VOL 2 (2018) NO 4

e-ISSN : 2549-9904
ISSN : 2549-9610

INTERNATIONAL JOURNAL ON INFORMATICS VISUALIZATION

Data Mining Usage and Applications in Health Services

Mehmet Akif CIFCI [#], Sadiq Hussain [*]

[#] *Istanbul Aydin University, Computer Engineering Program, 34295, İstanbul, Turkey*
[*] *System Administrator, Dibrugarh University, Assam, India*
E-mail: wwwakif@msn.com, sadiq@dibru.ac.in

***Abstract*— Data Mining (DM), used to extract large amounts of hidden, valuable, useful information in large quantities and to provide strategic decision support, has created a new perspective on the use of health data. It has become a rapidly growing method of responding to problematic areas of data in large quantities in almost all sections. Although in health services it seems to be slow, a major leap has come to the scene. The aim of this study is to provide a new perspective on decision-making processes by creating an infrastructure for the health data and to provide examples for healthcare workers in the healthcare industry using DM techniques. Forasmuch as, the conceptual framework of data discovery in databases, Data Warehousing, DM, Business Intelligence (BI) has been given. DM applications and usages are given as examples of priority issues and problem areas in the health sector.**

***Keywords*— Data mining, knowledge discovery in databases, data warehouse, business intelligence**

I. Introduction

The basis of health system policies and administrative decisions lay on the data and the knowledge gained from that data [1]. Health policies and decisions are based on reliable, up-to-date and accurate data which should be appropriate and effective. The aim of Health Information Systems (HIS) is to produce useful information from large amounts of health data. This information is used to provide better healthcare at the patient level, better management of health institutions, effective use of resources and the creation of health policies [2]. As health data is collected by hospitals, other health institutions, insurance companies and many other public institutions; it can be less reliable, so data security and cleansing become inevitable. Nowadays, the increase in the volume of digital data has created new problem areas [3]. These include: developing methods or systems for processing large amounts of multidimensional and complex data; developing methods or systems for processing new types of data; developing methods, protocols or infrastructure for processing distributed data; developing models for the use and security of data.

DM has roots which have been used thoroughly. Financial organizations have benefitted from DM to score the credits and to detect fraud [4]. The dealers, retailers for direct marketing and cross-selling they are into DM areas. Manufacturers take benefit of DM for quality control and maintenance schedule. In recent years in healthcare, DM has begun to gain popularity [5]. A lot of factors have motivated the use of DM applications in healthcare. Some of the examples are such the existence of medical insurance fraud and abuse.

The purpose of the study is to provide an overview of the DM process and to provide health professionals with a new perspective on the decision-making process by presenting examples of the use of DM in the healthcare industry [6]. To this aim, descriptive information is respectively given in the articles on Knowledge Discovery in Databases (KDD) [7], Data Warehousing (DW), DM, Business Intelligence (BI) and DM Methods; examples are given to the applications of DM by considering priority issues and problem areas in the health sectors.

The rest of the paper is organized as follows. In the next section, we highlight KDD. In Section 3, we provide some background on DW. Section 4 is about DM and in Section 5 we investigate DM and BI. In Section 6, we assess the performance of the proposed approaches in Health Service applications and usages. Finally, Section 7 concludes the paper.

II. Literature Review

Neesha Jothi et. al. reviewed various papers related to healthcare with respect to methods, tasks, algorithms and models and the findings were summarized in the conclusion. They had searched various papers from the year 2005 to 2015 in the online databases of IEEE Explore, ScienceDirect and Springerlink. The period was crucial as newly developed methodologies emerged during that time specially meant for healthcare industries. 50 articles were selected for review out of 205 articles. They categorized the papers with respect to discipline, model, tasks, and methods. If they categorized papers model wise, then there were 47 predictive model papers in contrast to 3 descriptive model papers. If we looked at the papers task-wise, 42 papers were based on classification, 5 papers based on association rules, 2 papers performed the clustering tasks while only one paper described anomaly detection. Machine learning based papers counted 40 from the discipline point of view. They concluded that there was no accurate data mining technique to tackle the issues related to the healthcare industry. They proposed for a hybrid model for prediction of different types of diseases [8].

Olegas Niaksu et. al. investigated the practice of different data mining applications in the field of medicine. They surveyed seven hospitals from five different countries with the different economic setup from underdeveloped to developed countries. They also explored the publications related to applications in the field of healthcare for the last 8 years. They found that there were about 400 publications in the area of medicine whereas, at the start of the 90's, it was only 5 publications. The search was based on the Web of Science, Google Scholar, and PubMed databases. They found that although the medical communities were aware of the practical usage of data mining application in the healthcare industry, only 29% respondents were actually able to cite such applications. The interdisciplinary approach was not enough for proper usage of academic advancements in the healthcare. They concluded that the domain-specific problems may be tackled carefully by using self-explanatory models for the DM applications in real practice of this industry [9].

