The Role of Open Standard Electronic Health Record in Medical Data Mining

Rajesh Khanal

Abstract—Electronic Health Record (EHR) has received significant attention of all the health service provider in the world. EHR contains electronic information of all the patient information such as demographics, medical history, family medical history, lab tests and results, and prescribed drug. There is not any consistency in type of the EHR software implemented by the hosting organization. So, the EHR is currently vendor dependent and is not transferrable to another health service provider. The open standard electronic health record makes it public available to both vendor and patient. It can further aid in creating a universal EHR database for medical data mining. Mining the EHR helps in developing the best standard of care and clinical practice. The following paper proposes a universal EHR database and medical data mining. The benefits and challenges of implementing a database system is also discussed in the paper. The following paper will also analyze the different application areas of the EHR data mining.

Index Terms—Electronic Health Record; Open Standard; Universal; Database; Medical Data Mining.

I. INTRODUCTION

Data mining is the process of analyzing big data sets to generate some useful information. According to Brown [1] Data mining (DM) is the process of studying large amount of data in order to find some implicit relationship between the data which reveals valuable information hiding in the data set. In today technological world, a huge amount of data is created in a small period of time from various data sources. It is not possible for the human being to analyze the large chunk of data. Data mining is the useful tool in such case[2]. The information generated by the data mining is not useful to everyone. So, it is very important to choose correct type of database sources. Usually the database which is reliable and less costly is chosen for the analysis. The data sources for the data mining can be flat files, data warehouses and database. The knowledge derived from data mining can be useful in different sectors such as business, marketing, medical research and government[3].

The data mining term is becoming more popular in health care sector too. The amount of data generated in medical care are too huge and complex in nature. It is almost impossible to analyze those data using the traditional analysis techniques. Data mining of patients' health record can provide useful insight in analyzing the health condition of the patient[4]. But analyzing the medical data is not an easy task due to its complexity. Cios and More [5] has mentioned some major uniqueness characteristics (Heterogeneity, ethical legal and social issues, statistical boundaries, and special status) in the

medical data. Different developed nation such as United States, Australia and United Kingdom have already established electronic health record systems[6]. Electronic Health Record (EHR) is a digital document which contains the list of patient's diagnosis and treatment measures. These records can be easily shared among different hospitals or medical care organizations. EHR will keep track of all the past diagnosis which can be helpful in proper diagnosis of patient health condition. The different attributes of EHR is shown in table below:

Electronic Health Record Structure

SEGMENT	EXPLANATION
No.	Number of physical examination users
Gender	Gender of physical examination users
Age	Age of physical examination users
Project	Physical examination projects
Indicator	The project indicators
Result	Physical examination results
Reference	Reference Results
Lower Bound	Physical indicators lower bound
Upper Bound	Physical indicators upper bound

Fig 1. Sample Electronic Health Record Structure

The term Project in the record structure refers to some physical examination such as Routine Pressure or Blood checkup. The indicators point to the project such as RBC count for Blood checkup. The description of other attributes is mentioned in the table. Some of the basic benefits of EHR are improved patient care, improved care condition, increase patient participation, improved diagnostics and cost savings[7].

The model of EHR implemented in different nation are in different in nature. The data used by one EHR is completely different from another EHR provider which does not support the cause of creating the EHR. Moreover, there are many synonyms of medical terminologies which is also one of the main problem in analyzing the health record[8].

II. TYPES OF DATA

A. Structured Data

Structured data refers to the organized data which can be easily used in the data analytics. The data are already stored in some definite manner. The data of the patient such as age, sex, test reports and so on are stored in structured table of the database[9]. Relational Database is one of the common example of structured data. The relational database is very to use and analyze. It can be directly fed as input for data analysis. If all the medical organization have same standard schema, then it is very easy in analyzing the data across the

universe to get some valuable insight of information[10].

B. Unstructured Data

Unstructured data are the data which are not organized in a logical manner. It is very difficult to understand the data stored in this format. There are different methods such as text mining, text analytics which can be applied to get valuable information from the unstructured data. Some of the common example of unstructured data are books, journal images, videos and so on[11]. The medical reports include the summary or the text of the diagnosis or the medical directives. Natural language processing tools is the most common tools for analyzing the text. Specific NLP methods should be designed which can interpret the medical notes and derive some useful information[12].

