Nat. Volatiles & Essent. Oils, 2021; 8(5): 1198 - 1203

NVEO
Natural Volatiles &
Essential Oils

Application of Data Mining In Healthcare and Current Issues

[1]Dr. Sanjivani Jadhao, [2]Malhar Jadhao, [3]Gauri Ajay Chaudhary, [4]Priyanka Deshmukh, [5]Dattu Hawale

[1]Prof and Head, Dept of Rachna Sharir, Datta Meghe Ayurvedic Medical College Hospital & RC, Wanadongri, Nagpur.
[2]CS Graduates from Shri Ramdeobaba College of Engineering and Management
[3]Computer Technology Yeshwantrao Chavan College of Engineering, Nagpur

[4]Assistant Professor Dept. of Anesthesiology Datta Meghe Medical College, Nagpur.
[5]Tutor Dept. of Biochemistry Jawaharlal Nehru Medical College, Datta Meghe Institute of Medical Sciences, Sawangi (Meghe), Wardha

Abstract
Quantitative and Qualitative exploration of clinical and diagnostic data using advanced analytics could unravel hidden medical knowledge by finding correlations, causations, and associations between apparently independent variables. Therefore, the scope and use and of Data Mining techniques in the current healthcare system is increasing steadily. In regard to this, we will discuss the disciplines, methods, models, algorithms, and results, and how these techniques would help in performing studies including but not limited to long-term prospective and retrospective studies, population studies, correlation studies, multicentric, multiracial, phased studies, meta-analysis, pharmacovigilance, etc. on Ayurvedic drugs and methods. We have discussed the applications of Data Mining on healthcare that are being implemented in developed countries. We have also discussed the issues like lack of quality data and record-keeping, and other issues and challenges in conducting ayurvedic studies in India, and how the National Digital Health Blueprint (NDHB) would be a game-changer in the current healthcare system of India.
Keywords: Data Mining, Healthcare, Ayurveda, Traditional, NDHB

Introduction:

Patient Health and diagnostic data are being gathered and stored in the form of Electronic Health Records (EHR) in the developed countries. In the healthcare domain, Data Mining has remodeled the current processes of assessing treatment effectiveness, diagnosis, identifying fraudulent activities, making health policy, dealing with real world evidence, managing pharma industries and hospitals, and maintaining customer relationship.

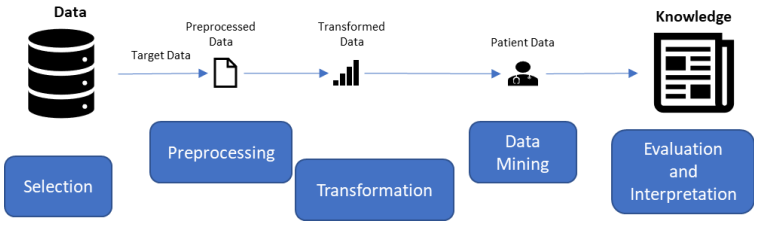
With data mining, useful information is derived by observing trends in huge datasets. These trends and information help the organizations make informed decisions. Different data mining methods have different applications in pattern discovery and extraction [1].

By adopting NDHB and leverage into AI and ML technologies in a country like India, we will see an evolution in Predictive treatment than a reactive treatment for Ayurvedic medicines. An increase in the availability and accessibility of patient data would decrease the cost and increase the efficiency of the healthcare providers, thus improving the quality of life of patients.

Data Mining: An Overview

Deriving information from data by observing patterns and trends in huge datasets is called data mining [3]. Using data mining we can extract valuable information which isn't readily available and visible in the dataset. With Data Size growing from day to day, this knowledge and information is becoming more important to aid the ayurvedic studies.

Initially the target data is selected from the entire dataset and then brought to a state where it could be easily understood by the machine. Later-on the data is transformed into a usable format. Here-after various findings are used to derive insights, with which we get information. This is how we convert data into information.



Data Mining unites the concepts and methods from other disciplines like Statistics, ML, AI, Probability, Database Management etc.

The Interpretation step mainly involves association analysis, Extrapolative modeling, divergence detection etc., to either classify the data, build associations between attributes, cluster the data or find anomalies in the data, which are achieved with the help of various Data mining algorithms.

