

# A Study on Application of Data Mining in Ayurinformatics

C. Gunaseelan  
Research Scholar  
Department of Computer  
Science and Applications  
SCSVMV University,  
Kanchipuram, Tamilnadu, India

V. Ramesh, PhD  
Assistant Professor  
Department of Computer  
Science and Applications  
SCSVMV University,  
Kanchipuram, Tamilnadu, India

## ABSTRACT

Ayurveda - the traditional medicine system of ancient India, which nourishes an individual's physical strength and intelligence along with maintaining a wonderful balance with the environment. Deviant to its nature, it has now become an *inactive magnificence*. Neither the research methodology nor the text book has changed over the past 50 years and paving way for the modern medicine system to progress. This work gives a convincing solution for Ayurveda to regain its lost glory with the incorporation of Information Technology, Data mining, Statistics and Bioinformatics, thus forming a new venture named *Ayurinformatics*. To prove this, relevant information's were collected from various literatures of Ayurveda and data mining algorithms. The findings reached by the implementation of data mining algorithms like k-Means algorithm, Canny Edge detection algorithm, Decision support systems, etc., for finding a better research methodology to identify a better treatment method, has made a significant cutback in both time and money than modern medicine system does. This was apparent by the identification of using methods like k-means algorithm to cluster ayurvedic medicines, prediction methods for Bronchial Carcinoma- a type of cancer, Canny Edge Detection Algorithm for Medicinal Plants Disease Identification and AVD with Transitive text mining algorithm for extracting ayurvedic texts from biological databanks. This work strongly enforces that Ayurinformatics gives a new way of life to Ayurveda with its innovative approach to research methodology and better treatment identification methods.

## Keywords

Ayurinformatics, Ayurveda, Bioinformatics, Data mining, Data mining algorithms, Information Technology, Research Methodology, Statistics..

## 1. INTRODUCTION

### 1.1 Preamble

AYURVEDA - A kind of medical practice follows its ancestry from India and its subcontinents. In general, it has a domestic meaning as *Ayus(r) = Life* and *Veda = Knowledge*. Its literature has been derived from Veda's such as Rig-Veda and Atharva-veda. The holistic or celestial knowledge acquired is termed as it is actually gained from gods to gurus to practitioners. Even in modern day's era, this has been considered as the greatest heavenly therapy for around two millennia. Ayurveda does put forth mainly three elements, i.e., the dosha's (Kapha, Pitta and Vata) and affirms that a superior stability of the dosha's produces high quality health and instability produces pitiable diseases.

## 1.2 Doctrines and Terms

Despite the fact that some of its concepts were retrieved from *non-vedic* systems, i.e., Jainism and Buddhism, the vital thoughts of Ayurveda are the results of vedic values. Its importance is sited on the amount of food intake and sufficient sleep. Ayurveda points out seven basic terms, such as blood(rakta), bone(asthi), fat(meda), muscles(māmsa), marrow(majja), plasma(rasa) and semen (shukra), five classical elements(air, earth, ether, fire and water) and twenty guna's in pairs of ten antonyms; cold/hot, dull/sharp, heavy/light, minute/gross, non-slimy/slimy, smooth/coarse, soft/hard, stable/mobile, unctuous/dry and viscous/liquid.

## 1.3 Observation and Identification

The observation of the illness is done through the integration of physical and mental aspects. This holistic approach possesses numerous channels to transport fluids. These channels made open by the treatment of massage using plant oils and foment. In turn, the diseases occur due to insanitary channels. There are eight ways to identify all types of illnesses, namely Aakruti(appearance), Druk(vision), Jihva(tongue), Mala(stool), Mootra(urine), Nadi(pulse), Shabda(speech) and Sparsha(touch) and five senses (hear, sight, smell, taste and touch).

