

Using Data Mining to Improve Healthcare Decision-Making: A Systematic Review

Amit Kumar Mishra

Department of Computer Science & Information Technology,
Graphic Era Hill University, Dehradun Uttarakhand India 248002

Abstract

Data mining is a process utilized to analyze large databases to find patterns and relationships. It involves using various techniques such as machine learning and statistical analysis to identify trends and patterns. Due to the massive amount of medical data that is available, data mining has become an increasing trend in healthcare. This process involves extracting knowledge from these data sources. This review aims to analyze the current state of the research regarding the use of data mining in the healthcare industry. This review thoroughly searched for studies related to various aspects of healthcare diagnosis and prognosis. The results indicated that data mining techniques can help improve the accuracy of diagnoses and the prediction of treatment outcomes. Although the review acknowledged the potential of data mining in improving the quality of healthcare, it also highlighted various issues that need to be resolved in order for it to be successful. These include the need for standardization and privacy concerns. Despite these obstacles, the potential of this technology is still immense. The findings of this review provide a comprehensive analysis of the current state of the art in data mining in healthcare. It shows the potential of this technology to improve the quality of healthcare and inform clinical decisions. Future research must address the issues identified in this review in order to advance the use of this technology in the field.

Keywords: prognosis, technology, comprehensive, diagnoses, healthcare industry

Introduction

Data mining is a process utilized to extract meaningful information from vast data sets. The main objective of this process is to find hidden relationships and insights that can be utilized to make educated decisions. There are many applications of data mining in various fields, such as finance, healthcare, marketing, pattern recognition and business[1]–[3].

- In business intelligence, data mining is used to identify trends and behaviors in customers. This process can then be utilized to develop effective marketing and sales strategies.
- In fraud detection, data mining can help identify patterns of anomalous behavior that can lead to financial losses.
- In the healthcare industry, data mining can help improve the quality of patient care by identifying potential disease risk factors and developing new treatment methods.

There are many data mining applications in healthcare, and some of the most common ones are in the areas of disease risk assessment and treatment.

- In the diagnosis of diseases, data mining can be utilized to identify patterns in patient data that can help with the identification of various conditions. By analyzing vast amounts of information, the process can identify risk factors, symptoms, and other characteristics associated with certain diseases.
- In the discovery of drugs, data mining is also used to identify potential targets and develop new treatments. Through the process, it can analyze large amounts of information and find compounds that can be used to create new drugs.
- In the field of patient care, data mining can be used to identify patterns in the data that can help improve the quality of care. Through the process, healthcare providers can then provide more personalized treatment plans. Data mining has various applications in different fields, and it can be used in healthcare to improve the quality of patient care and develop new treatments. Due to the increasing popularity of data mining in healthcare, the need for more research on its use in decision-making has become more prevalent. This review aimed to identify the current research on its use in the field. The findings of the review revealed that data mining can help improve the diagnosis, prognosis, and treatment of various diseases. One study revealed that it can help predict the likelihood of patients developing diabetes. Another study revealed that it can help identify high-risk individuals who are more prone to hospital readmission[4]–[6].

The review identified various challenges that affect the use of data mining within healthcare. One of these is the lack of standardization within the collection of data. This can lead to the creation of inconsistencies and errors in the results of the analysis. In addition to this, specialized skills are also needed to properly interpret and implement the techniques involved in data mining. Despite the various challenges that it faces, data mining can still provide valuable and beneficial services to healthcare. The findings of this review suggest that it can help improve the quality of healthcare and inform clinical decisions. However, it also highlighted the need for further research to improve its use. The findings of this review provide valuable information on the current research regarding the use of data mining technology in healthcare. It can help develop new strategies and procedures that will benefit the healthcare industry.

Literature Review

The process of data mining involves identifying patterns and relationships in large datasets. This is useful in healthcare due to its ability to extract meaningful information from medical records and other sources. It can also help improve the diagnosis and treatment of diseases.

Kaur et al.[7] explored the use of data mining to develop policies that can improve the efficiency of healthcare systems in Washington DC. The study used data collected from patient records to identify trends and patterns that can be used to improve the management of healthcare. The researchers concluded that by utilizing data mining, healthcare providers can make informed decisions and develop customized interventions to improve the quality of patient care.