Boris Milovic et. al. suggested in their study that the physician may be able to take the help of DM methods and models for the diagnoses of diseases and better treatment of their patients. The electronic health records of the patients may be mined to save lives of the patients and cut down the costs of the medical bills. The data produced in this industry are huge and heterogeneous. These data need to be preprocessed and integrated before mining to get valuable and useable knowledge. These hidden patterns are very useful for the healthcare industry to treat the patients better as well as for prediction of some of the contagious diseases. The extracted knowledge leads to important decision making in clinical diagnosis as well as different healthcare issues [10].

Sandeep Kautish et. al. studied that data mining may be used effectively in detection and prediction of several diseases like diabetics, lung cancer, breast cancer, heart disease, kidney problem, liver failure etc. The data mining application had a major role to play in the field of hospital management and diagnosis by reducing and minimizing different errors and diagnosis of diseases based on the clean electronic health records of the patients and thereby reducing the death rates [11]. M. Durairaj et. al. presented a comparative study of the different tools, techniques and methodologies prevailed in the healthcare industries and medical sector. As the medical records are massive, irrelevant and noisy, the data mining tools become handy for such records. They found that single technique was not enough for healthcare sector data, but a combination of different tools and techniques performed better results. They also observed an accuracy of 97.77% for cancer data prediction and the success rate of IVF treatment was estimated as 70% [12].

Veenu Mangat used Weka 3.6.0 data mining and machine learning software for enhancing strategic decisions based on medical records. The medical datasets were converted to arff format for analysis and comparison of different association rules. It was found that for strategic decision making, the predictive Apriori algorithm was the best suited to its highest and lowest values. The authors noticed that the limitations of the method were that success of extracting association rules depended on the datasets as some of the datasets were not suited for any kind of association rules [13].

Fei Jiang et. al. surveyed the usage of AI applications in healthcare industries at present and also explored the possibilities in near future. As the AI is based on the human cognitive functions, it has great potential in the area of healthcare as the analytical techniques improved to a new height. The field of AI expanded from the machine learning to modern deep learning techniques with both structured and unstructured data. AI is used extensively for the data related to cardiology, neurology, and cancer. The authors used real applications of AI by using detection and diagnosis of stroke, treatment, and prediction of stroke in a broader sense. They studied IBM Watson and hurdles of real-time deployment of AI applications [14].

III. Knowledge Discovery in Databases

A basic concept that needs to be conveyed before talking about DM is KDD [15]. The relationship between KDD and DM is undeniable. DM is among one of the most important steps of KDD as can be seen in Fig.1 KDD is a multi-step process that provides the conversion of data into useful information [9]. As seen in Fig.1 health data is processed then transformed and applied ML methods, finally assessed.

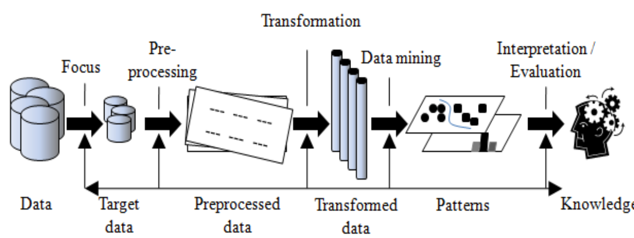

Fig. 1 Knowledge discovery in databases

Historically, the concept of finding useful patterns in data has been given various names such as DM, information transfer, information discovery, information harvesting, data archaeology and data pattern processing [16]. DM is mostly used by statisticians, data analysts, and management information systems communities [17]. At the same time, it has also gained popularity in the database fields. With the development of Artificial Intelligence (AI) and Machine Learning (ML), it has become widespread [18]. From the perspective above-mentioned, KDD is at the heart of the discovery process of useful information [19], and DM is a step in this process. DM is the implementation of special algorithms for the transmission of data patterns [20]. KDD continues to evolve and develop as an intersection of ML, pattern recognition [21], databases, statistics, AI, expert systems, data visualization and high-performance have been computing. One of the goals is to deliver a high level of information at a low level in a large dataset [22]. DM components of KDD relies heavily on known techniques such as ML, statistics, and pattern recognition to find patterns in KDD's. A natural question in this subject may be as: 'How does KDD differ from pattern recognition or ML or related areas?' The answer as seen in Fig. 2 is that these areas, having association with KDD, focuses on all the processes of exploration of verbose information, how to store and retrieve the data, how algorithms can be scaled to huge data sets and still work effectively [23, 24], how the results can be interpreted and visualized, and how the whole human-machine interaction can be usefully modeled and supported.

