care System Using Data Mining Techniques: A Review

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

The traditional style of healthcare is based on visiting a hospital or clinic to meet a doctor face-to-face and to pass through certain monitoring and analysis procedures using the professional medical equipment and direct expertise. This form of healthcare operation is expensive. Health policies and decisions are based on reliable, up-to-date and accurate data which should be appropriate and effective. In this paper we review the various data mining techniques for the health care industries, also using some smart techniques such as the internet of things, evolutionary techniques and the other methods.

Keywords: Data mining, Healthcare, Supervised techniques, Internet of things, Smart healthcare.

INTRODUCTION

Smart Health Care Prediction using Data Mining is a new powerful technology which is of high interest in computer world. It is a sub field of computer science that uses already existing data in different database to transform it to new researches and result. The actual task is to extract data by automatic or semi-automatic means. The different parameters included in data mining include clustering, forecasting, path analysis and predictive analysis. Data mining is a new discipline with different applications known as one of the ten leading sciences influencing technology. Wherever the data exists, data mining is also meaningful, for instance: Market Basket Analysis, Education, Manufacturing Engineering, Customer Relationship Management, Fraud Detection, Intrusion Detection, Lie Detection, Customer Segmentation, Financial-Banking, Corporate Surveillance, Research Analysis, Criminal Investigation, Tele-communication and Healthcare [2].

With the growing researching the field of health informatics a lot of data is being produced. The analysis of such a large amount of data is very hard and requires excessive knowledge. Smart health care applies data mining techniques for health diagnosis. Data mining refers to extracting meaning full information from the different huge amount of dataset. It is the process of determining the unseen finding pattern and knowledge from the massive amount of data set [1].

Today, the healthcare industry generates large amounts of complex data on patients, hospital resources, diagnosis of diseases, electronic patient records and medical devices. More copious amounts of data are an essential resource for data mining. There is a vast potential in healthcare data mining applications, and some of the most critical applications in healthcare data mining are prediction and diagnosis, treatment effectiveness, healthcare management, fraud and abuse, customer relationship management, and the medical device industry. Choosing the wrong treatment for patients will not only waste time and money but also can cause adverse effects such as the death of patients. Therefore, a method for diagnosing and selecting the appropriate treatment is essential for patients. Data mining can help with the prediction and determination of the diseases in this area [2].

Data mining is significant research doings in the field of medical sciences since there is a requirement of well organized methodologies for analyzing, predict and detecting diseases. To detect and predict diseases Data mining applications are used for the management of healthcare, health information, patient care system, etc. It also plays a major role in analyzing survivability of a disease. Data mining classification techniques play a vital role in healthcare domain by classifying the patient dataset. Data mining classification technique is used to analyses and predicts many diseases.

Clinical Decision Support System is a healthcare system, which is intended to assist physicians and other health professionals on decision making tasks. In healthcare sector, the term information mining can mean to analyze the clinical information to predict patient's health status. So discovering interesting pattern from healthcare data, different data mining techniques are applied with statistical analysis, machine learning and database technology. It can be also defined as "A computer system that uses two or more patient data to generate case specific or encounter specific advice." Most CDSS (Clinical Decision Support System) comprises of three parts, the knowledge base, inference engine, and mechanism to communicate [6].

It can be said that there has been vast potential for data mining applications in healthcare [9]. Commonly, all of these can be seen as normal healthcare applications but still the usage and exploiting the data is essential. The usage of data mining is mostly for detection of fraud and abuse. Nonetheless, specialized medical data mining, such as predictive medicine is the main focus. KDD is often referred to as data mining, which aims to discover useful information from a large volume of data collections. Databases are now expressed in peta-bytes. This large volume lies in hidden information that is of strategic importance within the database. But the most important question is how to bring out the important information contained in such a large volume of data. The most up-to-date answer to this important question is data mining, which increases both revenue and costs. data mining is a process that explores patterns and relationships in data with the use of many analysis tools and uses them to make valid estimates. Data mining, by its simplest definition, automatically determines the associated patterns in databases. Assuming data mining as magic is not sensible.

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 techniques. In section II we discuss about the classification of various data mining techniques, in section III we discuss about the related work. In section IV we discussed about the proposed method and flow graph, and finally in section V we conclude and discuss the future scope.