Baitharu et al.[8] analyzed the various techniques used in data mining for healthcare systems for the prediction of liver disorders. They found that neural networks and decision trees were the most effective tools for performing this process. They also emphasized the importance of selecting the right features for developing accurate models. Qiao et al.[9] analyzed the various techniques used in data mining to identify process abnormalities and improve the efficiency of a healthcare facility. The findings showed that the process could be utilized to enhance management and operations within the healthcare sector.

Malik et al.[10] analyze the various advantages and challenges of data mining in the healthcare industry. The researchers found that it can help improve the diagnosis of various diseases and develop new treatments. But, it can also expose patients to harmful effects.

In a study, Shen et al.[11] proposed a framework that would allow healthcare organizations to collect and analyze data from various sources to improve the delivery of healthcare services. They noted that this method could help improve the quality of care for patients from different regions and countries.

Ahmed et al.[12] presented a data summarization method that can help healthcare organizations improve the efficiency of their data mining efforts. The researchers noted that this method can help them identify rare events in the data collected.

Abouelmehdi et al.[13] discussed the importance of protecting the privacy and confidentiality of healthcare data. They noted that it is crucial to maintain the integrity, confidentiality, and availability of such information. To ensure that such data is secure, the authors recommend utilizing advanced security techniques and technologies.

Shouman et al.[14] utilized data mining techniques to diagnose and treat heart disease. They found that these techniques could help improve the quality of care and provide a more accurate and timely diagnosis. The researchers used a combination of statistical and computational methods, such as decision tree and the artificial neural network, to analyze the data collected from patients. The researchers noted that these techniques could help physicians make more informed decisions and improve the quality of care.

In 2016, Chadha et al.[15] conducted a study that analyzed the use of data mining techniques to predict the likelihood of heart disease. The researchers used a large dataset consisting of various clinical data points, such as medical history, gender, and age. The findings of the study revealed that the techniques can accurately identify individuals who are at risk of developing heart disease.

K. K et al.[16] discussed the various aspects of data mining in healthcare in order to improve the quality of patient care and prevent heart attacks. They presented an overview of the techniques used in the study, such as clustering and classification. The researchers concluded that these techniques can help improve the outcomes of patients by identifying those who are at high risk of experiencing heart attacks.

Ertl et al.[17] presented a study that analyzed the organic substituents using cheminformatics. The author was able to identify the most common substituent groups, calculate their properties, and identify bioisosteric groups that are similar to those of drugs. The study demonstrates how cheminformatics can be used in the design and discovery of new drugs. The author concluded that the techniques can help in identifying new drug targets.

These studies show the importance of utilizing the latest security protocols and techniques to safeguard the confidentiality and integrity of healthcare data. They also highlight how data mining techniques can help in identifying individuals who are at high risk of developing heart disease and how cheminformatics can be utilized in the design and discovery of drugs. The outcomes of these investigations provide valuable information that can be utilized to improve the treatment and discovery of new drugs. The potential of data mining to improve the diagnosis and treatment of diseases in healthcare has been acknowledged. However, its implementation in the field requires the proper algorithms and high-quality data. More research is needed to develop effective data mining techniques in the healthcare industry.

Various stage of extracting information from data mining

Data mining in healthcare involves several stages in extracting information from large datasets. These stages include as shown in figure.1:

- Data collection: This involves the collection of large datasets from various sources, such as electronic health records, medical devices, and clinical trials.
- Data preprocessing: This stage involves cleaning and transforming the raw data to eliminate errors and inconsistencies. This step involves data cleansing, data integration, data transformation, and data reduction.
- Data analysis: This stage involves applying various data mining techniques, such as classification, clustering, association rule mining, and regression analysis to identify patterns and relationships in the data.
- Interpretation and evaluation: This stage involves interpreting the results obtained from data analysis and evaluating the usefulness and effectiveness of the data mining process.
- Knowledge representation: This stage involves representing the knowledge obtained from data mining in a form that can be easily understood by healthcare professionals, such as decision trees, rules, and graphs.
- Application of results: The final stage involves applying the results obtained from data mining to improve the quality of healthcare. This may involve developing predictive models for disease diagnosis and treatment, identifying risk factors for certain diseases, and improving patient outcomes.