C. Semi-Structured Data

There is the type of the data which lies between the structured and unstructured data. Semi structure data model provides more flexibility in storing the data than the structured data model [13]. This type of model is more useful in EHR systems. The user can assign new attribute per his requirement. Some example of semi-structured data includes XML and JavaScript Object Notion.

Fig. 2. Types of Data with Examples[14]

III. UNIVERSAL ELECTRONIC HEALTH RECORD

A universal EHR can be solution to minimize the complexities of having different models of EHR. The data standard of the EHR should be open standard so that it is easily available to the public. The data and information of the patient can be easily moved to any organization or hosts. This ensures the health record is the property of the patient[15]. According to John[8], some of the major advantages of universal EHR to the patients are Consolidate, Comprehensive, Worldwide access, Patient Ownership and Ease in analysis by computer tools.

Fig. 3. Universal EHR and its relationship with its dependent variable.

A. Proposed Model of UEHR Database

According to Free Software Foundation Europe, an open standard is a format which is available for the general public use without any legal or technical constraints. It can be easily used by any business and modify or develop it per its requirement[16]. An open EHR will help in world wide access of the patient health record. As the universal EHR would contain all the medical history and keep track of other tests, reports and drugs use, it will be easier for doctor for medical diagnosis[8]. A data of the patient will be maintained in a database so that they can be easily accessed or analyzed[17]. A database is a pool of organized information which can be easily retrieved. A database contains many tables which are related to each other by some key[18].

“This paper proposes a universal EHR database which can be easily accessed and understood by the medical professional all around the world. The medical data of the patient is maintained using the open standard of the EHR. Each individual patient record will have a unique identifier which separates it from another. Similar to the Social Security Number or the passport number each individual record can be easily accesses by the patient and the authorized medical personnel[19].

Fig. 4. Sample Attributes of Universal EHR

IV. BENEFITS AND CHALLENGES OF UNIVERSAL EHR DATABASE

A. Benefits

Comorbidity Analysis - The universal EHR can help physicians in identifying the natural history of the disease. Comorbidity analysis is process of analyzing the relationship between diseases that occur concurrently[20]. A person suffering from diabetes might also suffer from high blood

pressure and high concentration of fats or lipids in the blood. Analyzing such different traits of diseases can lead to the new finding which can lead to the discovery of new treatment[21].

Cohort Identification - By proper analysis of Universal EHR database can aid in the cohort identification of the patient group fulfilling certain criteria. Lab tests, reports, medications etc. are the important attributes of EHR in cohort identification[22].

Risk Prediction - Data mining identifies the pattern in dataset which are either predictive or descriptive in nature. The predictive pattern helps in identifying the future value based on current data set whereas descriptive pattern describes the relationship among the different attributes present in the dataset[23]. Analyzing the EHR using logistic regression or Poisson regression can help in predicting the risk that might happen in the future. For example, a person who has abnormal hemoglobin can indicate a risk of having cardio-vascular problems[24]. The dataset can be thoroughly analyzed to identify the markers or indicators of certain diseases.

Predicting Future Complication - It is very hard to predict the future complication that might happen to the patient. The availability of a large uniform dataset help in analyzing the attribute which helps in prediction of some complexities that might arise in the future[25].

Evidence Based Instructions - These are the descriptive tools which can help medical officer in choosing the appropriate health care for some specific clinical circumstances. We need huge amount of data for preparing such guidelines. Different data mining techniques such as association rule mining, sequential rule mining and regression approaches can be useful in constructing those guidelines[26].

B. Challenges

Censored Data - The medical condition of the patient is only observed or recorded for only certain amount time. The other activities of the patient outside of the observation is not available. This can create problem in proper understanding of the patient health condition due to loss of some important information. This can lead to generate incorrect information from data mining[27].

No Uniform Data - According to American Diabetes Association, A1C test must be done at least 2 times a year for patient who has blood sugar level fairly stable at near-normal level. These recommendations are not followed by the patient in regular manner. These irregular collections of data can lead to irregular longitudinal data[28].

