International Journal of Scientific Research in Computer Science, Engineering and Information Technology

ISSN : 2456-3307 OPEN ACCESS

Available Online at : www.ijsrcseit.com

doi : https://doi.org/10.32628/IJSRCSEIT

A Significance of Data Mining in the field of Healthcare Sector

[1]Swati C. Tawalare*, [2]Dr. S. Y. Amdani

[1*]Research Scholar, Department of Computer Science & Engineering, Babasaheb Naik College of Engineering, Pusad, India

[2] Associate Professors, Department of Computer Science & Engineering, Babasaheb Naik College of Engineering, Pusad, India

ARTICLEINFO

Article History:

Accepted: 10 Feb 2024

Published: 24 Feb 2024

Publication Issue

Volume 10, Issue 1

January-February-2024

Page Number

217-230

ABSTRACT

Currently, there's huge quantum of data being collected and stored in databases far and wide across the globe running into terabytes of data and the tendency is to keep adding time after time. Moment, the healthcare assiduity which is one of the largest diligences throughout the world includes medical diligence having the large quantities of health- related and medical-affiliated data. It also includes thousands of hospitals, conventions and other types of installations that give primary, secondary and tertiary situations of care. The delivery of healthcare service is the most visible part of any healthcare system, both to druggies and the general public. Accurate, early and error-free opinion and treatment given to cases has been a major issue stressed in medical service currently. Quality service in health care field implies diagnosing cases rightly and administering treatments that are effective. To achieve a correct and cost-effective treatment, a system can be developed to fulfill the task.

Keywords : Data Mining, Healthcare Sector, Quality Service, MRI, ECG

I. INTRODUCTION

Moment, healthcare systems use sanitarium information systems to manage their healthcare or case data. These systems induce huge quantities of medical data which may contain electronic case records with their computer- readable entries like glamorous Resonance Imaging (MRIs), signals like Electrocardiography (ECG), clinical information like blood sugar, blood pressure, cholesterol situations, etc. as well as the croaker 's interpretation. These data may comprise a lot of inestimable information and knowledge hidden in similar databases, which can be used to support effective clinical decision- timber. But the discovery of retired patterns and knowledge from similar databases frequently goes unexploited. This explosive growth of data requires an robotic way to prize what's called "nuggets of knowledge" from the large sets of data for the opinion and treatment of complaint from the database and this decreasingly becomes necessary. Conventionally, the data is anatomized manually using retrospective tools typical

of decision support systems and numerous retired and potentially useful connections may not be honored by the critic because it lies outside their prospects. The exploration purpose is to find druthers to the result of complex medical opinion in discovery of heart complaint where mortal knowledge is restrained in a general fashion. Successful operation exemplifications shown preliminarily reveal that mortal individual capabilities are significantly worse than computer-backed individual systems.

Advanced data mining ways in drug can deal with this problem and remedy the situation. Data mining can ameliorate the operation position of sanitarium information, saving time and cost. Using data mining ways, we can prize intriguing knowledge and discrepancies. The discovered knowledge can also be applied to the medical data to increase the working effectiveness and help medical interpreters in their decision- timber. This provides near- endless openings for symptom trend discovery, before discovery of illness, and DNA trend analysis, etc. therefore adding effectiveness in healthcare.

While developed countries like the United States have their systems automated formerly and Intelligent Systems are decreasingly being stationed in drug and healthcare, to virtually prop the busy clinician and to ameliorate the quality of patient care, developing countries like Nigeria still make use of traditional statistical methods., But with the help of data mining, we can also join the moving trend.

According to [1], Data Mining consists of five major rudiments viz

• prize, transfigure and load data onto data storehouse system.
• Store and manage the data in multidimensional database.
• give data access to judges.
• dissect the data by operation software.
• Present the data in useful format.

Steps in a Data Mining Process are

1) Data integration first of all, data is collected from and integrated from all different sources.
2) Data selection data is named on behalf of some criteria.
3) Data preprocessing data collected may contain crimes, inconsistencies that need to be removed.
4) Data metamorphosis the data indeed after preprocessing may not be ready for mining so it needs to be converted into a form applicable for mining.
5) Knowledge discovery meaningful patterns are uprooted from large data. At last, meaningful patterns help in decision- timber [1].

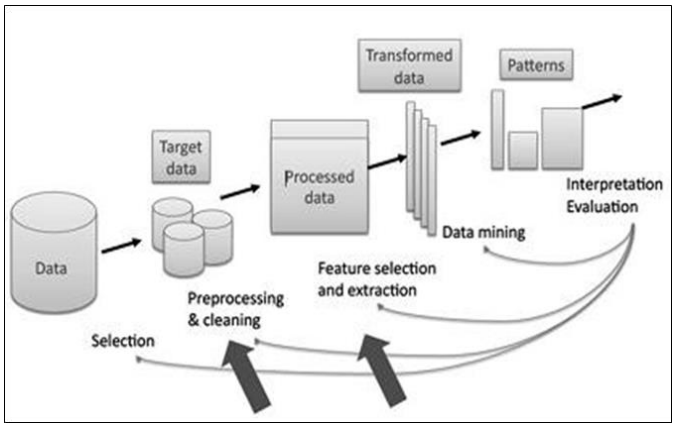

Figure 1. Data Mining Process

Data mining is extensively being used in working multitudinous exploration problems. Data mining is getting decreasingly popular and essential in healthcare sector [2]. Data mining operations can give advantage to all parties involved in the healthcare assiduity [3] [4]. For illustration, data mining can help healthcare insurer descry fraud and abuse, croakers identify effective treatments and stylish practices and cases admit better and further affordable healthcare services [5]. Huge quanta of data generated by healthcare deals are too complex and substantial to be reused and dissect by traditional styles. Data booby-trapping provides the methodology and technology to transfigure these quantities of data into useful information for decision- making in healthcare sector.