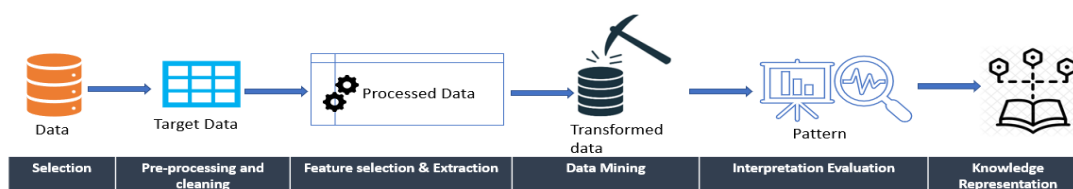


Figure 1 Various stage of data mining in healthcare

Various areas where datamining is applied in healthcare

The application of data mining in healthcare has the potential to revolutionize the way healthcare is delivered by providing actionable insights to doctors and healthcare providers. It can help doctors identify patients at risk of developing certain diseases and develop personalized treatment plans based on their unique medical history. Additionally, data mining can help healthcare providers identify patterns and trends in large datasets that would be difficult to detect through traditional methods. Few major areas where data mining plays a significant role are depicted in figure .2. Figure-3 and table-1 represent the accuracy of various medical conditions when datamining is implemented.

- Clinical decision[5]- Data mining is a process that involves extracting and analyzing large clinical datasets to improve the decision-making process for clinicians. It can help them identify potential risks and improve the treatment options for their patients. Electronic health records (EHRs): Through the use of electronic health records, data mining can help improve the quality of care and reduce the cost of treating patients. For instance, it can identify individuals who are at high risk of developing conditions and monitor their medication usage.

- Fraud detection[18]: In fraud detection, data mining can help identify healthcare providers who overcharge or bill for unnecessary procedures.
- Patient safety and quality improvement[19]: Healthcare organizations can use data mining to identify quality improvement opportunities and safety threats. It can then analyze incidents related to patient safety and medication errors.
- Predictive modeling[20]: A predictive model is created using data mining to identify individuals who are at risk of experiencing adverse events or developing conditions. It can also help predict which patients will be admitted to the hospital and which ones will benefit from preventive measures.
- Public health research: Large public health datasets are analyzed using data mining to find patterns and trends. This process can also help determine the effectiveness of certain public health initiatives.



Figure 2 Various area where data mining involves

Table 1 Accuracy of various medical condition[21]

Type of disease	Data mining tool	Algorithm	Accuracy
Heart disease	ODND, NCC2	NAİVE	60
Cancer	WEKA	Rule Decision Tree	97.77
HIV/ AIDS	WEKA	J48	81.8
Blood Bank Sector	WEKA	J48	89.9
Brain Cancer	K-mean Clustering	MAFIA	85
Tuberculosis	WEKA	KNN	78
Diabetes Mellitus	ANN	C4.5	82.6
Kidney Dialysis	RST	Decision Making	75.97
Dengue	SPSS	C5.0	80
IVF	ANN, RST		91
Hepatitis	SNP	Decision Tree	73.2

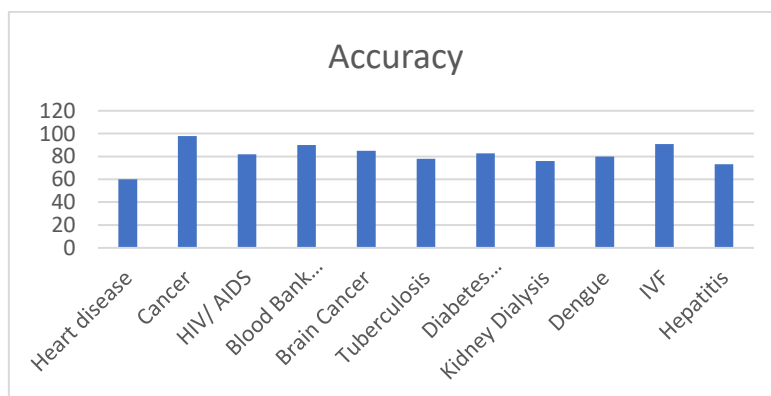


Figure 3 Accuracy of various medical conditions

Potential of data mining in improving the quality of healthcare

Through data mining, a powerful technology that can improve the quality healthcare, one can identify patterns and trends in medical data, which would previously have been impossible to do. This information can be utilized to develop new treatments and improve the current ones. Through data mining, healthcare providers can receive timely and accurate information, which can help improve the quality of patient care. For instance, it can help identify individuals at high risk of developing a particular disease. This information can also be utilized to develop effective treatment methods[22], [23].