Missing Data - Some of the diagnosis data of the patient is omitted from the health report. There is no any conformity that the patient takes the medication prescribed by the doctor. The informal advices given by the doctor is also not included in the report. A doctor asks patient to take some tests analyzing his medical condition. The reason for not asking some other test is not mentioned in the report[29]. For example, when a person has an injury and bacterial test has been ordered then it can indicate that a patient might be suffering from infection.

Heart Association is used to show the efficacy of the EHR data mining. The simple demographic information of the patient such as gender, age, height and weight is kept in the database. The lab test values such as systolic blood pressure and Diastolic blood pressure of the patient is also recorded. The person medical history and his family health history is also recorded.

Diabetes Risk Test

1 How old are you? (Less than 40 years (0 points), 40–49 years (1 point), 50–59 years (2 points), 60 years or older (3 points))

2 Are you a man or a woman? (Man (1 point), Woman (0 points))

3 If you are a woman, have you ever been diagnosed with gestational diabetes? (Yes (1 point), No (0 points))

4 Do you have a mother, father, sister, or brother with diabetes? (Yes (1 point), No (0 points))

5 Have you ever been diagnosed with high blood pressure? (Yes (1 point), No (0 points))

6 Are you physically active? (Yes (0 points), No (1 point))

7 What is your weight status?

Write your score in the box: []

Fig. 5. Diabetes Risk Prediction Tool (American Heart Association)

For each of the attributed mentioned in this assessment, some point is given. The points range from 0 to 3 according to the questions mentioned in the risk test. The weight status point is given according to the body mass index of the patient. After each attributed is assigned some points, the total score is calculated. If the total score is more than 5 then the patient is at more risk of Type 2 diabetes. The assessment tool was made by analyzing the data of different patients.

Fig. 6. Sample Database schema for Patient record

V. RESEARCH METHODOLOGY

For this journal, the simple tool prepared by the American

VI. RELATED STUDY

A database is a collection of data and information which can be easily utilized to gather some useful information. The world is moving towards the computerization of each and every data and information[30]. EHR is one of the recent development in the medical field which helps in patient's improved care. EHR an electronic module which contains all the medical history of the patient is popular in most of the developed countries in the world[31]. United Kingdom, Australia, Sweden and Canada are some of the countries which have already adopted the Electronic Health Record and working towards further minimizing the issues related to it[32]. Interoperability between the EHR systems is one of the major problems associated with the implementation of the electronic health record[8].

The primary objective of the universal electronic health record is to make it publicly available[8]. Implementing open standard of electronic health records has various advantages to the patient also. In the market, we can find many types of electronic health record. There is not any uniformity in the type of the electronic health record system deployed by the medical organizations. A failure of communication is one of the major issues due to different kinds of EHR system deployed in the hospitals and clinics. The electronic data from one clinic cannot be moved to another clinic if they have implemented different types of vendor[33].

According to openEHR, the health information of the patient should be individual dependent and vendor independent. There are numerous incentives provided to the health service providers in the United States who implement digital medical records. These incentives are provided to improve overall health of the patient[34]. But there are a lot of problem due to lack of common standard of the electronic health record. A common open standard of the electronic health record if implemented has many advantages. The application of different data mining techniques would be easier than now[35]. The decision-making process can be assisted with the help of analysis of electronic record. The medical professional uses their knowledge and experience to diagnosis certain disease. Sometime the treatment process can go wrong due to many reasons. The result from data mining can provide some details about clinical decision making to the medical professionals[36]. The patient data can be compared with the records present in the database to find out possible diseases that might happen in future. The biomedical and the molecular data can also help in researching about some diseases. Microarray data analysis and clustering techniques are some of the mining techniques that are used by the data analyst for molecular data analysis. The clustered micro array data can be useful in predicting the vulnerable groups which could get affected by the disease in the future[37]. The type, pattern or cause of some diseases can be analyzed according to the geographical area so that proper measures can be taken to eradicate the disease. Data mining techniques has been used in health data to analyze the certain traits which affects the cancer patients[38].