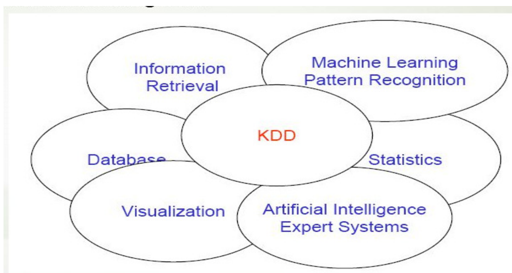

Fig. 2 KDD relation with AI methods

KDD process involves the implementation of DM methods to extract the desired selection, preprocessing, sub-sampling, transformation, patterns in databases using the interpretation of DM products to identify the exposed patterns [25]. KDD process should be evaluated as an algorithmic statement of which DM components are to be transferred where patterns are to be taken into consideration [26]. The whole of KDD process also includes a possible interpretation of what evaluation and mined patterns will be.

KDD process is interactive and iterative, consisting of steps that require the user to make decisions. The framework of some basic steps of the process is given in Table 1.

TABLE I
THE IMPORTANT BASIC STEPS OF KDD PROCESS

Step 1: To develop an understanding of the application domain and related prior knowledge and to define the objective of KDD process from the client's point of view.
Step 2: Creating the target dataset: Selecting the dataset to be probed or focusing on a subset of the variables or instances of the dataset.
Step 3: Data cleansing and pre-processing: Basic operations involving removal of noise.
Step 4: Data reduction and projection: The task is to find useful properties to represent the data depending on the target. The number of variables considered by size reduction or transformation methods can be reduced or the invariant representation of the data can be found.
Step 5: Matching of KDD process objectives of DM methods: Summarization, classification, regression, clustering etc.
Step 6: Choosing of DM algorithm(s): Descriptive analysis, model and hypothesis selection: This process involves determining which model and parameters may be appropriate and whether DM methods match all the criteria of KDD process.

A. Data Mining Steps in Knowledge Discovery

Information discovery objectives are defined according to the intended use of the system [27]. The goals can be divided into two categories.

a) Verification: With verification, the system is limited to validating the user's hypotheses. With the discovery, the system finds new patterns independently [28]. In the future, the discovery goals will be divided into two subgroups: prediction when the system is used to predict future behaviors of some assets, and identification when the system is used for a form that the handlers can understand for presentation to the clients.

b) Discovery: DM requires to construct the observed data model or to identify the patterns in the observed data. The model fitting assumes the role [29] of information extraction: The model is a part of the interactive KDD process which the model typically refers to as a subjective human judgment, whether it points to useful information or not. Two basic mathematical structures are used statistically and logically in constructing model [30]. As such most of DM methods are based on tried and tested techniques from ML, pattern recognition.

IV. DATA WAREHOUSE

Data Warehouses (DW) are special databases on which the DM process is performed. By definition, DW is a repository of data from many different sources, often in different structures, and is expected to be used under the same combined roof [31]. In addition, DW has the ability to analyze data from many different sources under the same roof. A related field developed from databases is DW whose name is also referred to as the popular business trend to collect transaction data and make them suitable for online analysis and decision support purposes [32]. DW helps in two important ways for KDD phase of data aggregation: these are data cleansing and data access.

A. Data Cleansing

Organizations are obliged to represent and handle the missing data and to address noise and misstatements as well as to consider the results of the mapped data as a single name since the extensive data and databases they have are hard to imagine being in a unified logical view [33].

B. Data Access

Generally, proper and well-defined methods should be created to provide access to the database, in particular, access to databases that are difficult to obtain in terms of history. First of all, organizations and individuals should solve the problems of storing and accessing data [34]. Naturally, the next step is the question 'What do we do with all the data?' This question naturally raises the opportunity for KDD. One popular approach to analyzing DW is Online Analytical Processing (OLAP). OLAP tools are focused on providing multidimensional data analysis that is superior to SQL (Structured Query Language) in computational summaries and definitions in many dimensions. OLAP tools are aimed at providing and simplifying interactive data analysis [35]. But the goal of KDD tools is to automate the process as much as possible.