Through data mining, doctors can also identify patients who are prone to experiencing adverse reactions to certain treatments. This information can then be used to adjust the treatment for them. Medical research can also be improved through data mining. Researchers can identify trends and patterns in clinical trial data, which they might not have noticed in smaller studies. This data can then be utilized to develop new and better treatments.

Data mining can also help improve the efficiency of the healthcare system by identifying inefficiencies. For instance, by reducing the number of unnecessary tests and procedures, healthcare providers can save money. Data mining has the capacity to improve the healthcare system's quality by identifying inefficiencies and providing more accurate information, as well as improving medical research and providing insight into patient behavior. As the field continues to expand, it can expect to see more advancements that will benefit both the providers and patients.

- Through data mining, healthcare providers can improve the quality of their patient care by identifying and analyzing hidden relationships and patterns in their data.
- Data mining techniques can help healthcare providers identify potential health threats and improve the quality of patient care. They can then use this information to develop effective treatment plans.
- By mining data, healthcare providers can identify individuals who are at high risk of developing chronic illnesses, such as cancer and diabetes. They can then create customized treatment plans that cater to each patient's specific needs.
- Healthcare providers can use data mining techniques to identify patterns and trends in their patient data to improve the quality of their care and reduce costs.
- In addition, data mining can help healthcare providers identify errors in the medication they use and possible drug reactions, which can lead to better patient outcomes.
- In addition, data mining techniques can help healthcare organizations identify opportunities to improve their operations.
- In population health management, data mining techniques can be utilized to identify high-risk individuals and develop effective interventions for better health outcomes.

Despite the advantages of data mining, implementing it in the healthcare industry can be challenging due to various factors. Some of these include the security concerns and privacy issues that can arise when collecting and sharing patient data.

various issues that need to be resolved in successful implementation of datamining in health care

Despite the potential of data mining techniques to improve the outcomes of healthcare patients, there are still many issues that need resolved before it can be implemented successfully. The quality of the data collected by healthcare organizations is one of the most critical issues that they need to address in order to improve the efficiency of their operations. This is because the information collected is often incomplete and fragmented, which makes it hard to use for analysis. In addition to this, the security of patient data is also a major concern. To ensure that this information is protected, healthcare organizations have to implement effective measures. One of the biggest challenges that healthcare professionals face when it comes to implementing data mining techniques is the lack of skilled individuals who can handle the necessary skills needed to analyze the massive amount of data that they collect. In addition to these, the regulatory and legal environment surrounding the use of data mining in healthcare has also become an issue. Due to the varying laws in different countries, it is important that the organizations follow proper legislation in order to effectively utilize the techniques[19], [24], [25].

Data mining techniques can be very time consuming and require a huge amount of resources, such as software, hardware, and skilled personnel. This is why healthcare organizations usually face significant logistical and financial issues when it comes to implementing such methods. In order to successfully implement data mining

techniques in the healthcare industry, the providers must first address the various issues that can prevent them from achieving their goals. Doing so will allow them to utilize the data collected by the systems to improve the quality of care and deliver more effective services.

- **Data quality:** One of the most important factors that healthcare organizations must consider when it comes to implementing information mining techniques is the quality of the data they collect. A poor quality of data can lead to unreliable and inaccurate results.
- **Data integration:** One of the most important issues that healthcare organizations must address when it comes to data mining is the integration of the various formats and systems used to store the collected information. This process can be very time consuming and complex.
- **Data privacy and security:** The security and privacy of patient data are two of the most important issues that healthcare organizations must address in order to prevent unauthorized access and use. When using data mining techniques, it is important that the information is de-identified to protect the privacy of the patients.
- **Lack of standardization:** One of the biggest issues that healthcare organizations face when it comes to implementing data mining techniques is the lack of standardization. This is because the data collected by the systems is often not standardized. This can prevent them from properly comparing and analyzing the results of the analysis.
- **Lack of data governance:** One of the most important factors that healthcare organizations must consider when it comes to implementing information mining techniques is the proper governance of their data. This is because improper use of this data could have detrimental consequences for patients.
- **Resistance to change:** Despite the potential of data mining techniques to improve the quality of care and deliver more effective services, healthcare professionals are still reluctant to adopt new technologies. This is why it is important that they are provided with the necessary training and support in order to make them comfortable with the new techniques.
- **Costs :** The implementation of data mining techniques can be expensive, and it may be a barrier to their adoption in the healthcare industry. A cost-benefit ratio must be considered when weighing the pros and cons of the project.
- **Interpretation of results:** One of the most important factors that healthcare organizations must consider when it comes to implementing data mining techniques is the interpretation of the results. This is because it can be very challenging to interpret the data collected by the methods. Having the right tools and expertise can help them make informed decisions.