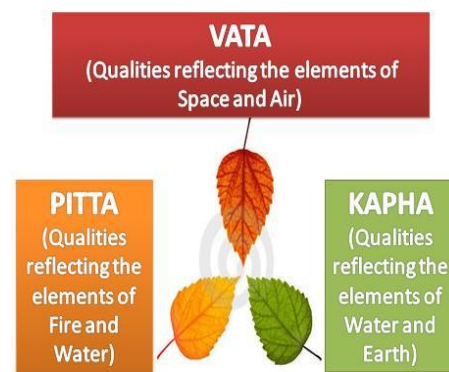


Fig 1: Constituents of Ayurveda

## 1.4 Sources Exercised

There are primarily three kinds of sources exercised in Ayurveda, namely Animal products, Minerals, and Plant bases. Animal products are comprised of bones, fat, gallstone and milk. Minerals are comprised of arsenic, copper sulphate, gold, lead and sulphur. Plant bases are comprised of bark, leaves, fruits, roots and seeds. These sources are essential in making ayurvedic products which has necessary components in it for creating novel ayurvedic medicines.

## 1.5 Current Status

Until now, nearly 80 percent of the inhabitants in India use some form of this traditional medicine. From the eve of 1970's the government of India has made great attempts to standardize Ayurveda by formulating numerous qualifications for Ayurveda practitioners and necessitated accreditation policies to institutions across states. Some of the initiatives are, Indian Medical Central Council Act-framed in 1970, Central Council of Indian Medicine (CCIM) under the Department of Ayurveda, Yoga and Naturopathy, Unani, Siddha and Homoeopathy (AYUSH) - established in 1971,

Ministry of Health and Family Welfare-framed to monitor higher education in Ayurveda and Central Council for Research in Ayurvedic Sciences (CCRAS)-designed to pursue research in Ayurveda. Similar to India, there are many countries practicing Ayurveda as their fundamental medical practice. Approximately 75% - 80% of the people in Nepal, Sri Lanka, China, European and Western countries use some form of ayurvedic products and the governments have established various ministries, medical regulations and universities to offer Ayurveda practice to common people too, so as to enable this medicine system an everlasting practice.

**Table 1. Chief sources pertaining to Ayurveda**

www.ayurveda-herbs.com	www.ayurveda.gov.lk
www.ayurveda-foryou.com	www.ayurveda.org/index.aspx
www.nia.nic.in	www.maharishiayurvedaindia.com
www.ayurveda-in.com	www.ayurveda.md
www.ayurindus.com	www.ayurvedaschool.in
www.blissayurveda.com	www.ayurhelp.com
www.soukya.com	www.healthandayurveda.com
www.ayurvedanextdoor.com	www.ayurveda-seminars.com
www.planetayurveda.com	www.alwaysayurveda.com
www.ayurveda.in	www.ayurved-int.com
www.ayush.com	www.ayushportal.nic.in

## 2. INTRODUCTION TO DATA MINING

### 2.1 Preamble

The quantity of data being piled up in databases have been striving an incredible pace during these days. Consequently the requirement for modern tools and techniques in assisting human for automatically and cleverly investigating huge sets of data to attain constructive information has been in greater demand. This rising requirement has put forth a new era of research named Data Mining or Knowledge Discovery in Databases (KDD), which eventually acquired the notice of numerous researchers in multiple fields together with machine learning, pattern recognition, data visualization and database design. Hence the course of learning intuitive, motivating, novel pattern, in association with expressive, comprehensive and predictive models from massive data is being phrased as *Data mining*. It engages the duties resembling association, anomaly detection, classification, clustering, regression and summarization.

### 2.2 Data Mining Steps

With the context of Data Mining, to achieve valuable information, it follows three major steps: data collection, data shrink and valuable data quest.