V. Data Mining

It can be said that there has been vast potential for DM applications in healthcare [36]. Commonly, all of these can be seen as normal healthcare applications but still the usage and exploiting the data is essential. The usage of DM is mostly for detection of fraud and abuse [37]. Nonetheless, specialized medical DM, such as predictive medicine is the main focus. KDD is often referred to as DM, which aims to discover useful information from a large volume of data collections. Databases are now expressed in petabytes. 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 DM, which increases both revenue and costs. DM is a process that explores patterns and relationships in data with the use of many analysis tools and uses them to make valid estimates. DM, by its simplest definition, automatically determines the associated patterns in databases [38]. Assuming DM as magic is not sensible. For years, the statisticians have looked for statistically significant associations in manually digging databases whereas in DM, this process is performed automatically.

DM has a vital role in uncovering new trends in healthcare organizations which in turn useful for all the parties related to the field. DM dataset depicts a collection of techniques aimed at finding undiscovered patterns, the goal of which is to create decision-making models for predicting future behaviors based on analysis of past activities. Some examples are shown in Table 2.

TABLE III
SOME EXAMPLES OF AI

1. Estimate the risk of a male patient getting prostate cancer based on health status, lifestyle habits, and genetic factors.
2. To estimate the likelihood of a company's financial crisis based on financial performance measures and economic data,
3. To define the numbers and letters written by hand from the magnetic image,

VI. Data Mining and Business Intelligence

Business Intelligence is a general term used for all processes, techniques, and tools that support decision making at work and that are based on information technology [39]. DM is a new and important component of BI. Fig. 3 shows the logical positions of different BI solutions according to their potential values (from a data source, DW, DM, data presentation and decision making) on the basis of tactical and strategic business decisions [40]. DM is frequently well-defined as the analysis of data for relationships and patterns not previously revealed by traditional and conventional methods.

BI defines processes and procedures for methodically collecting, keeping, examining, and offering entree to data so as to assist initiatives in making enhanced decisions. BI applications contain the actions of decision support systems, management information systems [41].

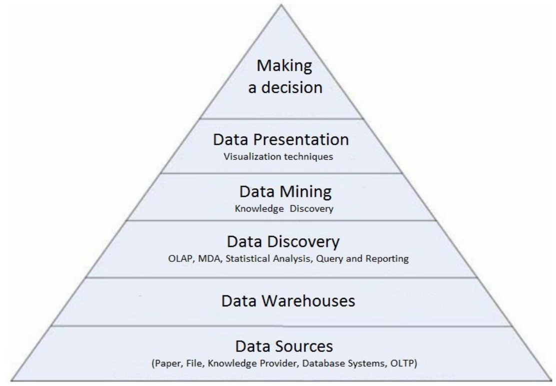

Fig. 3 DM and Business Intelligence

VII. Applications of Data Mining and Health Services

The health sector is one of the fastest changes in information content and structure [42]. For health services to be offered in the fastest, most accurate, highest quality and most responsive way, healthcare professionals need access to the most accurate and up-to-date information and use this information through decision support systems. DM is used to extract large amounts of hidden, valuable, usable information within the data and to provide strategic decision support; is a method that creates decision-making models based on the analysis of data [43]. Therefore, the use of DM as a decision support tool in the delivery of health services, management of all health institutions and health policies at all levels will help health professionals to make the most optimal decisions [44].

Applying the right DM solution here is very important in terms of achieving the effective results. In this part of the study, examples of DM solutions are presented to provide a perspective to back decision-support of all health professionals in the public and private sector. While such examples have been identified, the priority issues of healthcare professionals and workers are considered, for this reason, some terms are focused and explained here.

A. Creating Data Warehouse

Accessing to health data, the provision of analytical and cleansing data for health sector which the enterprises are taking benefit for different purposes constitutes one of the most important parts of DM. In addition, compiling valuable data such as clinical and demographic information within large data helps to make a better decision and taking right steps toward the problems.

Easy access also seen as one of the fundamental requirements for the implementation of DM methods is to provide data and analyzable clean data which is resolved by DW. All data in the hospitals can be cleaned and cached and downgraded to a single manufacturer database or data warehouse. Since large volumes of data produced by hospitals are often kept randomly arranged in a hierarchy, "Data Dumps" are created that are far from analyzable. With the DW that will prepare the DM sub-base, a clean, analytical database access that is needed in the DM process is gained [45].

B. Creating Electronic Patient Files

All records of the patient's story; diagnostic treatment processes; laboratory results; X-ray, MR, etc., are indexed in a single time-indexed manner and are of great importance in the evaluation of data and in the provision of services [44]. To be consistent with the DW logic, it is necessary to create an integrated structure that can be used to consolidate a single database from many databases of available and high-quality data or to provide access from a single center.