Conclusion and future scope

Data mining has been shown to provide actionable insights from various sources of information related to healthcare, such as electronic health records and databases. It can help improve the quality of patient care and develop effective interventions for public health. The use of data mining techniques has gained widespread interest in the field of healthcare due to their ability to extract valuable information from vast amounts of data. This review aims to provide an overview of the current research on this subject. Through data mining, researchers can identify patterns in the data collected by healthcare organizations, which can be useful in developing predictive models for identifying individuals at risk of developing certain conditions. In addition, it can help improve the diagnosis and treatment of diseases. Despite the immense potential of data mining, there are still many challenges that need to be resolved in order to realize its full impact. These include the quality of the data collected, the interpretability of results, and privacy concerns. To maximize the potential of data mining in healthcare, research must develop methods and algorithms that can handle the immense amount of data collected by various organizations. In addition, policymakers, data scientists, and healthcare professionals must work together to ensure that the use of such techniques is ethical and safe.

References

- [1] S. Bhattacharya, "Intelligent Frequent Pattern Analysis in Web Mining," vol. 2, no. 3, 2013.
- [2] M. Gera and S. Goel, "Data Mining - Techniques, Methods and Algorithms: A Review on Tools and their

- Validity,” *Int. J. Comput. Appl.*, vol. 113, no. 18, pp. 22–29, 2015, doi: 10.5120/19926-2042.
- [3] W. C. Lin, S. W. Ke, and C. F. Tsai, “Top 10 data mining techniques in business applications: a brief survey,” *Kybernetes*, vol. 46, no. 7, pp. 1158–1170, 2017, doi: 10.1108/K-10-2016-0302.
- [4] D. Tomar and S. Agarwal, “A survey on data mining approaches for healthcare,” *Int. J. Bio-Science Bio-Technology*, vol. 5, no. 5, pp. 241–266, 2013, doi: 10.14257/ijbsbt.2013.5.5.25.
- [5] S. Patel and H. Patel, “Survey of Data Mining Techniques used in Healthcare Domain,” *Int. J. Inf. Sci. Tech.*, vol. 6, no. 1/2, pp. 53–60, 2016, doi: 10.5121/ijist.2016.6206.
- [6] S. H. Liao, P. H. Chu, and P. Y. Hsiao, “Data mining techniques and applications - A decade review from 2000 to 2011,” *Expert Syst. Appl.*, vol. 39, no. 12, pp. 11303–11311, 2012, doi: 10.1016/j.eswa.2012.02.063.
- [7] H. Kaur, R. Chauhan, and Z. Ahmed, “Role of data mining in establishing strategic policies for the efficient management of healthcare system – a case study from Washington DC area using retrospective discharge data,” *BMC Health Serv. Res.*, vol. 12, no. S1, pp. 2–3, 2012, doi: 10.1186/1472-6963-12-s1-p12.
- [8] T. R. Baitharu and S. K. Pani, “Analysis of Data Mining Techniques for Healthcare Decision Support System Using Liver Disorder Dataset,” *Procedia Comput. Sci.*, vol. 85, no. Cms, pp. 862–870, 2016, doi: 10.1016/j.procs.2016.05.276.
- [9] X. Qiao and H. Jiao, “Data mining techniques in analyzing process data: A didactic,” *Front. Psychol.*, vol. 9, no. NOV, pp. 1–11, 2018, doi: 10.3389/fpsyg.2018.02231.