Jyoti Soni [39] has illustrated the use of data mining techniques in the prediction of heart problem. They concluded that the result from Bayesian classification and decision tree has same amount of accuracy[39]. Satej [40] has shown some examples for tumor detection using various mining algorithm such as classification, neural networks and association. The use of the data mining techniques was illustrated by "Survey on Utilization of Data Mining

Approaches for Dermatological (Skin) Diseases Prediction" for the prediction of various skin diseases by them.

The data gathered from electronic health record is of specific type. Since the data is related to the medical field, proper assistance of the medical professional is continuously required during the analysis process[20]. The amount of data from electronic health record database is very large in amount and the structure can also be of any type. We have many big data technologies which can deal with large volume and variety of data. There are many existing data mining techniques which can analyze health record efficiently. Many new algorithm are also in the research process to provide better analysis of the health record[41]. In general, data are aggregated at beginning. The data is then maintained and integrated in to the system for data analysis. With the help of data analysis certain pattern are interpreted which can aid in decision making[41]. Association mining has been widely used in cancer diagnosis. Classification is also another popular method in the medical data mining. Clustering is also another major mining technique in medical image mining. Some techniques such as neural network, genetic algorithm and fuzzy classification are used to discover the pattern among the different data. There is a lot of data analysis done in the field of the breast cancer patients. The survivability of the cancer patient can be analyzed using Naives Bayes techniques. Using some simple medical data such as age, sex, blood pressure and family medical history a person's vulnerability of the heart diseases can be found out. American Heart Association uses this simple technique to find out the person vulnerability to the heart risk and diabetes. These assessment tools are found in internet easily through which any person can measure his risk of these diseases.

Apriori algorithm is the oldest algorithm in analyzing the frequent datasets. This method can help in predicting the rate of patients affected by the diseases. The prevalence of disease in some geographical area can be analyzed for different time period[42]. The researcher used this algorithm and WEKA tools to find out the frequency of diseases and their location using the electronic health record dataset from medical institution[43].

VII. RESULTS AND FINDINGS

The overall efficacy of the EHR mining was determined by conducting the survey with the stakeholders who directly deal with the Electronic Health Records on daily basis. The Google form is one of the major tool used in the survey process. The questions of the survey were posted in the Google form and responses were recorded online. The survey members were five medical professionals from Chitwan Medical College and Hospital, Nepal. Three of them were general physician and remaining two were nurse. All the members were asked a same set of questions.

All of the agreed on the open standard of the Electronic Health Record. They also responded that EHR should be patient dependent. The data of the patient should belong to the patient and they should be free to choose any vendor for medical services. As of now, there are three open standard of EHR which are very popular in market. They are openEHR, Virtual Medical Record and SMART (Substitutable Medical Apps, reusable technologies). We asked the respondents if they were aware about them. According to survey, only 80% of the respondents were aware about openEHR and 40% of them were aware about Virtual Medical Record. The EHR

software used in the Chitwan Medical College and Hospital was Dr. MIDAS. The respondents were also confident in using the EHR software for entering patient details. Majority of the respondent felt their documentation was complete and accurate as it is on paper method. All the respondents had the experience of more than 12 months of dealing with the patient directly. The figure displays the major advantages of the open standard EHR pointed out by the respondents.

Fig. 7. Respondent's view on advantages of Open EHR

Data Mining would be easier when we implement the open standard of EHR. The same algorithm and technique can be applied to the large set of data with similar attributes which help in correct analysis of the data. We can achieve larger dataset which will be helpful in data training procedure. The data model will also include wide variety of scenarios. The application areas of Health data mining pointed out by the respondent is shown in figure below:

Fig. 8. Application areas of EHR mining

The respondents were very positive about the outcomes of EHR mining. They were using VisualDx tool to aid them in proper clinical decision. VisualDx is a software that runs both on PC and Tablets has the large database of wide variety of diseases, their types and their attributes. The symptoms and medical history of the patient entered in to the tool and the proper suggestion based on previous medical case is provided by the tools. Some of the common support provided by these decision support systems are shown in figure below.

