ACHSM Asia-Pacific Journal of Health Management

REVIEW ARTICLE

NOVEL INSIGHTS INTO DATA MINING TO IMPROVE THE SPECIFICITY OF PHARMACOVIGILANCE AND PREVENT ADVERSE DRUG REACTIONS IN PSYCHIATRIC PATIENTS

Aarushi Jain[1], Arunava Ghosh[2]

1. Department of Information Systems, Indian Institute of Management Indore, India
2. School of Business, University of Petroleum and Energy Studies, India

Correspondence: Arunava Ghosh f15arunavg@iimidr.ac.in

ABSTRACT

The aim of this perspective is to provide a review upon the fundamental computational methods deployed in data mining as applied to healthcare data, with particular regards to patient records of psychiatric patients. Albeit clinical data mining has advanced over the years, further research is needed to improve the specificity of pharmacovigilance and prevent adverse drug reaction in psychiatric patients. From describing the main principles and present challenges of data mining to its most-novel applications in clinical psychiatry, this literature review highlights current research gaps that have to be filled to increase the efficacy of psychiatric drugs nowadays, thus improving patient outcome and decreasing hospitalization costs.

KEYWORDS

Data mining; healthcare; pharmacovigilance

INTRODUCTION

DATA MINING FOR FACILITATING KNOWLEDGE DISCOVERY IN HEALTHCARE

Data mining denotes a variety of computer-based information system (CBIS) techniques aimed at discovering novel knowledge (e.g., useful data patterns/features) as derived from data in several fields that involve big data, e.g., business and finance, scientific, clinical and industrial research, as well as education. [4, 16, 19] In case of healthcare, data mining has been extensively applied to large clinical datasets, thus aiding medical diagnosis, enabling to tailor treatments to individual patients and improving the so-called "Health Care Output" (HCO) worldwide, resulting in an improved quality of care, patient outcome and a considerable reduction in hospitalization costs. [4, 13, 19, 33] Data mining helps in planning healthcare activities and reducing the number of inpatients in the hospital. This improves the convenience in healthcare systems. A similar study has suggested multiple strategies for healthcare systems' service convenience flexibility. [39] We have focussed on clinical data as clinical work deals with direct patient care. Non-clinical work may support patient care, but the work does not provide direct diagnosis, treatment, or care for the patient. Hence, the non-clinical data acquired isn't from first-hand experience.

Data mining highly relies upon methods of data patterns extrapolation (i.e., feature extraction), association, clustering and classification. [19, 33] The results drawn from deploying data mining techniques are useful to assist healthcare professionals in their decision-making process, thus helping design patient-specific treatments and drugs, minimising adverse reactions to medicines and improving the life of patients substantially, also reducing healthcare-related costs. [13, 33]

Nevertheless, the knowledge of the usage of data mining that has so far been inferred from medical data is still negligible. [16, 37] Research aimed at designing customized algorithms in data mining will help improve the whole healthcare system, making treatments more patient-specific and efficient, reducing costs considerably. [13] The prediction of trends in the data can be achieved not only by deploying data mining techniques that make use of meaningful patterns derived from the data themselves but also by discarding unnecessary pieces of information within the data that may severely impair the predictive power of data mining-based computational models. Therefore, questions upon how to use the pre-existing data meaningfully without adding to the current data further data have not yet been answered. Data Mining itself is a technique to identify patterns in a pre-built database. A suitable understanding must be developed between the data mining techniques and the previous data available at our deployment without worrying about the new data being accumulated. [36]

Data mining has three high-level objectives in health management: gaining an understanding of medical data, assisting healthcare professionals in their decision-making processes and analysing the pre-existing data to draw additional, new knowledge from them to improve our understanding upon the pathophysiology of different diseases as well as patient outcome. [16]

But solely in the United States, deploying data mining-based techniques wisely can save up to $450 billion to the healthcare system each year due to the large datasets collected in hospitals on a daily basis. [13]

DEFINITION AND APPLICATIONS OF CLINICAL DATA MINING

Healthcare involves medical procedures such as diagnoses, treatments, assessment of prognoses, methodologies aimed at preventing pathologies, physical injuries and mental disorders from arising in humans. [4, 19] Healthcare services and industries generate a vast amount of data every day, involving electronic medical records, as well as other benchmarking reports and findings. [19] Nonetheless, such healthcare-related data have not been efficiently deployed. [4, 19]