There are various approaches adopted by variety of researchers. One such approach involves the following operations:

1. Association analysis: association among data to be identified
2. Extrapolative modeling: usage of neural networks and predictive models
3. Database fragmentation: to partition data in multiple clusters
4. Divergence detection: discovering deviating data

Another approach involves the following operations:

1. Fetching data from a huge database
2. Relevant subset identification
3. Data cleaning, missing field identification and choosing proper sampling design
4. Concerns on dimensionality reduction, projection and proper transformation
5. Model selections for preprocessed data

### 2.3 Data Mining Practices

#### 2.3.1 Decision Tree Model

These are tree like structures for representing collection of decisions to generate various rules for the classification of data set. Some of the examples are Chi Square Automatic Interaction Detection (CHAID) and Classification and Regression Trees (CART) to apply on unclassified data sets to predicate the records with fine outcomes.

#### 2.3.2 Artificial Intelligence

Major techniques used are knowledge acquisition and representation, machine learning, pattern recognition, code based reasoning, intelligent agents and neural networks.

#### 2.3.3 Neural Networks

Various segments of Neural Networks:

1. Construction of networks: Based on the count of attributes, classes and coding method, networks are constructed.
2. Pruning of networks: Removal of superfluous data units to control network error rate.
3. Extraction of rules: Extraction of rules for classification

These make Neural Network a better data mining practice.

### 2.3.4 Statistics

During data selection and sampling, statistics has become a vital constituent. It plays a crucial role in the processes of outlier detection, data cleaning, data smoothing, missing data and noise estimation.

### 2.3.5 Genetic Algorithm

A feasible solution to a problem is created at a random phase with the rules population. Thereafter set of rules are joined to make the next generation offspring's. To modify the genetic structures of each generation, a mutation process is used. More than hundreds of generations can be run using this system. When an optimal solution is found, the process terminates.

### 2.3.6 Visualization

Such techniques provide great platform for mining huge databases. It requires the conjunction of data mining process and human. Some of the techniques used for multidimensional data set visualization are, parallel coordinates, projection matrices, coplots, icon based, graph and dynamic techniques.

## 2.4 Applications

Applications of data mining in real life areas include, Health care, Banking, Retail, and Science etc.

## 3. INTRODUCTION TO AYURINFORMATICS

Ayurveda needs to be reorganized worldwide to meet the rising demands of this cyber mobile society with the application of information and communication technology. Therefore, *Ayurveda requires Ayurinformatics* to remain temperament with this modern world.

The primary solution to further commercialization of sequence data will be to develop informatics technology that transforms this data into information that is useful for diagnosis and therapy. Therefore, modern Ayurveda and Siddha practitioners need training in Ayurinformatics to keep pace with the modern system of medicine.

Ayurvedic data play a vital role while emerging as a business model for the medical and pharmaceutical sector.