Data Mining Algorithms in Healthcare

The Healthcare industry is a data-rich industry with massive amounts of data in the form of Electronic Health Records (EHR), Claims Data, Registry Data etc. These data sources are however are not being utilized to their full potential even in the developed countries and rarely utilized in India. Data Mining in Healthcare is mainly used for identifying medical conditions, making data-driven clinical decisions, performing clinical trials, and analyzing real-world evidence (RWE). As discussed previously, to find valuable information's in these large volumes of Data using Data Mining, the following methods and algorithms are used-

- Anomaly Detection: Anomaly detection is used in identifying the data points which deviate from the dataset's usual behavior [5]. This is done by identifying the extreme values on single as well as multiple dimensions. Gaussian Mixture model and various forms of Support Vector Data Description (SVVDs) are generally used for the same and their effectiveness is measured with AUC scores. The uncertain data points are prone to be available in all datasets, and they could be resolved with anomaly detection.
- Clustering: Through clustering, one aims to divide and group the data into finite number of clusters such that data points in each group are most similar [5]. Clustering is generally helpful in anticipating patient readmissions for a treatment. K-means, K-medoids, and X-means are the commonly used methods and its accuracy is measured using Davies-Bouldin Index.

- Discriminant Analysis: Linear discriminant analysis (LDA) is used in discriminant analysis to separate the data points with the linear combination of the features [5]. The algorithm captures linear associations among the independent variables; hence it is helpful in exploring and analyzing linear data with better accuracy.
- Decision Tree: decision tree gives better accuracy prognosis because it's tree like model can separate observation into branches with the conditional control statements. By recursively traversing through multiple conditions, the result state is achieved.
- Swarm Intelligence: Swarm intelligence method gives good accuracy when used for diagnosis models. The classification process is achieved by efficiently identifying optimal solutions in large datasets. Particle Swarm Optimization(PSO) does this job most efficiently.
- K-Nearest Neighbor: It is used for classification of data based on multiple dimensions. The algorithm is very effective since all the dimensions and data points are utilized in the training process, but it is slower than other classification techniques. But since its classification accuracy is better, and it has more significance in medical diagnosis.
- Logistic Regression: Logistic regression is the most basic binary classifier which is still used in healthcare for proving quality results in lesser time. It uses logistic function to separate the data points [6]. It works well even for the larger datasets.
- Bayesian Classifier: It builds a probabilistic model which is lesser resource intensive. It allows us to use more features without increasing bias. It's helpful in the situations where we don't have complete data as it handles the empty values.
- Support Vector: It is useful in comparative study purposes due to excellent performance on out-of-sample data. The performance is improved by the regularization functions and it gives better performance by mapping the features to higher-dimension, and hence it is widely used in medical diagnosis.

Applications of Data Mining in Healthcare.

Healthcare industries generate huge amounts of data about patients, clinical trials, electronic health records, devices, Real-world evidence, hospital resources, etc. Findings and insights from these Data enable support for decision making and cost optimization. Data mining applications in healthcare are as follows-

- Measuring Treatment effectiveness: Data mining methods help to assess the potency and effects drugs and medical procedures, and analyses which route, combination, dose etc. is most effective by comparing symptoms, quality of life and other health related metrics for patients with diverse treatment histories and demographics.
- Healthcare management: By analyzing huge amounts of public health data, epidemic breakouts, population at high-risk, existing chronic disease etc., could be identified so that a plan could be formulated to take appropriate steps to reduce the impact on the healthcare system. Early detection and management of pandemics and other unusual or novel diseases is possible with Data Mining.
- Public health Policymaking: By combining GIS (Geographic Information System), public health data and data mining technologies, the local health departments and authorities

could detect location-specific unusual health-related events like child mortality, toxicity etc., find the causes, and contain and reduce them before they worsen. Based on these results, policies on subjects like sanitation, guidelines on permissible limits etc. could be updated.

- Evidence-based medications: By identifying and marking the previously undiscovered health hazards and other causes of medical complications due to drug combinations and procedures followed, through the past treatment data, the complications and deaths due to hospital and human error could be drastically reduced.
- Non-Invasive diagnosis: Several predictive systems based on clustering methods have proven to show better results than the invasive, costly, and painful diagnosis systems. By analyzing various health attributes, it could be determined if a patient actually needs a biopsy or not. Data-driven results have shown higher accuracy than the conventional diagnosis systems.
- Real World Evidence: The drugs approved as safe after clearing the 3rd phase of clinical trials are regarded as harmless but are later found to have various side and adverse effects. The pharma industries study and discover the side effects and adverse effects of the drugs using data mining. Using Data Mining, now they are able to find these results years before they could do it with the conventional methods
- Customer relationship management: Interactions between patients and commercial organizations like banks, insurance, retailers, etc. could be initiated and managed by mapping the patients' expenditure patterns with their financial needs and suggesting patients a suitable health cover accordingly.
- Fraud and abuse Detection: By identifying anomalies and unusual patterns in insurance claims, prescriptions, or referrals, the patients or the regulatory authorities could be notified about the fraudulent insurance claims, improper prescriptions, or referrals, thus making it difficult to conduct fraudulent activities in the healthcare systems.
- Medical Device Industry: With technological advancements and a better understanding of patient diagnosis needs, the medical devices have drastically improved. With limited resources, they can monitor and track the vital signs of the patients in the real time by using Data Stream mining algorithms [8].
- Pharmaceutical Industry: Pharmaceutical firms have very recently started managing their registers and operations using Data Mining methods. By understanding their competitive position, prescription and expenditure patterns in the market etc. they optimize their operations and the marketing and distribution strategy.
- Hospital Management: Temporal and spatial behavior of Hospital activities are globally visualized with the huge amount of hospital data collected. It is further used to optimize the services for Hospital Management, Medical Staff, and Patients.
- Bioinformatics: Systems biology could be understood better with health data and hence its need and demand are increasing in the international sciences. Applying data mining on the Biological data will give us new insights.