Enabling to retrieve useful knowledge from pre-existing data, without requiring the collection of further medical data, data mining-based techniques can be used to diagnose several pathologies and aid physicians in their decision-making processes regarding patient treatments and assessment of prognoses, thus improving patient outcome, and reducing the length and costs of hospitalization. [4, 16, 19] Therefore, clinical data mining (CDM) is the application of Artificial Intelligence (AI)- and data mining-based methodologies deploying clinical data to improve the quality of healthcare. [16, 33] Software-based applications aimed at storing patient data electronically have considerably facilitated an extensive use of data mining techniques and helped retrieve useful patterns from current data to diagnose and cure several pathologies. [16]

PHASES OF CLINICAL DATA MINING

The fundamental phases involved in CDM include data collection, pre-processing (e.g., sampling), analysis, maximization (e.g., feature extraction/selection), modelling, classification, clustering, outlier detection, prediction, ranking and holistic evaluation. [16]
All the above-mentioned steps are essential to retrieve meaningful and novel pathophysiological patterns from patient data (e.g., electronic records and data). [16] The major phases of clinical data mining will be examined in the following section.

LEARNING AND VALIDATION

The CDM modelling framework involves an initial learning phase in which the computational algorithms replicate the observed/learnt phenomenon as derived from the clinical data available, followed by a testing phase to validate the accuracy and reliability (e.g., robustness) of the computational model designed. [16] The most widely used performance assessment tools as applied to data mining techniques are the following: accuracy, sensitivity, specificity and receiver operator characteristic (ROC) curves. [16]

The learning stage can be achieved either via a supervised or an unsupervised methodology, respectively depending upon whether the class labels of the training data are preliminarily known or unknown. [16]

FUNDAMENTAL METHODS DEPLOYED IN CLINICAL DATA MINING

Models required for designing data mining techniques are either predictive, e.g., classification, regression, generalization, categorization, or descriptive, e.g., characterization, anomaly detection, clustering, association, pattern matching, data visualization, meta-rule-based methods and correlation analysis. [4, 16, 19]
Whilst predictive models tangentially deploy supervised learning-based methodologies to predict the future behaviour of specific variables [19], descriptive models make use of unsupervised learning algorithms to retrieve meaning patterns to describe the inputted data in order for them to be easily interpreted by human operators. [19] Therefore, due to their practical and clinically viable nature, predictive models are the most commonly utilized data mining-based techniques in the field of healthcare. [4, 19]

The techniques used for performing anomaly detection are standard support and density-induced vector data description, as well as the Gaussian mixture. [4, 19] Whilst the vector quantization technique is widely used for clustering, the following methods are deployed for classification [4, 19]: statistical, discriminant analysis, decision tree, Markov based, swarm intelligence, K-nearest neighbour, genetic classifiers, artificial neural network, support vector and association rule. Below are the data mining techniques available for healthcare management.

LOGISTIC REGRESSION

Logistic regression (LR) is a technique of data mining that deploys either continuous, discrete or hybrid types of datasets and the corresponding binary target, calculating a linear sequence of inputs and conveying it to a mathematical function named "logistic". [4, 19] Results attained in previous research works are not so promising owing to the considerably reduced size of the input datasets. [4, 19] Therefore, it is recommended to use a dataset of larger size to improve the accuracy of the LR-based learning type of data mining algorithm. [4, 19]

FEATURE RELEVANCE ANALYSIS

Feature relevance analysis [16] is a stage of data processing in data mining that allows scientists to discard some predictors of pathologies from a pre-existing dataset to facilitate data exploration, as their relative contribution to discerning the required classes in the data would not be significant [16], thus decreasing computational time and complexity substantially. [16]
Feature selection facilitates the visualization and understanding of clinical data, as well as their measurement and storage, also considerably reducing the time in training and testing data mining-based algorithms, improving their out-of-sample classification accuracies. [16]

ALTERNATIVE METHODS TO FEATURE SELECTION: GINI IMPORTANCE AND SWARM INTELLIGENCE

An alternative method to univariate statistics-based feature selection, named "Gini Importance" aimed at capturing population discrepancies in functional connectivity was designed to detect the most robust and highly predictive functional connections by summarizing multivariate patterns of interaction. [16] Differently from the univariate features that resulted in a considerably high variance across subsets of the data partitioned having a low classification/predictive power, the Gini Importance was able to accurately and reliably assess the extent of changes in functional connectivity as induced by Schizophrenia, thus enabling an early diagnosis of the condition. [16]