This will provide vast advantages as in the diagnosis and treatment process of the patients, it is necessary to provide clean access to the decision-support system and to prepare the infrastructure in accordance with the DM methods to be used.

C. Data Compliance Solution

One of the important steps in the use of health data is the solution of the data problems and an approach is suggested as such. When it comes to data problems, the basic senses are missing data, inconsistent data, contradictions value, extreme value [46]. As the solution of the mentioned problem, a non-parametric, dynamic approach should be followed, based on the DM profiling approach aiming at direct results.

Furthermore, it will have many advantages. Researchers who will implement applications in the health sector with a DM based profiling approach will have the possibility of eliminating the data problems in each different pattern by considering the characteristics of the patterns. Thus, researchers will have reached a solution that will allow less error, more organized data.

D. Detection of Early Warning Signs for Chronic Diseases

Along with the increase in the average lifespan, the frequency of chronic illnesses and the financial burden that it has brought in parallel is increasingly seen. From this point forward, proactive solutions should be developed to prevent the emergence of chronic diseases. The solutions can be viewed as considering all the social, economic, demographic, geographical, etc. variables for each chronic disease, the variables that have an effect on the occurrence of the disease can be determined with Key Component Analysis, Factor Analysis or Logistic Regression. Then risk signals can be developed that can indicate the occurrence of the disease by taking the limit values into account for which the affected variables are applicable. This provides such advantages as rather than testing hypothetical acceptance hypotheses with common acceptance, it may be possible to determine normative values for different groups. Thus, according to different groups, different policy solutions can be developed [47].

E. Error and Abuse Detection for Laboratory Tests

It is a very important issue in terms of patient safety to make a difference between errors and abuses that occur in the presentation of the health services, to minimize the risks and to take the necessary precautions according to this separation. Such solutions as joining a large volume data account, a sequential process needs to be designed. Firstly, by separating the standard values from the K-means-clustering analysis, then the findings obtained by using the Decision Trees and Association Rules (DTAR) methods to determine whether the anomaly values indicate a specific health condition or an abuse [48], are compared with the standard values obtained by the Clustering Analysis; it can be determined that the anomalous signifies a specific health condition or abuse. In case of ambiguity, the analysis process is repeated with the updated data, taking advantage of the expert opinions [49]. Some advantages are listed as abuse can be encountered in all health payments, in particular, the payments of social security institutions and insurance companies.

F. Clinical Decision Support Systems

There is a need for DW-oriented databases to help the diagnosis and treatment of the patient's problems and for systems that can be provided during diagnostic or treatment [50] depending on the data, solution, risk and recommended automation. As to solutions; systems that can work on electronic patient records and present the methods according to the needs of the physicians in a way that will respond to the request without going into the theoretical confusion should be designed. In other words, it is the development of decision support systems where appropriate DM methods are hidden behind the users graphical interface [51]. This brings some advantages such as the intelligent system will have easy-to-use and early-warning features that can be instantly accessed, analyze and correlate variables instantly.

G. Improving the Quality by Patient-Focused Health Service

Quality of healthcare is a concept influenced by many factors and varies according to the perception level of the patient. Therefore, when determining quality indicators, variables such as disease groups, patient demographics, patient's insurance status, clinical and service quality should be considered together and evaluated [52]. Key components analysis, factor analysis, logistic regression and decision tree algorithms can be applied by combining the variables obtained with a questionnaire of patient and management opinions [53], with administrative data via key variables, without reducing the precedence with Reliability and Question Analysis. All variables affecting quality can be handled together, and the quality variables for each focus group can be individually determined by automatically clustering according to the patients, disease or target clusters.

H. Risk Analysis to Optimize Service Delivery

Optimization of service delivery is required to create optimal service components for institutions and to be able to work effectively for resource budgeting. The theoretical definition of the risk is the expected value of the loss function [54]. Therefore, risk identification is also defined as; Model identification. Risk indicators can be considered. When it comes to advantages it can be concluded that all variables that constitute components for service providers can be handled together risk factors can be determined.