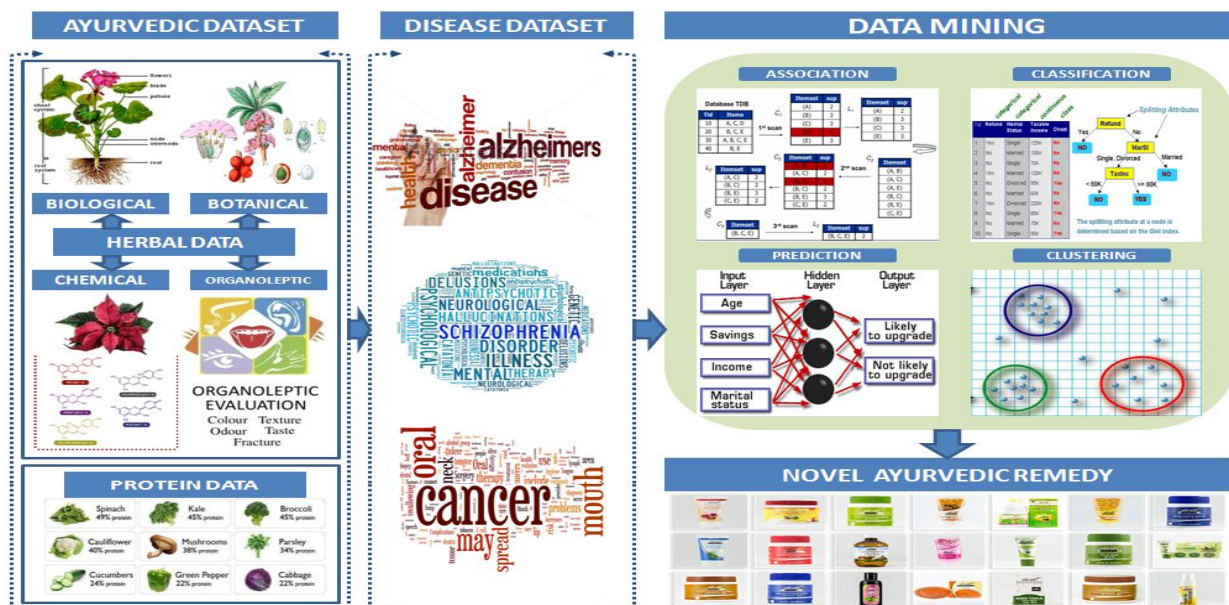


Fig 2: Analyzing Ayurvedic and Disease datasets using data mining techniques for preparing novel ayurvedic remedy

IT has major role to play in these areas in bringing the tools to manage the high throughput experiments, the data they generate and sharing & integrating all the data in a meaningful way resulting into the detailed models of complex systems.

*Bioinformatics* data play an imperative role and emerges as a business model for the medical and the pharmaceutical sector as well. Key areas such as gene prediction, data mining, protein structure modelling and prediction, protein folding and stability, macromolecular assembly and modelling of complex biological systems are thriving. Information Technology with Biological aspects brings up innovative domain culture with huge amount of experimental results which produces an integrated collection of data in a sustained reproduction of biological pathways.

## 4. REVIEW OF LITERATURE

When an individual consumes an incorrect medication or extreme quantity of medication and departs life, is there anything to be charged against the medication? Indeed, it is the fault of the *statistic discrimination* that occurs in an ignorant individual. He distrusts this indiscrimination and

believes in the ancient practices of traditional medicines. Though Ayurveda provides a remedy for any illness, still the basic research methodology being adopted has not found any variation over 2000 years now. *Neither the research methodology nor the text book has changed.*

In the context of bringing out a new research methodology for Ayurvedic science, many have worked on what is actually there and what is still need to be adopted. Excerpts from Dr. M.S Baghel[3] on the need of new research methodology for Ayurveda, stated that a research must always convert data into information, information into knowledge and knowledge into wisdom. This transformation must remain fair, complete and stress the literary, clinical and experimental researches to create an impact in academic, pharmacy and day to day practices. He further strongly annotated that the research methodologies being carried out in Ayurveda has not changed for the last 50 years. Hence, new research methodologies are needed to be planned and adopted for this traditional medicine system to have sustenance. In an account on the similar note Prof. RH Singh[4] has stated on the exploring issues in the development of Ayurvedic research methodology and has

identified that the enduring research is leading to a manner that it provides further strengthening of the modern medicine than Ayurveda. These kinds of development of the modern medicine are of no use to the Ayurvedic practitioners or academicians. He further stated that even WHO in the year 2000, has made sincere efforts to bring out new research methodologies for Ayurveda. He strongly suggested that an in-depth education to the practitioners, support from the government and increased amount of Ayurvedic research institutions are required for generating new research methodologies for Ayurveda. To supplement the requirement of new research methodology, Dr. Nandini Kumar in a personal communication to Dr. M.S Baghel[3], has also stated that one has to urge himself/herself on the current practices carried out on Ayurveda to overcome the challenges offered by modern medicines. She powerfully speaks out that Ayurvedic research protocols must always be conferred with Ayurvedists to produce interpretive evidences.

Nevertheless, these discussions are not the end, it is just the beginning. Unless there arises a new research methodology which considers all the fundamental principles of Ayurveda, it will always remain an *exciting inactive magnificence*. This kind of magnificent illusion about Ayurveda is appropriate to a number of stoppages of modern medicine in providing remedy to diseases like cancer, skin disorders, metabolic disorders, etc. However, Ayurveda has much more to offer than just a magnificent illusion. The identifications and approaches put forth by these eminent scholars remain suitable for laying a roadmap to new research methodologies.