Fig. 9. Common Usage of Decision Support Systems

VIII. FUTURE DIRECTION

The open standard EHR is the vendor independent form of medical record which can be used across the world without any compatibility issues. The open aims to mitigate the issues that arises due to the lack of interoperability between the different software vendor of EHR. The open standard of EHR can leads us towards the development of distributed universal EHR database. The medical data of the patient can be accessed irrespective of the geographical location. The patient can also change the health service provider according to his choice. The medical data of the patient can move with him to different service provider. The data in the Universal EHR database can only be uploaded and updated by the authorized person of the health service provider. The security of the data should also be given more priority. The actual database should be separated from the web servers. The data stored in the database can also be encrypted using different techniques. Apart from the actual database, the backup database should also be encrypted. This will help in preventing the patient data from identity theft. The web application firewall tools can be used to prevent from different SQL injection attacks.

There are numerous tools available in market which are used in mining of large amount of data. Mining of universal EHR can provide us new insight about the use of medical data mining has been proved effecting in treating the breast cancer patient. Moreover, it also helps in making clinical decision. The mining tool can be integrated with the database system to provide the alerts about the risk the patients are vulnerable to. The automated alert can help medical professional in doing job more efficiently. This might can decrease the overall error rate in medical practice. The medical history of one patient can be used as basis of treatment for another patient with similar symptoms. The health service provider can also choose among different health practices and choose the best one effective for the patient. This will overall help in developing the best standard of care and clinical practice.

IX. CONCLUSION

This paper presents the overall advantages of the open standard Electronic Health Record. The study reveals the need of open standard in Health Record throughout the world. This will transfer the complexity of the patient to the vendor.

The medical data of the patient will be the sole property of the patient and it can move with him to any hosting vendor. The idea of having universal EHR will be the stepping stone for creating the universal EHR database which can be universally accessed. The universal EHR database will aid in the effective analysis of medical data for better clinical decisions. Different data mining tools such as Weka, Orange, Rapid Miner, R and so on can be utilized to get new insight of the disease and clinical procedure. Some common use of medical data mining is drug-drug interaction, drug allergy interaction, dose range checking and relevant data displays.

APPENDIX

This survey is the part of academic project to find the role of open standard Electronic Health Record in medical data mining. This survey will take less than 5 minutes and covers the basic questions related to open standard of EHR and its usefulness in data mining. This paper also proposes in creating a universal electronic health record database which can be accessed universally.

Open Standard Electronic Health Record(EHR)

- 1) Role at the practice (please check the appropriate box):
 - A. Physician
 - B. Midlevel provider
 - C. Clinical staff (nurse, medical assistant).
 - D. Other: _____
- 2) Do you regularly work with patients in this hospital?
 - A. Yes
 - B. No
- 3) How long have you used the EHR (in Months)?
- 4) My documentation is as complete and accurate as it was on paper/using dictation.
 - A. Strongly disagree
 - B. Disagree
 - C. Agree
 - D. Strongly agree
- 5) EHR should be
 - A. Vendor Dependent
 - B. Patient Dependent
- 6) Do you support EHR Open Standard?
 - A. Yes
 - B. No
 - C. Maybe
- 7) What do you think will be the major advantage of EHR Open Standard?
- 8) Have you heard about following open standards of EHR?
 - A. openEHR
 - B. Virtual Medical Record
 - C. SMART
 - D. None of the Above
- 9) What is the name of the EHR solution software used in your organization?

Medical Data Mining Application Areas

- 1) Do you think Health Data Mining can help you in clinical decision making?
 - A. Yes
 - B. No
 - C. Maybe
- 2) Do you use any tools for helping in clinical decision? (If yes, please mention the name of the tools.)
- 3) What do you think are the major application areas of Electronic Health Data mining?
 - A. Understanding the Natural History of Disease
 - B. Risk Prediction
 - C. Predicting the next complication: What and When
 - D. Constructing Evidence Based Guidelines
 - E. Adverse Event Detection
 - F. Other: _____
- 4) Please list some of the use of Clinical Decision Support systems.