I. Abuses Detection and Billing Corruptions

As happened in all areas, abuse in healthcare is quite widespread, and most of the burden on the economy is due to the fact that the services are based on public finance. However, since there is a system based on cash, it is difficult to detect abuses such as bill corruption and so on. Developments in information technologies have led to the adoption of classical human-based inspection methods by automation-based surveillance and control systems [55]. Priority has come to mean systems for detecting and preventing potential risks with smart algorithms begun to place human-made computations and inquiries into systems supported by information technology and almost think and decline. Abuses can be evaluated as outliers or extreme values. With DM methods abuses will decrease and stop. DM in healthcare will bring some new features as using the descriptive statistics, outlier detection is determined subjectively on an observation basis [52]. However, for an objective decision-making process, a norm that can be accepted by everyone is required to be valid in the scientific sense.

J. Factors Influencing Firms and Minimizing Cost

Determination of service costs is very important in terms of providing and purchasing health services. Healthcare providers need to know the factors that affect the cost of the services and to minimize the costs in order to control the costs. The relationships between them can be identified by identifying those who have a significant influence on all potential cost factors. In addition, not only the overall sense but also the decomposition of subgroups can be detected [56]. Decision trees, Key Component Analysis or Factor Analysis can be used for this purpose. In addition to fixed and variable costs, it is possible to evaluate the other variables together and to determine the effect levels of the variables.

K. Early Financial Warning Systems

It is necessary to determine the financial performance and determine the financial risks in terms of efficient use of the resources allocated to the health sector in all of the hospitals and health institutions. The most practical method that can be used to take precautions without financial performance deterioration or financial crisis is financial early warning systems. A profiling approach should be followed to determine the norm for financial performance [57]. Decision Tree can be used for this purpose. By preventing the managers from being confused in financial variables, it offers much more objective results than other statistics and financial methods. It also provides access to early warning signals using basic variables and risks.

L. Managerial Decision Support Systems

Healthcare supervisors need systems that best utilize available data and support decision-making processes to manage healthcare institutions more effectively, efficiently and without losing quality. In the definition of the required decision variable; Model identification, the determination of efficiency, good quality and risk indicators can be taken as a basis [58]. All variables that can be used for administrative purposes in hospitals can be considered as multi-dimensional, optimal values and road maps can be determined. This will provide efficient decision-making steps.

VIII. RESULT

In this study, it is aimed to analyze DM applications and usages in healthcare to display an infrastructure for DM and to give a new perspective to the health professionals in terms of decision-making processes by presenting examples about the use of DM in the health sector [59]. As understood it is vital to use DM methods in the health sector to have better solutions to the arising problems when compared to the traditional methods.

Other uses of DM in the health sector are; Monitoring of health employees' performance. Patients flow planning, Optimization of medical treatment processes, Identification of early warning signals for drug use errors and side effects, profiling of patients and drug use in chronic diseases, drug usage habits and risk identification based on DM, calculation of drug unit costs, determination of drug innovation costs, Establishing smart health databases against bioterrorism, Determination of priority and minimum costs in disaster compensation [60].

DM is a decision support tool that will enable health authorities to access the most accurate and up-to-date information, using the most objective and optimal solutions. DM, the digital decision making, and BI method of the future are recommended by the experts in the health sector in terms of more efficient presentation of health services, more efficient use of resources and scientific, comparable, transparent information access [61]. It can be summarized as DM applications will immensely contribute the healthcare industry though, some limitations are inevitable which, within time and with the help of future work, will reduce. Now the limitations are as the accessibility of data, as the raw inputs for DM repeatedly occur in different settings and systems, such as administration, clinics, laboratories. Overall the study will contribute to literature in a way; open a new horizon for future work.

REFERENCES

[1] Roski, J., G.W. Bo-Linn, and T.A. Andrews, *Creating value in health care through big data: opportunities and policy implications.* Health affairs, 2014. **33**(7): p. 1115-1122.

[2] Drummond, M.F., et al., *Methods for the economic evaluation of health care programmes*. 2015: Oxford university press.

[3] Jagadish, H., et al., *Big data and its technical challenges.* Communications of the ACM, 2014. **57**(7): p. 86-94.

[4] Kuna, H.D., R. García-Martinez, and F.R. Villatoro, *Outlier detection in audit logs for application systems.* Information Systems, 2014. **44**: p. 22-33.

[5] Young, K.M. and P.J. Kroth, *Sultz & Young's Health Care USA*. 2017: Jones & Bartlett Learning.

[6] Firouzi, F., et al., *Internet-of-Things and big data for smarter healthcare: From device to architecture, applications and analytics*. 2018, Elsevier.

[7] Ruiz, P.P., B.K. Foguem, and B. Grabot, *Generating knowledge in maintenance from Experience Feedback.* Knowledge-Based Systems, 2014. **68**: p. 4-20.

[8] Jothi, N. and W. Husain, *Data mining in healthcare–a review.* Procedia Computer Science, 2015. **72**: p. 306-313.