1.1 Statement of Problem

Data Mining as an logical process is designed to prize useful data, patterns and trends from a large quantum of data(generally business, medical or request related data) by using ways like clustering, bracket, association and retrogression. The ultimate thing of data mining is prediction. Research statistics have shown that utmost healthcare- related conditions make use of data mining ways that don't produce optimal results.

Below are some of the colorful inconsistencies associated with the use of the wrong data mining ways

1) Not having veritably high delicacy in decision.
2) Deficit of moxie.
3) Difficulties in knowledge upgrade.
4) Time-dependent performance (veritably time-consuming).

Because of these problems, there's necessity to emplace data mining to give the backing medium in opinion procedure. The conclusion is clear humans and their statistical styles cannot ad hoc dissect complex data without crimes. In drug and healthcare where safety is critical, it's important if data mining ways are to be extensively accepted in clinical practice (6). The thing of the process is to take the medical data which contain numerous attributes and determine which bones are actually applicable to the opinion, symptoms and result of heart complaint. Without automatic styles for rooting this information, it's virtually insolvable to booby-trap for them, seeing that we're looking at a veritably huge quantum of data running into terabytes of data.

II. Literature Review

2.1. Conceptual Framework

The healthcare assiduity battles with millions of digitally recorded data and patterns being collected at enormous speed due to the wide operation of important computer bias currently [7]. The data collected are substantially unorganized and haven't been used duly for applicable operations, therefore, assessing new challenges regarding their operation including their modeling, storehouse, and reclamation capabilities. There's frequently intriguing knowledge in the data that isn't readily apparent. The spread of electronic case records, with their computer- readable entire glamorous Resonance Imaging (MRI), signals like ECG(Electrocardiography), clinical information like blood sugar, blood pressure, cholesterol situations, etc. as well as the croaker 's interpretation is opening new possibilities for medical data mining and a world of virtual exploration[8].

Knowledge Discovery in Databases (KDD) and Data Mining (DM) give a result to the information flood tide problem by rooting valid, new, potentially useful, and eventually accessible patterns from data [9]. Patterns constitute compact and rich in semantics representations of raw data [10] compact by means that they epitomize, to some degree, the quantum of information contained in the original raw data and rich in semantics by means that they reveal new knowledge hidden in the cornucopia of raw data.

Different data mining tasks achieve different perceptivity over the data bracket captures the class of data or a new item, clusters reveal natural groups in data, decision trees descry characteristics that prognosticate(with respect to a given class trait) the gets of unborn records, and so on [11]. This unorganized data requires processing to be done to induce meaningful and useful information from the large databases. In order to organize large quantum of data, you apply the conception of Database Management Systems (DBMS) similar as Oracle, and SQL Garcon. These Database Management Systems bear you to use SQT, a technical query language to recoup data from a database. Still, the use of SQT isn't always acceptable to meet the end stoner conditions of technical and sophisticated information from an unorganized large data bank. Database experimenters pay further attention to the issues related to the volume of data and also concerned with the effective use of the available database ways similar as effective

data reclamation mechanisms. This thus necessitates you to look for certain indispensable ways to recoup information from large and substantially unorganized sources of data. Currently, data stored in medical databases are growing in a decreasingly rapid-fire way. As saying that data is pivotal for medical decision-timber and operation [12]. It has been extensively honored that medical data analysis can lead to an improvement of health care by perfecting the performance of patient operation tasks. There are two main aspects that define the need for medical data analysis

1) Support of specific knowledge- grounded problem working conditioning through the analysis of cases' raw data collected in monitoring.
2) Discovery of new knowledge that can be uprooted through the analysis of representative collections of illustration cases, described by emblematic or numeric descriptors. For these purposes, the increase in database size makes traditional homemade data analysis to be inadequate. To fill this gap, new exploration fields similar as knowledge discovery in databases (KDD) have fleetly grown in recent times. KDD is concerned with the effective computer-backed accession of useful knowledge from large sets of data.