Having been into the era of Information technology, the world is running faster than ever and IT gets incorporated with all the fields and breaks down all the predefined rules and brings out new ways to approach a problem and forms multiple traditions of research methodologies. IT did not leave Ayurveda either. Hence, the formation of *Ayurinformatics*-the process of incorporating Information technology, Data mining, statistical methods and Ayurveda. It utilizes the potentials of varied number of result oriented data mining techniques and information technology solutions, which are available to analyze ayurvedic databases and provide a new range of research methodology solutions. Predictions propose that this development will persist for a predicted future. Research in the area of Ayurinformatics has gone beyond a number of paths these days.

This new venture has already provided quite a few good results. On the results of clinical research, by applying interpretation of statistical hypothesis testing, Sanjeev B.Sarmukaddam[5] stated that clinical significance and statistical significance must always be kept in brain when interpreting the results of any clinical research. He observed that to obtain a better estimation rate on the complication of identifying a better treatment method, confidence intervals must be computed. This confidence interval rate will be better if one obtains a specific random sample for the observation. *Statistics is all about observations*. He enforced that some of the statistical methods like, simple random sampling, binomial distribution, etc. have got a certainty with which one can estimate the size of treatment effect. He affirmed that practice of medicine depends greatly on the sample size. Larger sample size is much more reliable than a smaller sample size. One can strongly believe that such a kind of statistical approach in research must be incorporated in Ayurvedic research as well. In a literary research by Athikajan, Shreevathsa, Rajendra V [6], from biostatistical perspective, interpretation of Ayurvedic terminologies in

research must be inculcated. They further identified that any medicinal research must be practical, precise and uttered in figures, which is named as *statistics*. Now a day, Ayurvedic scholars have given outstanding regard to the importance of data mining based statistical methods. The data mining based statistical methods are the need of the hour to explore with new research methodologies of Ayurveda and to standardize the quality of research in Ayurveda. They further affirmed that numerical terminologies in Ayurveda are not considered much for better treatment method identification. Hence, to standardize the Ayurvedic variables, data mining based statistical mining is required for Ayurvedic scholars to attain accurate results and global acceptance.

It is a strong view that the management of data mining based statistical methodologies in ayurvedic research will definitely yield a significant development in Ayurvedic research. Hence, if the desired results of actual ayurvedic clinical practices are to be evidenced, the procedure should be prepared with the support of information technology and statistical methods through data mining.

Some of the following works have clearly indicated that data mining based ayurvedic research methodology has certainly yielded a better solution to various treatment methods. Early work by Kajal C.Agrawal and Meghana Nagori [7] was concerned with k-means algorithm to create Ayurvedic medicinal cluster. This algorithm is an iterative, non-deterministic, numerical and unsupervised method for identifying clusters. To form different ayurvedic clusters they compared k-means and improved k-means algorithm for four different diseases. They identified that the marginal distance of each ayurvedic data object to its cluster center was avoided by this improved method which produced a better accuracy, clustering speed and running time. In a report on a special case to identify ayurvedic medicinal plant diseases, detection of well built edges of plant leaves were required. Prof. T.N.Janwe and Vinita Tajane[8] dealt with Canny Edge Detection algorithm to sketch the edges for testing the diseases of plants. To identify whether a plant is healthy or infected, they used color histogram to separate the layers, for designing RGB layer histogram and to monitor the power of every pixel color. This is one such worthy way of incorporating IT with Ayurveda medicinal plants. There are many fundamental principles in Ayurveda which can be helpful in understanding the maladies in modern medicine. In an infinite research domain on the extraction of knowledge from huge unstructured documents, thus otherwise termed as *Text Mining*, Prof. Vaka.H.C.G and S.Mukhopadhyay [9] have explored documents belonging to Ayurveda, which is a databank called *PubMed*. They found new associations among various objects in biological nature. For extracting texts from the biological databank, they used Automated Vocabulary Discovery (AVD) algorithm. This algorithm identifies new associations among ayurvedic texts and allots a numerical value to them. They identified that if the score is high then there exists a potential new interesting association among ayurvedic objects.