[9] Algur, S.P. and B.A. Goudannavar, *Managing and Mining Web Multimedia–B-Tree Indexing Approach.* 2016.

[10] Milovic, B. and M. Milovic, *Prediction and decision making in health care using data mining.* Kuwait Chapter of the Arabian Journal of Business and Management Review, 2012. **1**(12): p. 126.

[11] Kautish, S. and R.K.A. Ahmed, *A Comprehensive Review of Current and Future Applications of Data Mining in Medicine & Healthcare.* algorithms. **6**(8): p. 9.

[12] Durairaj, M. and V. Ranjani, *Data mining applications in healthcare sector: a study.* International journal of scientific & technology research, 2013. **2**(10): p. 29-35.

[13] Mangat, V., *Swarm intelligence based technique for rule mining in the medical domain.* International Journal of Computer Applications, 2010. **4**(1): p. 19-24.

[14] Jiang, F., et al., *Artificial intelligence in healthcare: past, present and future.* Stroke and Vascular Neurology, 2017: p. svn-2017-000101.

[15] Atzmueller, M., *Subgroup discovery.* Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery, 2015. **5**(1): p. 35-49.

[16] Hailu, T.G., *Comparing Data Mining Techniques in HIV Testing Prediction.* Intelligent Information Management, 2015. **7**(03): p. 153.

[17] Abbasi, A., S. Sarker, and R.H. Chiang, *Big data research in information systems: Toward an inclusive research agenda.* Journal of the Association for Information Systems, 2016. **17**(2).

[18] Witten, I.H., et al., *Data Mining: Practical machine learning tools and techniques*. 2016: Morgan Kaufmann.

[19] Kourou, K., et al., *Machine learning applications in cancer prognosis and prediction.* Computational and structural biotechnology journal, 2015. **13**: p. 8-17.

[20] Voyant, C., et al., *Machine learning methods for solar radiation forecasting: A review.* Renewable Energy, 2017. **105**: p. 569-582.

[21] Wang, Q., et al., *Addressing Complexities of Machine Learning in Big Data: Principles, Trends and Challenges from Systematical Perspectives*. 2017.

[22] Stimmel, C.L., *Big data analytics strategies for the smart grid*. 2016: Auerbach Publications.

[23] Hogenboom, F., et al., *A survey of event extraction methods from text for decision support systems*. Decision Support Systems, 2016. **85**: p. 12-22.

[24] Sen, R.S., et al., *Developing and Implementing the Data Mining Algorithms in RAVEN*. 2015, Idaho National Laboratory (INL), Idaho Falls, ID (United States).

[25] Tiwari, M. and R. Dixit, *Data Mining Principles, Process Model and Applications*. 2017: Educreation Publishing.

[26] McFee, B., *Statistical Methods for Scene and Event Classification*, in *Computational Analysis of Sound Scenes and Events*. 2018, Springer. p. 103-146.

[27] Larose, D.T., *Discovering knowledge in data: an introduction to data mining*. 2014: John Wiley & Sons.

[28] Holzinger, A. and I. Jurisica, *Knowledge discovery and data mining in biomedical informatics: The future is in integrative, interactive machine learning solutions*, in *Interactive knowledge discovery and data mining in biomedical informatics*. 2014, Springer. p. 1-18.

[29] Forkan, A., et al., *Bdcam: Big data for context-aware monitoring-a personalized knowledge discovery framework for assisted healthcare.* IEEE transactions on cloud computing, 2015.

[30] Draycott, T., I. Crofts, and I. Ash, *Simulation in Healthcare Education.* Health Informatics: An Interprofessional Approach, 2014: p. 468.

[31] Raghupathi, W., *Data mining in healthcare.* Healthcare Informatics: Improving Efficiency through Technology, Analytics, and Management, 2016: p. 353-372.

[32] George, J., V. Kumar, and S. Kumar. *Data Warehouse Design Considerations for a Healthcare Business Intelligence System*. in *World Congress on Engineering*. 2015.

[33] Greer, M., *A case study for integration of legacy healthcare electronic medical records into a trusted data warehouse utilizing master data management*. 2016, University of Arkansas at Little Rock.

[34] Visscher, S.L., et al., *Developing a standardized healthcare cost data warehouse.* BMC health services research, 2017. **17**(1): p. 396.

[35] Archenaa, J. and E.M. Anita, *A survey of big data analytics in healthcare and government.* Procedia Computer Science, 2015. **50**: p. 408-413.