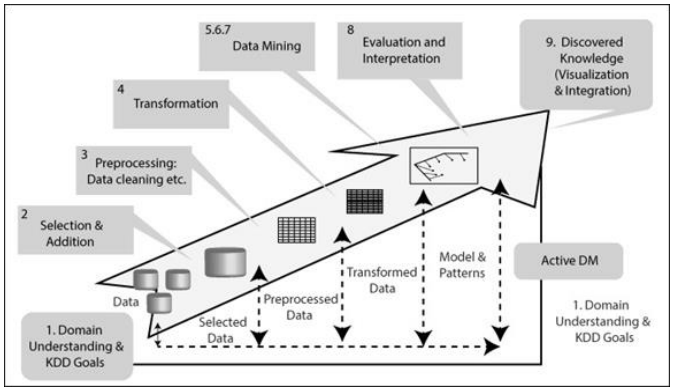

Figure 2. The KDD Process

It also includes the choice of garbling schemes, preprocessing, slice, and protrusions of the data previous to the data mining step. Data Mining refers to the operation of algorithms for rooting patterns from data without the fresh way of the KDD process (12). The KDD process is frequently to be nontrivial; still, we take the larger view that KDD is an each-encompassing conception. KDD is a process that involves numerous different ways. The input to this process is the data, and the affair is the useful information asked by the druggies. Still, the ideal may be unclear or inexact. The process itself is interactive and may bear important ceased time. To insure the utility and delicacy of the results of the process, commerce throughout the process with both sphere experts and specialized experts might be demanded.
Data mining is the step in the process of knowledge discovery in databases, that inputs generally gutted, converted data, searches the data using algorithms, and labors patterns and connections of interest in a particular emblematic form or a set of similar representations as bracket rules or trees, retrogression and clustering, to the interpretation/ evaluation step of the KDD process. The description easily implies that what data mining (in this view) discovers are suppositions about patterns and connections. Those patterns and connections are also subject to interpretation and evaluation before they can be called knowledge.

A simple data mining process model includes the following way (13)
1) Elect a target data set.
2) Data preprocessing.
3) Data metamorphosis.
4) Data mining.
5) Interpretation/ evaluation.
6) Donation.
7) Attestation Simply the attestation and reporting it to interested parties are done at this last step.

Whereas in unsupervised literacy no training set is used. Each data mining fashion serves a different purpose depending on the modeling ideal. The two most common modeling objects are bracket and Predication. Bracket models prognosticate categorical

markers (separate, unordered) while Predication models prognosticate nonstop- valued functions [14]. Several data mining algorithms are used in IQ opinion of heart complaint similar as Naive Bayes, Decision Tree, neural network, kernel viscosity, bagging algorithm, and support vector machine showing different situations of rigor. Indispensable names to data mining are knowledge discovery mining) in databases (KDD), knowledge birth, data patterns analysis, data archeology, data dredging, information harvesting, business intelligence, etc.

2.2. Techniques of Data Mining

According to https://www.geeksforgeeks.org/data-mining/, data mining combines many techniques from other fields, including statistics, machine learning, pattern recognition, database and data warehousing systems, information retrieval, and visualization. is integrated with. Collect more information about your data, predict hidden patterns, future trends, and behaviors to help businesses make decisions (see Figure 1 below). It's clear that over the past two decades we've been able to develop systems that collect large amounts of medical data, but what do we do with it now?

Data mining techniques use powerful computer software tools and large clinical databases. Sometimes in the form of a data repository or data warehouse, to identify patterns in data. Data mining techniques allow you to choose from a wide range of techniques and pattern types, including classification, clustering, and association rules, among others [12].

Technically speaking, data mining is a computational process that analyzes data from different perspectives, dimensions, and perspectives and categorizes/summarizes them into meaningful information. Data mining can be applied to any type of data, including: Data warehouses, transactional databases, relational databases, multimedia databases, spatial databases, time series databases, and the World Wide Web.

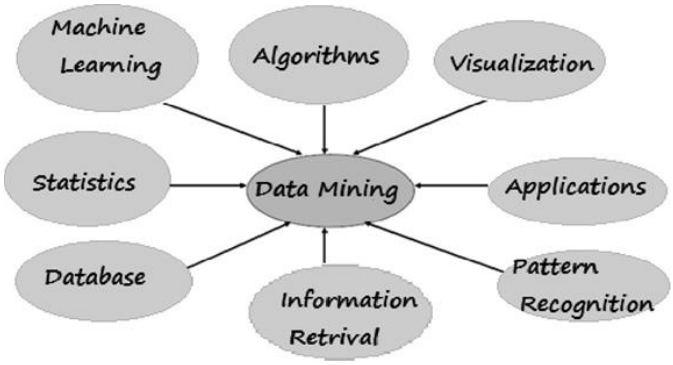

Figure 3. Data Mining Techniques.

There are three primary stages to the entire data mining process. As seen below in Figure 2, this is:

1) Pre-processing of the data involves cleaning, integrating, selecting, and transforming the data.

2) Data Extraction: Precise data mining occurs.

3) Evaluation and Presentation of Data: Conducting analysis and presenting findings.

2.3. Factors Driving Data Mining

Three factors have come together to make data mining a mature field: • First, massive amounts of data can be captured, stored, and processed at a low cost; • Second, database technology advancements allow data to be stored and organized in ways that facilitate quick answers to complex queries; and • Third, analysis methods can be effectively applied to these large and complex databases thanks to advancements and improvements in these methods [15]

2.3.1. Advantages of Data Mining in Medicine

According to [15], there are numerous advantages associated with data mining in medicine some of which are listed below:

· Earlier detection of illness.
· Symptom trends.
· Data analysis.
· Improved drug reactions.
· Provides new knowledge from existing data.
· Public databases.