Alternatively, the Swarm Intelligence-based method was used to perform the same diagnostic task. [4, 19] Via particle swarm optimization (PSO), the computational algorithm enables to discern pathological data across large search spaces. [4, 19] The classification/predictive procedure resulted to be faster and more accurate if the number of features used were reduced. [4, 19]

CLUSTERING AND CLASSIFICATION

The clustering technique deployed in CDM is a widely applied descriptive method that blends statistics and numerical analysis whereby a set of groups or clusters able to describe the input data is found. [19, 33] The clusters so identified can be used to analyse and find the drug which has high probability of risk. [33]

The main algorithms deployed in the vector quantization

method are K-means, K-mediods and X-means, which can be compared via the Davies-Bouldin Index. [4, 19]

The clustering algorithms aim at grouping elements (e.g., medical records of patients) whilst maximizing a similarity metric, e.g., proximity, between elements of the same class or cluster. [16]

A generalization of the clustering method comes under the umbrella term of "classification", which makes use of different mathematical functions to assign patient data with certain patterns/features to a predetermined class, e.g., healthy or pathological. [16] The Bayesian classifiers, artificial neural networks (ANNs) and the Support Vector Machines (SVMs) Learning are amongst the most commonly utilized Artificial Intelligence-based algorithms for classification. [4, 16, 19]

ADVANTAGES AND DISADVANTAGES OF DEPLOYING DATA MINING-BASED TECHNIQUES

The main advantage brought about by CDM is the efficient utilization of pre-existing clinical data regarding patient demographics along with their medical conditions to obtain new clinically viable knowledge, retrieving useful patterns/features and their relative relationships from them without requiring further data to be collected. [16]
Nevertheless, some CDM-based tools may not handle missing data satisfactorily for some medical conditions, such as mental disorders. [16]

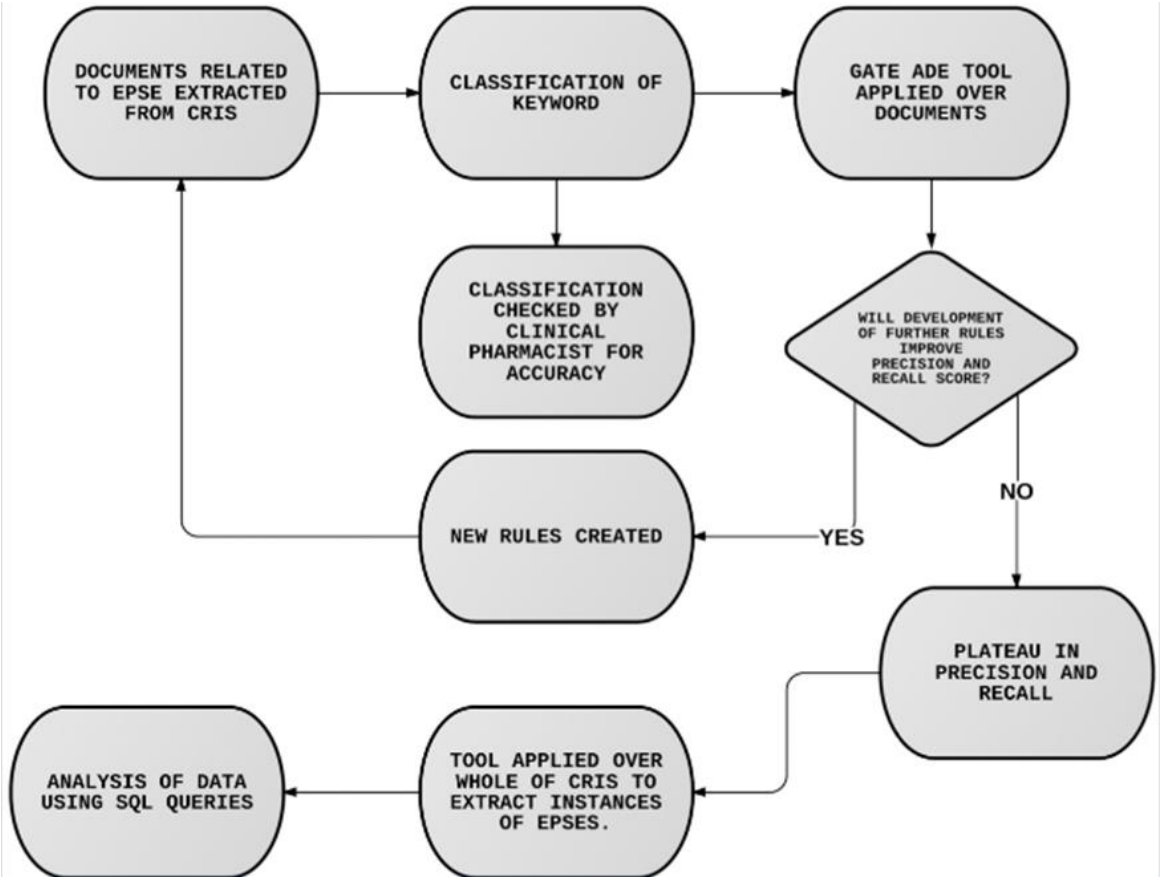
DATA MINING AS CLINICAL DECISION SUPPORT SYSTEM TO DETECT ADVERSE DRUG REACTIONS IN PSYCHIATRIC PATIENTS