[36] Sahni, N.R., et al., *The IT Transformation Health Care Needs.* HARVARD BUSINESS REVIEW, 2017. **95**(6): p. 129-136.

[37] Ekin, T., et al., *On the Use of the Concentration Function in Medical Fraud Assessment.* The American Statistician, 2017. **71**(3): p. 236-241.

[38] Islam, A. and T. Amin, *Data Mining Application for Finding Patterns: Survey of Large Data Research Tools.* American Journal of Neural Networks and Applications, 2017. **3**(2): p. 14-21.

[39] Hair Jr, J.F., et al., *The essentials of business research methods*. 2015: Routledge.

[40] Breiman, L., *Classification and regression trees*. 2017: Routledge.

[41] Zhang, Y., et al., *Health-CPS: Healthcare cyber-physical system assisted by cloud and big data.* IEEE Systems Journal, 2017. **11**(1): p. 88-95.

[42] Luo, J., et al., *Big data application in biomedical research and health care: a literature review.* Biomedical informatics insights, 2016. **8**: p. BII. S31559.

[43] Shukla, D., S.B. Patel, and A.K. Sen, *A literature review in health informatics using data mining techniques.* International Journal of Software and Hardware Research in Engineering, 2014. **2**(2): p. 123-129.

[44] Bates, D.W., et al., *Big data in health care: using analytics to identify and manage high-risk and high-cost patients.* Health Affairs, 2014. **33**(7): p. 1123-1131.

[45] Wang-Roveda, J.M., L. Powers, and K. Ren, *Pervasive Computing in Hospitals.* Wireless Computing in Medicine: From Nano to Cloud with Ethical and Legal Implications, 2016: p. 43.

[46] Gravenhorst, F., et al., *Mobile phones as medical devices in mental disorder treatment: an overview.* Personal and Ubiquitous Computing, 2015. **19**(2): p. 335-353.

[47] Ali, A., et al., *Big data for development: applications and techniques.* Big Data Analytics, 2016. **1**(1): p. 2.

[48] Bardy, G.H. and J.M. Bishay, *Health Monitoring Apparatus For Initiating A Treatment Of A Patient With The Aid Of A Digital Computer*. 2018, Google Patents.

[49] Xu, Y., M. Pinedo, and M. Xue, *Operational risk in financial services: A review and new research opportunities.* Production and Operations Management, 2017. **26**(3): p. 426-445.

[50] Iliff, E.C., *Computerized medical diagnostic and treatment advice system including network access*. 2015, Google Patents.

[51] Fernandez, D.S., *Integrated biosensor and simulation systems for diagnosis and therapy*. 2017, Google Patents.

[52] Bauder, R.A. and T.M. Khoshgoftaar, *Multivariate outlier detection in medicare claims payments applying probabilistic programming methods.* Health Services and Outcomes Research Methodology, 2017. **17**(3-4): p. 256-289.

[53] Butcher, H.K., et al., *Nursing Interventions classification (NIC)-E-Book*. 2018: Elsevier Health Sciences.

[54] Dangl, T., O. Randl, and J. Zechner, *Risk control in asset management: Motives and concepts*, in *Innovations in Quantitative Risk Management*. 2015, Springer. p. 239-266.

[55] Mühlbacher, A. and F.R. Johnson, *Choice experiments to quantify preferences for health and healthcare: state of the practice.* Applied health economics and health policy, 2016. **14**(3): p. 253-266.

[56] Hossein-Zadeh, I., *Beyond Mainstream Explanations of the Financial Crisis: Parasitic Finance Capital*. Vol. 186. 2014: Routledge.

[57] Berner, E.S. and T.J. La Lande, *Overview of clinical decision support systems*, in *Clinical decision support systems*. 2016, Springer. p. 1-17.

[58] Lian, J.-W., D.C. Yen, and Y.-T. Wang, *An exploratory study to understand the critical factors affecting the decision to adopt cloud computing in Taiwan hospital.* International Journal of Information Management, 2014. **34**(1): p. 28-36.

[59] Muscat, D.M., et al., *Qualitative insights into the experience of teaching shared decision making within adult education health literacy programmes for lower-literacy learners.* Health Expectations, 2017. **20**(6): p. 1393-1400.

[60] Popovska, K., V. Gerasimovska, and M. Zdravkovska, *Report from implementation of INNIC multidimension hand hygiene approach (IMHHA) among health care workers (HCWs) in nephrology ICU.* 2014.

[61] Stoddart, G. and R. Evans, *Producing health, consuming health care*, in *Why are some people healthy and others not?* 2017, Routledge. p. 27-64.