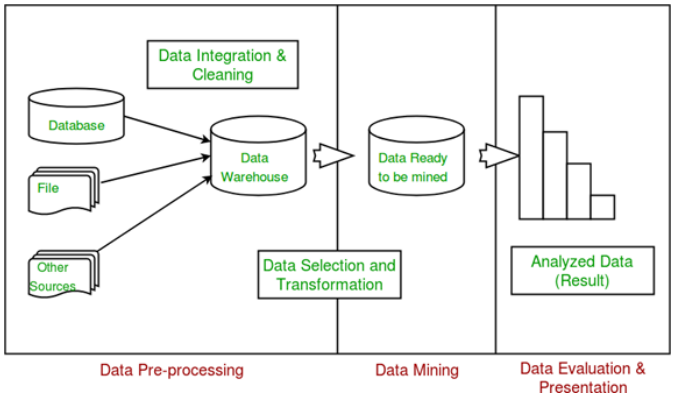

Figure 4. Data evaluation and presentation

· Government sources.
· Company or healthcare Databases.
· Old data can be used to develop new knowledge.
· New knowledge can be used to improve services or products.
· Improvements lead to: Bigger profits. More efficient service.

2.3.2. Disadvantages of Data Mining in Medicine
· No uniform language-Medical.
· Incomplete records.
· Privacy, etc.

2.4. Methodological Studies
Intelligent medical systems have benefited greatly from data mining [16]. Users can quickly assess the relationships between diseases, their true origins, and the consequences of symptoms that patients experience on their own using the software that has been designed for this purpose. Because the software is extensible, it can be used to apply large databases as input data. Healthcare providers can more easily identify patients who are at a high risk of developing heart disease by being aware of the risk factors linked to the condition. Healthcare practitioners can diagnose cardiac disease with the aid of statistical analysis and data mining techniques. Numerous algorithms have been developed over time to extract so-called nuggets of insight from massive data collections. There are several different methodologies to approach this problem like Classification, Regression Trees, Decision Trees, Support Vector Machines, etc.

2.4.1. Classification
In their opinion, [1]states that bracket is the process of prognosticating affair grounded on given input data. The purpose of bracket is to directly prognosticate the target class of each case in the data [17]. Reuse the training set and Predication set to prognosticate the data. First, connections between the attributes of the training data set are constructed. Next, a Predication dataset is handed that contains analogous attributes but different data values. It also analyzes the given data and makes prognostications by classifying different datasets into different classes grounded on the relationship of their attributes (18)(19) For illustration, for medical databases. As shown in Tables 1 and 2, the training set contains applicable patient information grounded on former records, and the prophetic trait is whether the case is at threat of a heart attack.

Table 1. Training set. Source: Kamna Solanki et al., 2016.

AGE	HEART RATE	BLOOD PRESSURE	HEART PROBLEM
62	79	145/70	YES
35	82	115/75	YES
79	65	110/68	NO

Table 2. Predictive set. Source: Kamna Solanki et al., 2016.

AGE	HEART RATE	BLOOD PRESSURE	HEART PROBLEM
45	96	143/69	?
63	54	108/73	?
83	95	115/68	?

Bracket uses prophetic rules expressed in the form of IF- also rules. The first part(IF part) consists of a combination of conditions, and the alternate part(also part) predicts a certain prognosticated trait value that exceeds the first part when it's satisfied. Using the illustration over, the rule to prognosticate the first row of the training set can be expressed as IF(age = 62 and heart rate> 72) or(age> 60 and blood pressure> 140/ 70), also heart problems = yes. This fashion provides a Predication rate of 80. Still, the optimal result is the 100 Predication rate rule. That's veritably delicate to achieve. Below is bracket ways used in drug.

2.4.2. Decision Trees

A decision tree and a flow chart are similar in that each non-leaf node indicates a test on a specific property, and each branch shows the test's result. The decision tree's root node is at the top. For instance, we may determine whether or not a patient needs to be readmitted using the use of the readmission tree. A decision-maker can select the optimal option by using a decision tree, and a traversal from root to leaf shows distinct class separation depending on maximum information gain [20] [21]. Decision trees are simple to understand and self-explanatory. A decision tree can also be used to create a set of rules.

2.4.3. Support Vector Machines (SVM)

The Support Vector Machine concept was first presented by Vladimir Vapnik [22]. Compared to all other current procedures, its accuracy is superior. Although it can be further expanded to multi-class issues, it was initially introduced for binary classification problems. To divide data points, it builds hyper-planes [23].

There are two methods to put it into practice:

1) Programming using mathematics.

2) Making use of kernel functions.

Non-linear functions are easily transferred to high dimensional space with the use of training data sets. The only way to do this is to use kernel functions like sigmoid, Gaussian, etc.