In the field of psychiatry, data mining techniques can be useful in pharmacovigilance to detect adverse drug reactions (ADRs), thus assisting physicians in their decision-making processes aimed at identifying the most suitable drug for a psychiatric patient, decreasing the potential adverse drug reactions and, therefore, reducing the length of hospitalization and enabling recovery. [16]
ADRs, which is any harm occurring when a certain drug is used, may occur following a single dosage or prolonged administration of a drug or result from the combination of two or more drugs. [3, 10, 25, 33] Adverse Drug Reactions account for more than 5% of all hospitalizations and are one of the leading aetiologies of injury or death amongst psychiatric patients undergoing pharmaceutical treatments. [7, 9, 12, 29]

CDM can be deployed to mine readily available observational data such as electronic medical records (EMRs) [27, 29], where co-morbidities and concomitant drug use are present, to provide quantifiable patient-specific metrics of harm perceived by psychiatric patients. [9] Noteworthily, the information that can be retrieved from electronic healthcare records (EHRs) can be helpful for future research use, such as via the GATE Natural Language Processing (NLP) software developed. [35] Fig. 1 outlines a scheme upon how the software works. CDM improves patient safety and quality of care for multiple medical conditions by integrating individual patient data and evidence databases to facilitate clinical decision making. [38]

This procedure would be analysed concurrently with spontaneous reporting systems to provide a more reliable assessment of the patient condition and, hence, facilitate early diagnosis and treatment of mental disorders.
The knowledge derived from deploying CDM in this instance is essential for providing further informed medical care and, therefore, for preventing ADRs, thus helping psychiatric patients rehabilitate more and hopefully facilitating their reintegration into society. [7, 12, 20, 22, 28]

FIGURE 1. A SCHEME UPON HOW THE SOFTWARE "GATE NATURAL LANGUAGE PROCESSING (NLP)" WORKS. [35]



HIGHLIGHTS UPON RESEARCH GAPS IN DATA MINING AS APPLIED TO PSYCHIATRY

Clinical data outlining phenotypes and patient treatment are currently under-utilized; these resources could be used concurrently with data mining techniques to infer new medical knowledge. Data mining of Electronic Patient Records (EPR) can help unveil novel discrepancies amongst several psychiatric disorders and, therefore, assist physicians in tailoring treatments to individual patients, hence improving patient outcome and reducing hospitalization time and costs. An EPR is the systematized collection of patients' and population's electronically stored health information in a digital format. These records can be shared across different health care settings. Integrating the information from EPR with genetics would lead to unveil new pathophysiological aspects of mental illnesses.

The human body is a complex biological entity, which is composed of various levels (from genetic to cellular, molecular, tissue, organ and system level). [13] Hence, scientific research in tailoring data mining techniques to psychiatric patient data must take account of these different levels and be scalable to represent a correct pathophysiological condition and predict patient outcome further to adopting a specific treatment designed on a patient-specific basis. [13] Furthermore, such algorithms need to consider discrepancies perceivable amongst different populations of patients being considered. [13]

CONCLUSION

Clinical data mining is a research-based tool whereby physicians can retrieve and interpret pre-existing clinical data from patient records and infer new knowledge that can aid them in various decision-making processes regarding diagnosis and several treatment strategies. [16] Considering the high volume of medical records, data mining of readily available clinical data is more important and much more preferred than adding further data to the pre-existing ones. [16]

With particular regards to psychiatry, EPR data can be used to prevent ADRs via different techniques such as the GATE software [35], and, therefore, improve patient outcome, thus reducing the time and costs of hospitalization. [9, 27, 29, 33] This work is nifty in case of cancer patients who are the frequent victims of adverse drug reactions. Through our study, we believe that we could throw some light on the existing literature on the usage of data mining in case of pharmacovigilance. Future study can focus on the feature relevance analysis in order to provide more insights into the clinical dataset. A similar study can also be conducted by studying non-clinical datasets.