II DATA MINING CLASSIFICATION

In data mining, categorization is formulated to make a forecast of the memberships in a group for data instances. This process utilizes complex analysis of data to determine data connections in huge datasets. Due to its complex features, medical databases provide complications for pattern extortion [12]. There are two approaches to data mining: statistical and machine learning algorithms. The processes in data mining are classified into descriptive and predictive see below figure. Descriptive mining tasks provide the general data properties in the database. For Predictive mining tasks, inference is made on the data for predictions whereby forecast is made on explicit values based on patterns identified by known results. Descriptive data mining, without having any predefined target, provides characteristics and descriptions for the data set. Data mining techniques are effective and predictive for future patterns because: a) it is user friendly and prediction is based on past circumstances b) it operates by learning from past data c) data from numerous resources is managed and only required data is extracted d) models are easily updated by relearning, past information and

change in trends. These are what make it reliable and practical in the medical image categorization.
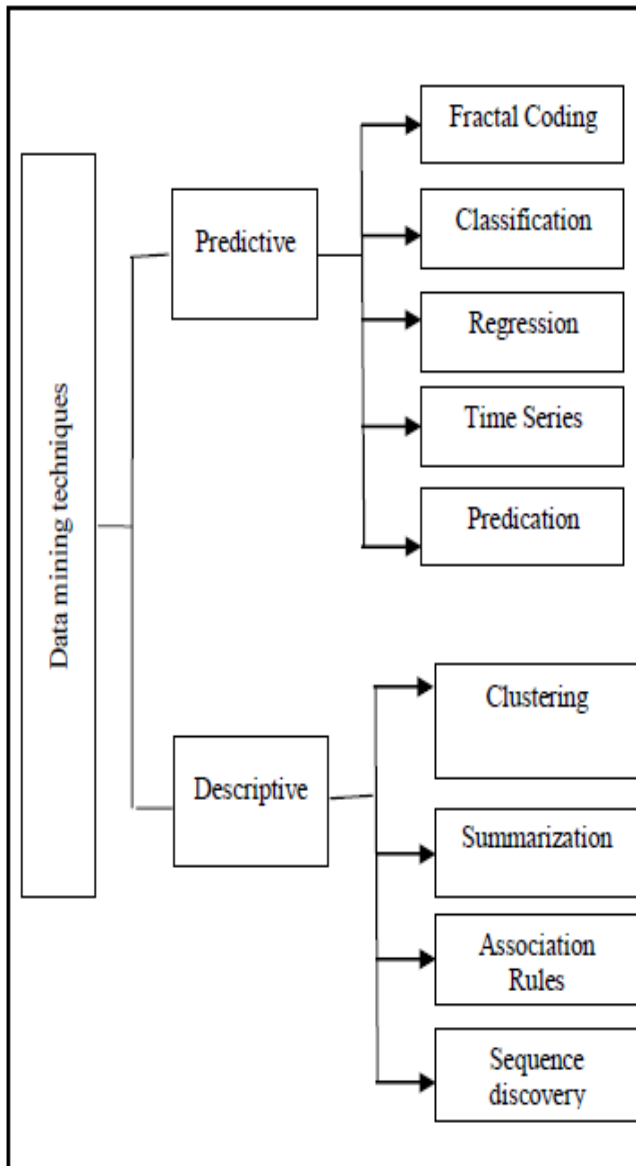
Figure 1: Data mining techniques [12].