2.4.4. Neural Network (NN)

It was created in the twentieth century. Prior to the development of decision trees and SVM, which produce significantly superior results, neural networks were thought to be the greatest classification algorithms. This was the driving force behind NN's widespread adoption as the most popular classification algorithm across a range of biomedical and healthcare domains. For instance, neural networks (NNs) have been employed as the algorithm to enable the detection and prediction of diseases such as cancer. Basic components of NN are neurons or nodes. Because of their connections, these neurons cooperate to produce the output functions inside the network. They are fault resistant because, in the event that a few network neurons fail, they can still generate new observations based on the preexisting observations. Every neuron in the NN has an activation number given to it, and every edge has a weight. Because NN is adaptive by nature, its primary characteristic is its ability to minimize error through weight and structural adjustments. The ability of NN to effectively handle noisy training data and categorize novel data that differs from training data is one of its main advantages. NN has a number of drawbacks as well. The parameters chosen have a significant impact on the classification performance. Secondly, the process of training or learning it is quite costly and slow.

2.4.5. Regression

A data mining technique called regression aids in finding the functions that are helpful in illustrating the link between various variables. It is a tool for mathematics that is simple to build with training data sets. Depending on the number of independent variables, regression can be divided into linear and non-linear categories.

Researcher	Technique used	Purpose
1. Hu *et al.* [12]	SVM, decision tree, bagging and boosting.	To analyze micro array data.
2. Huang *et al.* [13]	Hybrid SVM based diagnosis model	For breast cancer.
3. Khan *et al.* [14]	Decision tree	For breast cancer.
4. Chang *et al.* [15]	Integrated Decision tree model.	For skin diseases in adults and children.
5. Curiac *et al.* [16]	Bayesian method	For psychiatric disease.
6. Moon *et al.* [17]	Decision tree algorithm	To characterize the smoking behavior among smokers by assessing their psychological health conditions and consumption of alcohol.
7. Chien *et al.* [18]	Hybrid decision tree classifier.	For chronic disease.
8. Shouman *et al.* [19]	K-NN classifier.	For heart disease.
9. Liu *et al.* [20]	Fuzzy-NN classifier.	For thyroid disease.
10. Er *et al.* [21]	Artificial Neural network	For chest disease.

Table 3. Usage history of classification techniques in HealthCare Sector. Source: Kamna Solanki et al.

Linear regression is used to evaluate the association between two types of variables, where one is an independent variable and the other is a dependent variable. The inability to employ this technique with categorized data is one of its drawbacks. Logistic regression

can be used to make use of the categorical data. Table 4 provides an overview of the regression's application in the health industry.

In order to estimate association between two types of variable in which one is dependent variable and another one is independent variable, linear regression is used. One of the disadvantages of this technique is that it cannot be used for categorized data. The categorical data can be used with the help of logistic regression. Usage of Regression for health sector has been summarized in Table 4.

2.4.6. Clustering

Unlike classification techniques (supervised learning methods), it is an unsupervised learning technique. It works well with enormous volumes of data. By observing independent variables, it functions. Creating clusters from huge databases based on similarity measures is the primary challenge. Table 5 defines different kinds of clustering algorithms, while Table 6 describes different clustering algorithms utilized in the medical field.

2.5. Applications of Data Mining

Finding high-risk patients: Healthcare practitioners can use data mining to find people who are more susceptible to certain illnesses or ailments.

Figure 5. Application of Data Mining

Customizing treatment plans: Medical professionals can use data mining to identify patterns in patient information that will enable them to craft personalized treatment plans that are tailored to each patient's particular requirements.

Enhancing patient outcomes: Healthcare professionals can utilize data mining to identify factors connected to better patient outcomes.

Healthcare professionals can use data mining to identify patterns in data that will assist them in anticipating epidemics.

According to https://www.geeksforgeeks.org/data-mining/, Data Mining can be

Researcher	Technique used	Purpose
1.Divya et al. [22]	Weighted SV Regression	To provide better healthcare services by continuously monitoring patients.
2. Xie et al. [23]	Regression decisiontree algorithm	To study number of hospitalization days.
3.Alapont et al. [24]	Linear regression	For effective utilization of hospital resources.

Table 4. Usage history of regression techniques in HealthCare sector. Source: Kamna Solanki et al.

For example, data mining can identify individuals who are susceptible to diabetes, cancer, heart disease, or other illnesses based on their lifestyle choices, medical history, family history, and other relevant data.

Technique	Description
1.Partitioned Clustering	With the help of “n” data points maximum possible of “k” clusters is obtained by relocating objects to “k” clusters.
2. Hierarchical Clustering	Data points are partitioned in tree form either top-down or bottom-up.
3. Density-based Clustering	It can handle cluster of any arbitrary shape whereas above two can handle only spherical shape clusters.

Applied in the following Areas as shown in Figure 3 below:

1. Financial Analysis 2. Biological Analysis 3. Scientific Analysis

4. Intrusion Detection 5. Fraud Detection 6. Research Analysis

Researcher	Technique used	Purpose
1. Chen et al. [25]	Hierarchical clustering	To analyze micro-array data.
2. Chipman et al. [26]	Hybrid Hierarchical clustering.	Toanalyze micro-array data.
3. Bertsimas et al. [27]	Clustering algorithm	To predict health care cost.
4. Peng Y et al. [28]	Clustering algorithm	To detect healthcare frauds.