References

1. Agrawal R, Mannila H, Srikant R, Toivonen H, Verkamo AI. Fast discovery of association rules. Advances in knowledge discovery and data mining. 1996 Feb 1;12(1):307-28.
2. Alzheimer FAQ. Frequently Asked Questions. Available: < alzheimer.wustl.edu/About_Us/FAQ.htm> (accessed 17/7/2021)
3. Rohilla A, Yadav S. Adverse drug reactions: An overview. Int J Pharmacol Res. 2013;3(1):10-2.
4. Bellazzi R, Ferrazzi F, Sacchi L. Predictive data mining in clinical medicine: a focus on selected methods and applications. Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery. 2011 Sep;1(5):416-30.
5. Bender S, Grohmann R, Engel RR, Degner D, Dittmann-Balcar A, Rüther E. Severe adverse drug reactions in psychiatric inpatients treated with neuroleptics. Pharmacopsychiatry. 2004 Mar;37(S 1):46-53.
6. Classen DC, Resar R, Griffin F, Federico F, Frankel T, Kimmel N, Whittington JC, Frankel A, Seger A, James BC. 'Global trigger tool'shows that adverse events in hospitals may be ten times greater than previously measured. Health affairs. 2011 Apr 1;30(4):581-9.
7. Coloma PM, Trifirò G, Schuemie MJ, Gini R, Herings R, Hippisley-Cox J, Mazzaglia G, Picelli G, Corrao G, Pedersen L, van der Lei J. Electronic healthcare databases for active drug safety surveillance: is there enough leverage?. Pharmacoepidemiology and drug safety. 2012 Jun;21(6):611-21.
8. Auslander GK. Clinical Data Mining: Integrating Practice and Research by Irwin Epstein: Oxford: Oxford University Press, 2010, 240 pages, $26.95 (paperback).
9. Eriksson R, Werge T, Jensen LJ, Brunak S. Dose-specific adverse drug reaction identification in electronic patient records: temporal data mining in an inpatient psychiatric population. Drug safety. 2014 Apr;37(4):237-47.
10. FDA - Medication Errors. Medication Errors Related to CDER-Regulated Drug Products <www.fda.gov/drugs/drugsafety/medicationerrors/> (accessed 17/7/2021)
11. Haerian K, Varn D, Vaidya S, Ena L, Chase HS, Friedman C. Detection of pharmacovigilance-related adverse events using electronic health records and automated methods. Clinical Pharmacology & Therapeutics. 2012 Aug;92(2):228-34.
12. Harpaz R, DuMouchel W, Shah NH, Madigan D, Ryan P, Friedman C. Novel data-mining methodologies for adverse drug event discovery and analysis. Clinical Pharmacology & Therapeutics. 2012 Jun;91(6):1010-21.
13. Herland M, Khoshgoftaar TM, Wald R. A review of data mining using big data in health informatics. Journal of Big data. 2014 Dec;1(1):1-35.
14. Iqbal E, Mallah R, Jackson RG, Ball M, Ibrahim ZM, Broadbent M, Dzahini O, Stewart R, Johnston C, Dobson RJ. Identification of adverse drug events from free text electronic patient records and information in a large mental health case register. PloS one. 2015 Aug 14;10(8):e0134208.
15. Iqbal E, Mallah R, Jackson RG, Ball M, Ibrahim ZM, Broadbent M, Dzahini O, Stewart R, Johnston C, Dobson RJ. Identification of adverse drug events from free text electronic patient records and information in a large mental health case register. PloS one. 2015 Aug 14;10(8):e0134208.
16. Jacob SG, Ramani RG. Data mining in clinical data sets: a review. training. 2012 Dec;4(6).
17. Jain T, Bhandari A, Ram V, Parakh M, Wal P, Nagappa AN. Drug interactions and adverse drug reactions in hospitalized psychiatric patients: A critical element in providing safe medication use. German J Psychiatry. 2011 Jan 1;14:26-34.
18. Jensen PB, Jensen LJ, Brunak S. Mining electronic health records: towards better research applications and clinical care. Nature Reviews Genetics. 2012 Jun;13(6):395-405.
19. Jothi N, Husain W. Data mining in healthcare–a review. Procedia computer science. 2015 Jan 1;72:306-13.
20. Landmark CJ, Johannessen SI. Safety aspects of antiepileptic drugs—focus on pharmacovigilance. Pharmacoepidemiology and drug safety. 2012 Jan;21(1):11-20.
21. LePendu P, Iyer SV, Bauer-Mehren A, Harpaz R, Mortensen JM, Podchiyska T, Ferris TA, Shah NH.