III LITERATURE SURVEY

Subasish Mohapatra, Prashanta Kumar Patra, Subhadarshini Mohanty, Bhagyashree Pati et al. [1] The objective of this study is to observe the application of data mining, classification algorithm and feature selection methods to analyze and predict Chronic kidney disease, Heart disease and Liver disease. It can be very well analysis and predict by data mining and feature selection methods. From the experimental result, we observe that data mining, classification and feature selection methods have been used to identify, analysis and predict for chronic kidney, heart and liver diseases. Many researchers have been conducted different data mining, classification algorithm like ANN, Naïve Bays, SVM, Decision tree (J48, C4.5) and feature selection to improve the performance accuracy of the algorithm. Healthcare institutions that use data mining applications have the possibility to predict future requests, needs, desires, and conditions of the patients and to make adequate and optimal decisions about their treatments. With the future development of information communication technologies, data mining will achieve its full potential in the discovery of knowledge hidden in the medical data. Ramin Ghorbani, Rouzbeh Ghousi et al. [2] This paper reviewed the predictive data mining approaches in heart disease, breast cancer, and diabetes diagnosis. The obtained results reveal that a significant number of studies have used classification technique. Also, researchers have achieved better prediction accuracy results with hybrid and ensemble models. Furthermore, in most research, the performance of different data mining models is compared to each other. Comparison of the different clustering methods has appeared that K-Means clustering is the most common clustering method. Additionally, the Decision Tree algorithm, Bayesian Network, and Neural Network are three widely used classification methods based on the comparison of the different classification methods. Jaswinder Singh, Sandeep Sharma et al. [3] This paper proposes a prediction model for diagnosing the correct stage of infection. The proposed model utilizes UCI data repository and machine learning classifiers for prediction. The data from sensors undergoes pre-processing where the feature extraction and validations updates the repository. Their proposed model considers ten features of cervical cancer relating to four stages. The pre-processed data is then made available to physician for verification followed by training of machine

learning classifiers. Hina Firdaus, Syed Imtiyaz Hassan, Harleen Kaur et al. [4] The study has raised various questions regarding the best machine learning technique on different disease dataset and solutions to overcome the problem of optimizing the feature selection for better performance. A comparison of the various algorithms is presented in tabular form based on two categories, namely the use of ML and also the use of ML with meta-heuristic algorithms in disease dataset. A model proposed for future work, which uses meta-heuristic algorithms for feature selections, appears to be better for any dataset. Beant Kaur, Williamjeet Singh et al. [5] The objective of their work is to provide a study of different data mining techniques that can be employed in automated heart disease prediction systems. Various techniques and data mining classifiers are defined in this work which has emerged in recent years for efficient and effective heart disease diagnosis. The analysis shows that different technologies are used in all the papers with taking different number of attributes. So, different technologies used shown the different accuracy to each other. K. Lakshmi, D.Iyajaz Ahmed, G. Siva Kumar et al. [6] In this paper they presented data mining algorithms clustering and classification algorithms like Decision Tree and K-Nearest Neighbor (KNN) Algorithms as supervised classification model for diabetes disease prediction. As their proposed system uses the Decision Tree and K-Nearest Neighbor (KNN) Algorithms as supervised classification model for diabetes disease because it is a great threat to human life worldwide finally our proposed system will reduce the time and cost of diagnoses. Shubham Banka, Isha Madan, S.S. Saranya et al. [7] The proposed system here consists of various medical devices such as sensors and web based or mobile based applications which communicate via network connected devices and helps to monitor and record patients' health data and medical information. The proposed outcome of the paper is to build a system to provide world-class medical aid to the patients even in the remotest areas with no hospitals in their areas by connecting over the internet and grasping information through about

their health status via the wearable devices provided in the kit using a raspberry pi microcontroller which would be able to record the patient's heart rate, blood pressure. The system would be smart to intimate the patient's family members and their doctor about the patient's current health status and full medical information in case any medical emergency arises. Dmitry Korzun, Alexander Meigal et al. [8] In this paper author introduces an approach to collecting and mining health-related information on the patient based on sensed data from various medical devices as well as from other digitally enabled sources. The regularly sensed data are semantically linked thus creating an additional information space semantic layer. On the latter, a linked knowledge rich structure semantic network is maintained and used to construct mobile services. The use of various medical devices and other data sources makes it possible to remotely monitor patients' vital physiological parameters and other important health-related events. It includes sensing the context of the physical environment, which is then coupled with the health state of the patient. Several patients and interested people can be virtually integrated into a group. Mehmet Akif CIFCI, Sadiq Hussain et al. [9] In this study, it is aimed to analyze data mining applications and usages in healthcare to display an infrastructure for data mining and to give a new perspective to the health professionals in terms of decision-making processes by presenting examples about the use of data mining in the health sector. As understood it is vital to use data mining methods in the health sector to have better solutions to the arising problems when compared to the traditional methods. Noble Mary Juliet Augusstine, Suba Rani Narayana Samy et al. [10] This paper adopted an innovative health monitoring system using the internet of things for accessing the patient's medical parameters in the local and remote area. The goal of this proposal is to transmit an emergency message to caretaker when the health condition goes critical. A cloud server records the data from the temperature sensor, and heartbeat sensor which is connected to the patient; the data are analyzed using support vector machine

learning algorithms to detect the abnormal conditions, issues an emergency message to the caretaker of the patient through a mobile application, and sends an alert message to the nearest hospital.