5. Belciug et al. [29]	Hierarchical, partitioned and density-based clustering.	Efficient utilization of healthcare resources.

Table 6. Usage history of clustering techniques in Healthcare sector. Source: Kamna Solanki et al.

Applications of Data Mining in Healthcare Domain

Patient care has already improved in certain quantifiable ways as a result of data mining applications in the healthcare industry [24]. Surgeon Dr. Nicholas Morrissey, who is involved in the effort, stated in an interview that New York-Presbyterian Hospital has used analytics software to lower the risk of potentially fatal blood clotting in patients. In 2010, the hospital began scanning medical data as patients were admitted using Microsoft software to check for risk factors like smoking, cancer, and length of bed rest. Doctors saved time and produced better ratings by allowing the program handle the assessment instead of the hospital staff, according to Morrissey. According to Morrissey, the rate of clotting has decreased to 0.23 events per 1000 patient days from 0.33 incidences per 1000 patient days before to the software's adoption. He remarked, "That was a significant drop. I won't be out there saying that we've solved the problem, but we're definitely making progress."

According to [25], data mining's successful application in a variety of industries, including engineering, banking, retail, and marketing, has led to the field's development into other areas, including public health and medical.

Figure 6. Application of data mining in Health care

These days, a growing number of data mining applications concentrate on examining healthcare facilities to improve patient care, identify hospital illness outbreaks, stop patient fatalities, and, of course, identify bogus insurance claims [21] [26] [27] [28].The process and technology of data mining enable the extraction of knowledge from vast amounts of data, which is then applied to decision-making. The author of [27] talks about how data mining can be used to enhance.

The standard of the pharmaceutical industry's decision-making procedure. Actually, a key issue with pharmacological data is informational deficiency. Anticipating drug behavior is crucial in determining whether treatment improves or deteriorates a patient's state of health. Experts in healthcare management [29] can use data mining to support their decision-making in the CRM space. Patients will benefit from improved and more reasonably priced healthcare services provided a substantial quantity of information regarding other patients' satisfaction levels with the medical field is properly evaluated and interpreted. Because biological databases have a large variety of data types and frequently intricate relational structures, they can be thought of as the raw material for multi-relational data mining algorithms [30].

A surveillance system that use data mining techniques (association rules) to find novel and intriguing patterns

in the infection control data was put into place at the University of Alabama [31]. Three alternative studies, each utilizing a different amount of data partition, were carried out on the data that were gathered over the course of a year (1996).

The research [32] presents a case study of American Healthways, which offers hospitals and health plans diabetes management services to improve the quality and reduce the cost of treating people with diabetes.

The current article's authors concentrate their study on using data mining techniques to categorize thyroid disease patients. The following data mining algorithms have been recognized by the writers of literature on thyroid illness diagnosis: decision trees, expert systems, support vector machines, artificial neural networks, etc. As an illustration, the use of ANNs in the diagnosis of thyroid problems is covered in [33] [34] [35]. Authors in [33] used information from James Cook University in Townsville, Australia, which was gathered in 1992 and tied to the UCI site. There were 215 laboratory samples in all. One property was the objective and five other attributes served as predictors in the data mining technique. The degree of classification accuracy in the instance of thyroid disease was 98.6% thanks to the selection of a hidden layer, the Logsig activation function for the hidden layer, and six neurons from this layer. MATLAB 2012 was utilized as the testing software for the model. Authors discuss their research on three ANN algorithms—the learning vector quantization (LVQ) networks, the radial basis function (RBF) networks, and the back propagation algorithm (BPA)—for the detection of thyroid disease in [34]. The LVQ network had the highest accuracy rate, or 98%, following the model evaluation. Utilizing support vector machines, thyroid nodule categorization was carried out in [36]. A comparative analysis of data mining classification algorithms (C 4.5, C5.0) for thyroid cancer is reported in [37]. A database including 400 entries and 29 attributes—extracted from the UCI thyroid database—was utilized by the authors of [37]. According to the analysis, the rule set produced by C5.0 has a confidence level more than 95%. The C4.5 technique was applied to the Java platform in [37] with the help of Eclipse and the Windows XP operating system. While a three-stage expert system based on support vector machines is explored in [39], an expert system for diagnostics based on fuzzy rules is detailed in [38]. Using a 10-fold cross-validation procedure, the system suggested in [39] achieved the highest accuracy recorded to date in the classification of thyroid illnesses, with a mean accuracy of 97.49% and a maximum accuracy of 98.59%.The authors will provide a case study of the classification models used on a database containing information on people with thyroid disorders in the section of the paper that follows. 756 entries that were taken from the UCI Machine Learning Repository made up the data set [40]. Along with a class attribute, we employed 21 attributes as predictors. CART and Tree Net models were employed in the trials that are detailed below. In most of our experiments, the accuracy of the CART model was over 93%, with the highest value being 96.86%, for the following settings: Priors = Equal, Costs = 0.5, Parent node min cases = 10, Terminal node min cases = 1, Partition = 0.6. These results were compared with those found in the previously mentioned studies. The Tree Net model's accuracy was 94.97%.