Pharmacovigilance using clinical notes. Clinical pharmacology & therapeutics. 2013 Jun;93(6):547-55.
22. McClellan M. Drug safety reform at the FDA—pendulum swing or systematic improvement?. New England Journal of Medicine. 2007 Apr 26;356(17):1700-2.
23. Meystre SM, Savova GK, Kipper-Schuler KC, Hurdle JF. Extracting information from textual documents in the electronic health record: a review of recent research. Yearbook of medical informatics. 2008;17(01):128-44.
24. Muench J, Hamer AM. Adverse effects of antipsychotic medications. American family physician. 2010 Mar 1;81(5):617-22.
25. Nebeker JR, Barach P, Samore MH. Clarifying adverse drug events: a clinician's guide to terminology, documentation, and reporting. Annals of internal medicine. 2004 May 18;140(10):795-801.
26. NHLBI Clinical Trials. https://www.nhlbi.nih.gov/health-topics/clinical-trials (accessed 17/7/2021)
27. Roitmann E, Eriksson R, Brunak S. Patient stratification and identification of adverse event correlations in the space of 1190 drug related adverse events. Frontiers in physiology. 2014 Sep 9;5:332.
28. Roque FS, Jensen PB, Schmock H, Dalgaard M, Andreatta M, Hansen T, Søeby K, Bredkjær S, Juul A, Werge T, Jensen LJ. Using electronic patient records to discover disease correlations and stratify patient cohorts. PLoS computational biology. 2011 Aug 25;7(8):e1002141.
29. Sampathkumar H, Chen XW, Luo B. Mining adverse drug reactions from online healthcare forums using hidden Markov model. BMC medical informatics and decision making. 2014 Dec;14(1):1-8.
30. Strom BL. What is pharmacoepidemiology?. Pharmacoepidemiology. 2019 Nov 6:1-26.
31. Szabo CP. Common adverse drug reactions with psychiatric medications and an approach to their management. CME: Your SA Journal of CPD. 2011 Jun 1;29(6):230-2.
32. VA Center for Medication Safety and VHA Pharmacy Benefits Management Strategic Healthcare Group and The Medical Advisory Panel. Adverse Drug Events, Adverse Drug Reactions And Medication Errors Frequently Asked Questions. < https://www.pbm.va.gov/PBM/vacenterformedicationsafety/tools/AdverseDrugReaction.pdf> (accessed 19/7/2021)
33. Viveka S, Kalaavathi B. Review on clinical data mining with psychiatric adverse drug reaction. In2016 World Conference on Futuristic Trends in Research and Innovation for Social Welfare (Startup Conclave) 2016 (pp. 1-3). IEEE.
34. World Health Organization. Reporting and learning systems for medication errors: the role of pharmacovigilance centres. < https://apps.who.int/iris/handle/10665/137036> (accessed 19/7/2021)
35. Wu H, Ibrahim Z, Iqbal E, Dobson RJ. Predicting Adverse Events from Multiple and Dynamic Medication Episodes–a preliminary result in a large mental health registry. InIJCAI 2016-Workshop on Knowledge Discovery in Healthcare Data 2016 Jul.
36. Wilson AM, Thabane L, Holbrook A. Application of data mining techniques in pharmacovigilance. British journal of clinical pharmacology. 2004 Feb;57(2):127-34.
37. Zaffalon M, Wesnes K, Petrini O. Reliable diagnoses of dementia by the naive credal classifier inferred from incomplete cognitive data. Artificial intelligence in medicine. 2003 Sep 1;29(1-2):61-79.
38. Nilan Y, Sellahewa D, Fernando S, Gamage L, Meedeniya D. A Clinical Decision Support System for Drug Conflict Identification. In2018 Moratuwa Engineering Research Conference (MERCon) 2018 May 30 (pp. 126-131). IEEE.
39. Kumar P, Bera S, Chakraborty S. An examination of the association between service convenience flexibility in healthcare delivery systems and patient satisfaction. South Asian Journal of Management. 2017 Oct 1;24(4).