Hence, all the above works stated are an extensive array of real life circumstances wherein one can apply ayurinformatics as a quicker way for predicting various ayurvedic treatment methods and enhancing the research methodology being planned and adopted in Ayurveda. This requires massive amount of further research institutions, scholars and organizations from both government and private sectors to extend and build this progressive, traditional and everlasting medicinal system.

## 5. CONCLUSION

Thus it becomes certain that to have an innate remedy for the current issues in health care sector; Ayurveda needs better research methodologies and user friendly Information Technology tools. Hence, Ayurveda along with data mining, statistics and bioinformatics, predominantly supports any ailment over allopathic medicine. It is identified that it just requires to be updated to the current smart mobile world with new research methodologies. This incorporation will definitely help nations to form a new era of informatics world called *Ayurinformatics – An active magnificence*. The future challenges are massive but the remedy is undeviating. The projected approach has attained a far better quicker way of identifying a treatment method than earlier techniques used. For future scope, as the practices adopted and planned in this paper are largely based on the assimilation of ayurvedic, information technology and data mining algorithms, further examination would be on the efficiency of utilizing machine learning algorithms or more statistical advances. With this work the present practice focuses mostly on using data mining algorithms alone. Hence, the intention is to further compare different data mining and statistical algorithms to be more *disease specific*, so as to make the cure for any ailment is no more be an *outlying objective*.

## 6. REFERENCES

- [1] Anand V. Saurkar, Vaibhav Bhujade, Priti Bhagat, Amit Kharpade, A Review Paper on Various Data Mining Techniques, International Journal of Advanced Research in Computer Science and Software Engineering, Volume 4, Issue 4, April 2014
- [2] Sushant Sud, Khyati S. Sud, Ayurinformatics: Need of the hour, International Journal of Research in Ayurveda Pharmacy, 4(2), Mar –Apr 2013
- [3] Dr. M.S. Baghel, Need of new research methodology for Ayurveda, Journal of research in Ayurveda, 2011 Jan-Mar; 32(1)
- [4] Singh RH. Exploring issues in the development of Ayurvedic research methodology, Journal of Ayurveda and Integrative Medicine 2010
- [5] Sanjeev B. Sarmukaddam, Interpreting “statistical hypothesis testing” results in clinical research, Journal of Ayurveda & Integrative Medicine | April-June 2012 | Vol 3 | Issue 2
- [6] Athikajan, Shreevathsa, Rajendra V, Interpretation of Ayurvedic Terminologies from Biostatistical perspective in Research, International Journal of Research in Ayurveda and Pharmacy, Mar-April 2013
- [7] Kajal C. Agarwal, Meghana Nagori, Clusters of Ayurvedic Medicines Using Improved K-means Algorithm, Proc. of the Intl. Conf. on Advances in CSEE 2013
- [8] Prof. N.J. Janwe, Vinita Tajane, Medicinal Plants Disease Identification Using Canny Edge Detection Algorithm, Histogram Analysis and CBIR, IJARCSSE, Volume 4, Issue 6, June 2014
- [9] Prof. Vaka H.C.G and S. Mukhopadhyay, Hypotheses Generation Pertaining to Ayurveda Using Automated Vocabulary Generation and Transitive Text Mining, International conference on Network based information systems 2009
- [10] Preenon Bagchi, Nikita Sinha, Ajit Kar, Mahesh. M, Ayur-Informatics: Establishing an Ayurvedic Remedy for Bronchial Carcinoma, International Journal of Bioscience, Biochemistry and Bioinformatics, Vol. 1, No. 1, May 2011