III. Summary and Conclusion

This research work discussed vital issues bothering on Data Mining as a Technique applied in Healthcare Industry. Some of these concepts discussed as Techniques include but are not limited to the following: Classification, Regression, Decision Trees, Support Vector Machines, etc. This work also concludes that data mining is of great importance in the solution of healthcare problems. Data mining, however, is not a “silver bullet” capable of solving all of Cardiovascular Diseases, but rather it aims at providing possible prevention methods, remedies and symptoms.

Some diseases pose a serious life-threatening risk globally across all races and age groups. But most of them could be controlled and risk reduced through a proper medical checkup, proper dieting, nutrition, healthy eating habits and exercising all of which are incorporated into our data mining software for easy access.

This work also highlighted that the main goal of achieving high accuracy and efficiency which is very important in healthcare sector still remains an open research issue.

The healthcare business has benefited greatly from data mining, particularly in terms of disease prediction. Diagnostics play a major role in medical diagnosis and are commonly employed in illness prediction. To sum up, no single data mining technique can address all of the problems with healthcare data sets. We must create a hybrid model that may address the aforementioned problems in order to get the best accuracy among classifiers, which is crucial for medical diagnosis with the features of the data being taken into consideration. Our next goals are to improve the forecasts by utilizing hybrid models.

IV. REFERENCES

[1]. Kamma, S., Parul, B., Sandeep, D. and Sudluv (2016) Analysis of Application of Data Mining Techniques in Healthcare. International Journal of Computer Applications, 148, 16-21. https://doi.org/10.5120/ijca2016911011

[2]. Koh, C.H. and Tan, G. (2011) Data Mining Applications in Healthcare. Journal of Healthcare Information Management, 19, 64-72.

[3]. Kaur, H. and Wasan, S.K. (2006) Empirical Study on Applications of Data Mining

[4]. Techniques in Healthcare. Journal of Computer Science, 2, 194-200. https://doi.org/10.3844/jcssp.2006.194.200

[5]. Obenshain, M.K. (2004) Application of Data Mining Techniques to Healthcare Data. Infection Control & Hospital Epidemiology, 25, 690-695. https://doi.org/10.1086/502460

[6]. Liao, S.H., Chu, P.-H. and Hsiao, P.-Y. (2012) Data Mining Techniques and Applications—A Decade Review from 2000 to 2011. Expert Systems with Applications, 39, 11303-11311. https://doi.org/10.1016/j.eswa.2012.02.063

[7]. Ubochi, C.I. (2017) Data Mining Technique for Detecting Cardiovascular Diseases. An Unpublished Msc Thesis.

[8]. Savage, N. (2011) Mining Data for Better Medicine. MIT Technology Review, 38, 235-237.

[9]. Joshi, S., Deepashenoy, P., Venugopal, K.R. and Patnaik, L.M. (2010) Data Analysis and Classification of Various Stages of Dementia Adopting Rough Sets Theory. International Journal on Information Processing, 4, 86-89.

[10]. Fayyad, U.M., et al. (1996) From Data Mining to Knowledge Discovery in Databases. AAAI Press/The MIT Press, Cambridge.

[11]. Rizzi, S., Bertino, E., Catania, B., Golfarelli, M., Halkidi, M., Terrovitis, M., Vassiliadis, P., Vazirgiannis, M. and Vrachnos, E. (2003) Towards a Logical Model for Patterns in ER. Springer, Chicago, Vol. 2813, 77-90. https://doi.org/10.1007/978-3-540-39648-2_9

[12]. Ntoutsi, I. (2008) Similarity Issues in Data Mining—Methodologies and Techniques. University of Piraeus, Piraeus, 31-32.

[13]. Li, J.-S., Yu, H.-Y. and Zhang, X.-G. (2011) Data Mining in Hospital Information System. In: Funatsu, K., Ed., New Fundamental Technologies in Data Mining, InTech, Shanghai, Vol. 1, 143-156.

[14]. Balasunda, V., Devi, T. and Saravanan, N. (2012) Development of a Data Clustering Algorithm for Predicting Heart. International Journals of Computer Applications, 48, 8-13. https://doi.org/10.5120/7358-0095

[15]. Han, J. and Kamber, M. (2006) Data Mining Concepts and Techniques. 2nd

[16].Edition, Morgan Kaufmann Publishers Inc., Burlington, 55-86.

[17].Chauraisa, V. and Pal, S. (2013) Data Mining Approach to Detect Heart Diseases. International Journal of Advanced Computer Science and Information Technology, 56-66.

[18].Aflori, C. and Craus, M. (2007) Grid Implementation of the Apriori Algorithm.

[19].Advances in Engineering Software, 38, 295-300. https://doi.org/10.1016/j.advengsoft.2006.08.011

[20].Deulkar, D.S. and Deshmukh, R.R. (2016) Data Mining Classification. Imperial Journal of Interdisciplinary Research, 2.

[21].Palaniappan, S. and Awang, R. (2008) Intelligent Heart Disease Prediction System Using Data Mining Techniques. IEEE Conference on Computer Systems and Applications, Doha, 31 March 2008-4 April 2008, 108-115. https://doi.org/10.1109/AICCSA.2008.4493524

[22].Srinivas, K.B., Rani, K. and Govrdhan, A. (2010) Applications of Data Mining Techniques in Healthcare and Prediction of Heart Attacks. International Journal on

[23].Computer Science and Engineering, 2, 250-255.