IV 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.
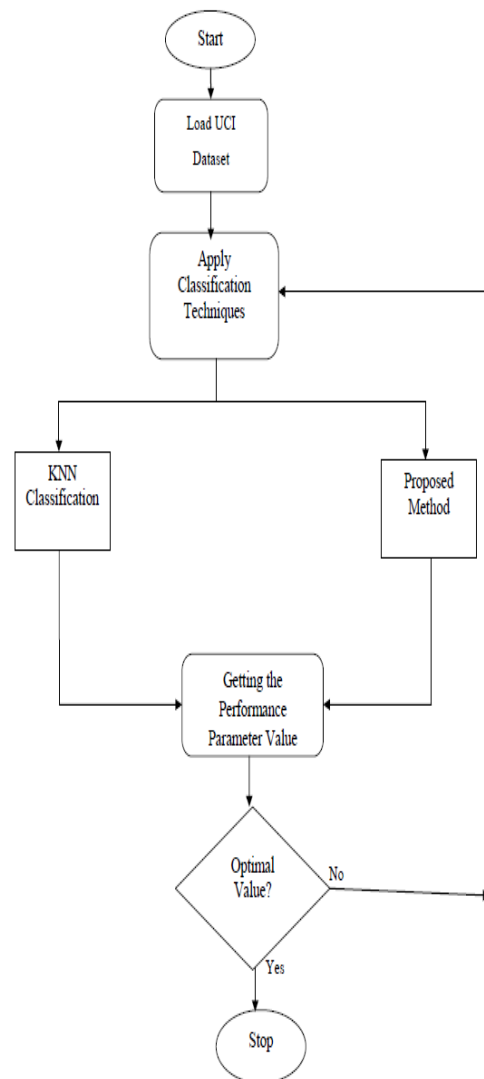
Figure 2: Proposed method flow graph.

SUPPORT VECTOR MACHINES

In this paper for the healthcare data classification we used the support vector machine, Support vector Machine is binary classifier, the performance of classification of support vector machine is high in compression of another binary classifier such as decision tree, KNN and other classifier. It is a novel machine learning method based on statistical learning theory developed by V.N.Vapnik, and it has been successfully applied to numerous classification and pattern recognition problems such as text categorization, data

classification for the various applications, image recognition and bioinformatics.

The concept of SVM is to transform the input vectors to a higher dimensional space Z by a nonlinear transform, and then an optical hyperplane which separates the data can be found. This hyperplane should have the best generalization capability. As shown in below figure, the black dots and the white dots are the training dataset which belong to two classes. The Plane H series are the hyperplanes to separate the two classes. The optical plane H is found by maximizing the margin value $2/\|w\|$.

Hyperplanes H_1 and H_2 are the planes on the border of each class and also parallel to the optical hyperplane H. The data located on H_1 and H_2 are called support vectors.
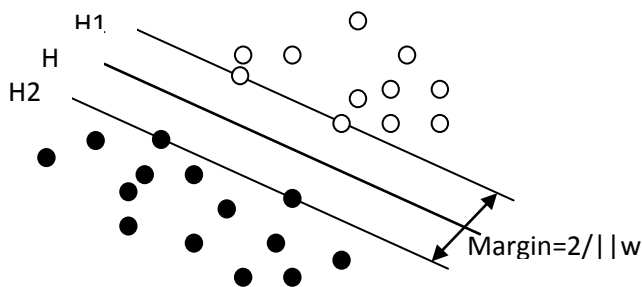
Figure 3: The SVM classifications.

V CONCLUSION

In our presented system the conventional disease diagnosis system uses the perception and experience of doctor without using the complex clinical data. In medical science, doctor's facilities introduced different data frameworks with a lot of information to manage medical insurance and patient information but unfortunately, data are not mined to discover hidden information for effective decision. In this review discuss with various techniques in the healthcare, we proposed the solution using improve efficiency, better solution from the existing techniques and accuracy.

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.