- [10] M. M. Malik, S. Abdallah, and M. Ala’raj, “Data mining and predictive analytics applications for the delivery of healthcare services: a systematic literature review,” *Ann. Oper. Res.*, vol. 270, no. 1–2, pp. 287–312, 2018, doi: 10.1007/s10479-016-2393-z.
- [11] C. P. Shen *et al.*, “A data-mining framework for transnational healthcare system,” *J. Med. Syst.*, vol. 36, no. 4, pp. 2565–2575, 2012, doi: 10.1007/s10916-011-9729-7.
- [12] M. Ahmed and A. S. S. M. Barkat Ullah, “Infrequent pattern mining in smart healthcare environment using data summarization,” *J. Supercomput.*, vol. 74, no. 10, pp. 5041–5059, 2018, doi: 10.1007/s11227-018-2376-8.
- [13] K. Abouelmehdi, A. Beni-Hessane, and H. Khaloufi, “Big healthcare data: preserving security and privacy,” *J. Big Data*, vol. 5, no. 1, pp. 1–18, 2018, doi: 10.1186/s40537-017-0110-7.
- [14] M. Shouman, T. Turner, and R. Stocker, “Using data mining techniques in heart disease diagnosis and treatment,” *Proc. 2012 Japan-Egypt Conf. Electron. Commun. Comput. JEC-ECC 2012*, pp. 173–177, 2012, doi: 10.1109/JEC-ECC.2012.6186978.
- [15] R. Chadha and S. Mayank, “Prediction of heart disease using data mining techniques,” *CSI Trans. ICT*, vol. 4, no. 2–4, pp. 193–198, 2016, doi: 10.1007/s40012-016-0121-0.
- [16] K. K. M. N. M, and S. R, “Applications of Data Mining Techniques in Healthcare and Prediction of Heart Attacks,” *Int. J. Data Min. Tech. Appl.*, vol. 7, no. 1, pp. 172–176, 2018, doi: 10.20894/ijdmata.102.007.001.027.
- [17] P. Ertl, “Cheminformatics analysis of organic substituents: Identification of the most common substituents, calculation of substituent properties, and automatic identification of drug-like bioisosteric groups,” *J. Chem. Inf. Comput. Sci.*, vol. 43, no. 2, pp. 374–380, 2003, doi: 10.1021/ci0255782.
- [18] V. Rawte and G. Anuradha, “Fraud detection in health insurance using data mining techniques,” *Proc. - 2015 Int. Conf. Commun. Inf. Comput. Technol. ICCICT 2015*, 2015, doi: 10.1109/ICCICT.2015.7045689.
- [19] B. C. Delaney *et al.*, “Translational Medicine and Patient Safety in Europe: TRANSFoRm - Architecture for the Learning Health System in Europe,” *Biomed Res. Int.*, vol. 2015, 2015, doi: 10.1155/2015/961526.
- [20] I. Yoo *et al.*, “Data mining in healthcare and biomedicine: A survey of the literature,” *J. Med. Syst.*, vol. 36, no. 4, pp. 2431–2448, 2012, doi: 10.1007/s10916-011-9710-5.
- [21] M. R. Mia, S. Akhter Hossain, A. C. Chhoton, and N. Ranjan Chakraborty, “A Comprehensive Study of Data Mining Techniques in Health-care, Medical, and Bioinformatics,” *Int. Conf. Comput. Commun. Chem. Mater. Electron. Eng. IC4ME2 2018*, no. February, 2018, doi: 10.1109/IC4ME2.2018.8465626.
- [22] P. Ahmad, S. Qamar, and S. Qasim Afser Rizvi, “Techniques of Data Mining In Healthcare: A Review,” *Int. J. Comput. Appl.*, vol. 120, no. 15, pp. 38–50, 2015, doi: 10.5120/21307-4126.
- [23] A. Ahmed and S. A. Hannan, “Data Mining Techniques to Find Out Heart Diseases : An Overview,” (*sem qualis*) *Int. J. Innov. Technol. Explor. Eng.*, vol. 1, no. 4, pp. 18–23, 2012.
- [24] Y. A. A. S. Aldeen, M. Salleh, and M. A. Razzaque, “A comprehensive review on privacy preserving data mining,” *Springerplus*, vol. 4, no. 1, pp. 1–36, 2015, doi: 10.1186/s40064-015-1481-x.
- [25] S. Gambhir, S. K. Malik, and Y. Kumar, “Role of Soft Computing Approaches in HealthCare Domain: A Mini Review,” *J. Med. Syst.*, vol. 40, no. 12, 2016, doi: 10.1007/s10916-016-0651-x.