[24].Ahmed, P., Qamar, S. and Rizvi, S.Q.A. (2015) Techniques of Data Mining in

[25].Healthcare: A Review. International Journal of Computer Applications, 120, 38-50. https://doi.org/10.5120/21307-4126

[26].Durairaj, M. and Ranjani, V. (2013) Data Mining Applications in Healthcare Sector: A Study. International Journal of Scientific and Technology Research, 2, 29-35.

[27].Vapnik, V. (1998) The Support Vector Method of Function Estimation. In: Suykens, J.A.K. and Vandewalle, J., Eds., Nonlinear Modeling, Springer, Berlin, 55-85. https://doi.org/10.1007/978-1-4615-5703-6_3

[28].Burges, C.J.C. (1998) A Tutorial on Support Vector Machines for Pattern Recognition. Data Mining and Knowledge Discovery, 2, 121-167. https://doi.org/10.1023/A:1009715923555

[29].Robertson, J. (2012) Data Mining in Doctor's Office Helps Solve Medical Mysteries. Vol. 1, Wal-Mart or Western Union United Healthcare Corp., New York.

[30].Ionita, I. and Ionita, L. (2016) Applying Data Mining Techniques in Healthcare. Studies in Informatics and Control, 25, 385-394. https://doi.org/10.24846/v25i3y201612

[31].Canlas Jr., R.D. (2015) Data Mining in Healthcare: Current Applications and Issues.

[32].Ranjan, J. (2007) Application of Data Mining Techniques in Pharmaceutical Industry. Journal of Theoretical and Applied Information Technology, 3, 61-67.

[33].Diwani, S., Mishol, S., Kayange, D.S., Machuve, D. and Sam, A. (2013) Overview Applications of Data Mining in Health Care: The Case Study of Arusha Region.

[34].International Journal of Computational Engineering Research, 3, 73-77.

[35].Desikan, P., Hsu, K.W. and Srivastava, J. (2011) Data Mining for Healthcare Management. SIAM International Conference on Data Mining, Arizona.

[36].Page, D. and Craven, M. (2016) Biological Applications of Multi-Relational Data Mining. http://www.kdd.org/exploration_files/Page.pdf

[37].Brossette, S.E., Sprague, A.P., Hardin, M.K., Waites, B., Jones, W.T. and Moser, S.A. (1998) Association Rules and Data Mining in Hospital Infection Control and Public Health Surveillance. Journal of the American Medical Informatics Association, 5, 373-381. https://doi.org/10.1136/jamia.1998.0050373

[38].Ridinger, M. (2002) American Healthways Uses SAS to Improve Patient Care. DM Review, 12, Article No. 139.

[39].Gharehchopogh, F.S., Molany, M. and Mokri, F.D. (2013) Using Artificial Neural Network in Diagnosis of Thyroid Disease: A Case Study. International Journal on Computational Sciences & Applications, 3, 49-61.

[40].Shukla, A. and Kaur, P. (2009) Diagnosis of Thyroid Disorders Using Artificial Neural Networks. IEEE International Advance Computing Conference, Patiala, 6-7 March 2009, 1016-1020. https://doi.org/10.1109/IADCC.2009.4809154

[41].Prerana, E., Sehgal, P. and Taneja, K. (2015) Predictive Data Mining for Diagnosis of Thyroid Disease Using Neural Network. International Journal of Research in Management, Science & Technology, 3, 75-80.

[42].Chang, C.Y., Tsai, M.F. and Chen, S.J. (2008) Classification of the Thyroid Nodules Using Support Vector Machines. International Joint Conference on Neural Networks, Hong Kong, 1-8 June 2008, 3093-3098. https://doi.org/10.1109/IJCNN.2008.4634235

[43].Upadhayay, A., Shukla, S. and Kumar, S. (2013) Empirical Comparison by Data Mining Classification Algorithms (C 4.5 & C 5.0) for Thyroid Cancer Data Set. International Journal of Computer Science & Communication Networks, 3, 64-68.

[44].Keleş, A. and Keleş, A. (2008) ESTDD: Expert System for Thyroid Diseases Diagnosis. Expert Systems with Applications, 34, 242-246. https://doi.org/10.1016/j.eswa.2006.09.028

[45].Chen, H.L., Yang, B., Wang, G., Liu, J., Chen, Y.D. and Liu, D.Y. (2012) A Three Stage Expert System Based on Support Vector Machines for Thyroid Disease Diagnosis. Journal of Medical Systems, 36, 1953-1963. https://doi.org/10.1007/s10916-011-9655-8

[46].UCI Machine Learning Repository. https://archive.ics.uci.edu/ml/machinelearning-databases/thyroid-disease

Cite this article as :

Swati C. Tawalare, Dr. S. Y. Amdani, "A Significance of Data Mining in the field of Healthcare Sector", International Journal of Scientific Research in Computer Science, Engineering and Information Technology (IJSRCSEIT), ISSN : 2456-3307, Volume 10, Issue 1, pp.217-230, January-February-2024.