REFERENCES

- [1] Brown, M. Data mining techniques. 2012; Available from: <https://www.ibm.com/developerworks/library/ba-data-mining-techniques/>
- [2] Goyal, N., et al. A High Performance Computing Framework for Data Mining. in 2016 IEEE 23rd International Conference on High Performance Computing Workshops (HiPCW). 2016.
- [3] Algwaiz, A., S. Rajasekaran, and R. Ammar. Data mining using Probabilistic Grammars. in 2016 IEEE International Symposium on Signal Processing and Information Technology (ISSPIT). 2016.
- [4] Sharma, R., S.N. Singh, and S. Khatri. Medical Data Mining Using Different Classification and Clustering Techniques: A Critical Survey. in 2016 Second International Conference on Computational Intelligence & Communication Technology (CICIT). 2016.
- [5] Cios, K.M., William G, Uniqueness of medical data mining. Artificial Intelligence in Medicine, 2002. 26.
- [6] Butt, N. and J. Shan. CyberCare: A Novel Electronic Health Record Management System. in 2016 IEEE First International Conference on Connected Health: Applications, Systems and Engineering Technologies (CHASE). 2016.
- [7] Wei, X., W. Zou, and S. Gao. Mining Electronic Physical Records, a trial. in 2016 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining (ASONAM). 2016.
- [8] John, M. Moving Toward an Open Standard, Universal Health Record. 2010; Available from: <http://www.smart-publications.com/articles/moving-toward-an-open-standard-universal-health-record/page-2>.
- [9] McCALLUM, A. Distilling Structured Data from Unstructured Text 2005; Available from: http://delivery.acm.org/10.1145/1110000/1105679/p48-mccallum.pdf?ip=173.3.97.41&id=1105679&acc=OPEN&key=4D4702B0C3E38B35%2E4D4702B0C3E38B35%2E4D4702B0C3E38B35%2E6D218144511F3437&CFID=917918683&CFTOKEN=36526551&__acm__=1490891086_b750e9b0617beead0c5ac079a014ceef.
- [10] Tatarinov, I., et al., Storing and querying ordered XML using a relational database system, in Proceedings of the 2002 ACM SIGMOD international conference on Management of data2002, ACM: Madison, Wisconsin. p. 204-215.
- [11] Gharehchopogh, F.S. and Z.A. Khalifelu. Analysis and evaluation of unstructured data: text mining versus natural language processing. in 2011 5th International Conference on Application of Information and Communication Technologies (AICT). 2011.
- [12] Gardner, J. and L. Xiong, An integrated framework for de-identifying unstructured medical data. Data & Knowledge Engineering, 2009. 68(12): p. 1441-1451.
- [13] Jiawei Han, J.P., Micheline Kamber, Data Mining: Concepts and Techniques. 3rd ed2011.