Issues and Challenges for Data Mining in Healthcare

Although Ayurveda literature contains systematic documentation of clinical experiences, the lack of structured data poses a challenge in proving the safety and efficacy of the ayurvedic drugs. The same data could be used to obtain real-world evidence and thus improving the credibility of the Ayurveda discipline.

Another issue that comes into the picture because of lack of Standardization and Quality Control is that the findings from data mining methods could be incorrect and misleading if the standardization and quality control aren't looked after.,

The Data Mining methods are good at describing the patterns and trends but not explaining the trends. This inability to be totally unambiguous reduces the credibility of Data Mining in this segment of healthcare. Moreover, overcoming biases due to majorities and finding minorities that do not conform to the patterns is another challenge. Furthermore, Ayurveda considers disease occurring due to multifactor and follows multitarget management [10]. Thus, Health Data with limited attributes might not be enough to perform Ayurvedic Studies.

Although the results from data mining methods are credible, the healthcare practitioners aren't welcoming enough to accept these results. The physicians choose to listen to any respected Key Opinion Leaders(KOL) over the Data Mining results. Thus, convincing the physicians to trust and follow those results would be difficult.

Since individual Health record is sensitive information and multiple parties would be involved in analyzing the data at various steps, ensuring privacy and ethical use of those records would be another big challenge.

Conclusion and Recommendations

This paper aimed to discuss a brief introduction on Data Mining, Methods used in Healthcare, currently used and Potential application of Data Mining in the healthcare system, and Issues and challenges in implementing data mining on ayurvedic studies. Health care organizations and agencies in countries like India should investigate into these applications and formulate a detailed plan on how to implement these in their country with the database of their patients. Healthcare organizations in collaboration with the Government of India should start NDHB soon as possible and start leveraging into sophisticated information systems and inculcate these results into the policymaking to eliminate issues like infant mortality and disease outbreak (Malaria, Dengue, etc.). The same systems could also be used for vaccination planning, identifying high-risk individuals, etc. This would also help in establishing the standardization, safety, and efficacy benchmarks of the Ayurvedic drugs, rather than just relying on the Empirical shreds of evidence and literature, thus improving the credibility of the Ayurveda Discipline. We also need strict laws and clear policies on the misuse of Health records to protect the privacy and security of the patients. Data Mining gives promising results with encouraging accuracies and it would change the landscape of the current Healthcare System.

REFERENCES

1. R. Agrawal and G. Psaila, "Active data mining," Current, pp. 3–8, 1995. Last Name, F. M.
2. S. H. Liao, P. H. Chu, and P. Y. Hsiao, "Data mining techniques and applications - A decade review from 2000 to 2011**,"**

3. G. E. Vlahos, T. W. Ferratt, and G. Knoepfle, “The use of computer-based information systems by German managers to support decision making,” Inf. Manag., vol. 41, no. 6, pp. 763–779, 2004.
4. J.-J. Yang, J. Li, J. Mulder, Y. Wang, S. Chen, H. Wu, Q. Wang, and H. Pan, “Emerging information technologies for enhanced healthcare,” Comput. Ind., vol. 69, pp. 3–11, 2015.
5. U. Fayyad, G. Piatetsky-Shapiro, and P. Smyth, “From data mining to knowledge discovery in databases,” AI Mag., pp.37–54, 1996.
6. P. J. García-Laencina, P. H. Abreu, M. H. Abreu, and N. Afonoso, “Missing data imputation on the 5-year survival prediction of breast cancer patients with unknown discrete values,” Comput. Biol. Med., vol. 59, pp. 125–133, 2015.
7. B. Zheng, S. W. Yoon, and S. S. Lam, “Breast cancer diagnosis based on feature extraction using a hybrid of K-means and support vector machine algorithms,” Expert Syst. Appl., vol. 41, no. 4 PART 1, pp. 1476–1482, 2014.
8. Mobile Data Mining for Intelligent Healthcare Support
9. David Page and Mark Craven, —Biological Applications of MultiRelationalData Mining‖.
10. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5954257/