- [14] Babu, M.C., Souwmya. Claims Fraud: A Big Opportunity for Big Data & Analytics. 2013 July 29, 2013; Available from: <http://www.claimsjournal.com/news/national/2013/07/29/233805.htm>
- [15] Schiller, D. EHRs and healthcare interoperability: The challenges, complexities, opportunities and reality. 2015; Available from: <http://www.healthcareitnews.com/blog/ehrs-healthcare-interoperability-challenges-complexities-opportunities-reality>.
- [16] Europe, F.S.F. Open Standards. 2008; Available from: <https://fsfe.org/about/basics/freesoftware.en.html>.
- [17] Flores, A.E. and V.M. Vergara. Functionalities of open electronic health records system: A follow-up study. in 2013 6th International Conference on Biomedical Engineering and Informatics. 2013.
- [18] Vieira, M. and H. Madeira. Towards a security benchmark for database management systems. in 2005 International Conference on Dependable Systems and Networks (DSN'05). 2005.
- [19] Banerjee, S., The Role of Global Educational Database in Educational Data Mining. *European Journal of Engineering Research and Science*, 2016. 1(6).
- [20] Pranjul Yadav, M.S., Vipin Kumar & Gyorgy Simon, Mining Electronic Health Records (EHR): A Survey, in Department of Computer Science and Engineering 2015, University of Minnesota.
- [21] William P. Castelli , J.T.D., Tavia Gordon , Curtis G. Hames , Marthana C. Hjortland , Ph. D , Stephen B. Hulley , Abraham Kagan , William , J. Zukel, HDL cholesterol and other lipids in coronary heart-disease - cooperative lipoprotein phenotyping study. *Circulation*, 1977. 55(5).
- [22] Jensen, P.B., L.J. Jensen, and S. Brunak, Mining electronic health records: towards better research applications and clinical care. *Nat Rev Genet*, 2012. 13(6): p. 395-405.
- [23] Ng, K., et al., PARAMO: A Parallel Predictive Modeling Platform for Healthcare Analytic Research using Electronic Health Records. *Journal of biomedical informatics*, 2014. 48: p. 160-170.
- [24] Tran, T.P., Dinh Luo, Wei & Venkatesh, Svetha, Stabilized sparse ordinal regression for medical risk stratification. *Knowledge and information systems: an international journal*, 2015. 43(3): p. 28.
- [25] George Hripcsak, D.J.A., and Adler Perotte, Parameterizing time in electronic health record studies. *Journal of the American Medical Informatics Association*, 2015.
- [26] Lohr, M.J.F.K.N., Guidelines for Clinical Practice: From Development to Use. National Academies Press, 1992.
- [27] Frydman, H., Nonparametric estimation of a Markov 'illness-death' process from interval-censored observations, with application to diabetes survival data. *Biometrika*, 1995. 82(4): p. 773-789.
- [28] Ramakrishnan, N., D. Hanauer, and B. Keller, Mining Electronic Health Records. *Computer*, 2010. 43(10): p. 77-81.
- [29] Padhukasahasram, B., et al., Joint Impact of Clinical and Behavioral Variables on the Risk of Unplanned Readmission and Death after a Heart Failure Hospitalization. *PLOS ONE*, 2015. 10(6): p. e0129553.
- [30] Kim, W., Object-oriented databases: definition and research directions. *IEEE Transactions on Knowledge and Data Engineering*, 1990. 2(3): p. 327-341.
- [31] Pereira, R., et al. Usability of an electronic health record. in 2012 IEEE International Conference on Industrial Engineering and Engineering Management. 2012.
- [32] Rihab A. HASANAIN, K.V.a.M.C., Electronic Medical Record Systems in Saudi Arabia: Knowledge and Preferences of Healthcare Professionals *Journal of Health Informatics in Developing Countries*, 2015. 9(1).
- [33] Palma, G. Electronic Health Records: The Good, the Bad and the Ugly. 2013; Available from: <http://www.beckershospitalreview.com/healthcare-information-technology/electronic-health-records-the-good-the-bad-and-the-ugly.html>.
- [34] Porter, M., Adoption of Electronic Health Records in the United States. Kaiser Permanente, 2013.
- [35] Ajami, S. and T. Bagheri-Tadi, Barriers for Adopting Electronic Health Records (EHRs) by Physicians. *Acta Informatica Medica*, 2013. 21(2): p. 129-134.
- [36] Gagnon, M.-P., et al., Multi-level analysis of electronic health record adoption by health care professionals: A study protocol. *Implementation Science*, 2010. 5(1): p. 30.
- [37] Selvaraj, S. and J. Natarajan, Microarray Data Analysis and Mining Tools. *Bioinformatics*, 2011. 6(3): p. 95-99.
- [38] Lisboa, P.J.G., et al., Data Mining in Cancer Research [Application Notes]. *IEEE Computational Intelligence Magazine*, 2010. 5(1): p. 14-18.
- [39] Jyoti Soni, U.A., Dipesh Sharma and Sunita Soni, Predictive Data Mining for Medical Diagnosis: An Overview of Heart Disease Prediction. *International Journal of Computer Applications*, 2011. 8(8).
- [40] Wagle, S., J.A. Mangai, and V.S. Kumar. An improved medical image classification model using data mining techniques. in 2013 7th IEEE GCC Conference and Exhibition (GCC). 2013.
- [41] Gotz, D., et al., Visual Cluster Analysis in Support of Clinical Decision Intelligence. *AMIA Annual Symposium Proceedings*, 2011. 2011: p. 481-490.
- [42] Zhang, C., Zhang, Shichao, Association Rule Mining. *Models and Algorithms 2002*: Springer-Verlag Berlin Heidelberg.
- [43] Robu, R. and C. Hora. Medical data mining with extended WEKA. in 2012 IEEE 16th International Conference on Intelligent Engineering Systems (INES). 2012.