International Journal of Science and Research (IJSR)
ISSN (Online): 2319-7064

A Fuzzy Pattern Application in Multiple Disease Diagnosis on Large Database: Review

Namrata D. Ghuse [1], S. R. Gupta [2]

[1, 2]Department of Computer Science & Engineering, PRMIT, Amravati, Maharashtra, India

Abstract: ***Basically, the Medical diagnosis process can be interpreted as a decision making process, during which the physician induce the diagnosis of a new and unknown case from an available set of clinical data and from his/her clinical experience. Data mining techniques can extract relationships and patterns holding in this wealth of data, and thus be helpful in understanding the progression of diseases and the efficacy of the associated therapies. Traditionally the enormous quantities of medical data are utilized only for clinical and short term use. Medi-Query puts to use this vast storage of information so that diagnosis based on this historical data can be made. There are systems to predict diseases of the heart, brain and lungs based on past data collected from the patients. We focus on computing the probability of occurrence of a particular ailment from the medical data by mining it using a unique algorithm which increases accuracy of such diagnosis by combining Neural Networks, Bayesian Classification and Differential Diagnosis all integrated into one single approach. The system uses a Service Oriented Architecture (SOA) wherein the system components of diagnosis, information portal and other miscellaneous services provided are coupled.***

Keywords: Differential Diagnosis, Bayesian Classification, SOA, LAMSTAR, K-NN Method**.**

1. Introduction

Since the advent of advanced computing, doctors have always made use of technology to help them in various possible ways, from surgical imagery to X-ray photography. Unfortunately, technology has always stayed behind when it came to diagnosis, a process that still requires a doctor's knowledge and experience to process the sheer number of variables involved, ranging from medical history to climatic conditions, blood pressure, environment, and various other factors. The number of variables counts up to the total variables that are required to understand the complete working of nature itself, which no model has successfully analyzed yet. To overcome this problem, medical decision support systems [29]–[31] are becoming more and more essential, which will assist the doctors in taking correct decisions. Medical decision is a highly specialized and challenging job due to various factors, especially in case of diseases that show similar symptoms, or in case of rare diseases. The factors leading to misdiagnosis may vary from inexperience of the doctors, habitual and repetitive diagnosis by experienced doctors, stress, fatigue, and other occupational conditions, and also due to factors including, but not limited to misinterpretation, ambiguous symptoms, and incomplete information. Conventional algorithms completely overlook various variables involved such as prevailing conditions, the build-ups resulting in the symptoms, medical history, family history, and other factors relating to the patient, due to sheer magnitude of available unknown variables. Experienced doctors generally classify such diseases based on the differential diagnosis [12] method. This involves doctors narrowing down the diseases to the root disease out of the list of diseases that show similar symptoms. This is done using their knowledge and experience, and it is later confirmed by performing various tests. In case of rare diseases or diseases with similar symptoms, due to the number of tests involved, it might not be always feasible. Especially in developing countries, the problem of lack of trained and experienced doctors leads to intensification of this problem [32]. This process of differential diagnosis has been emulated in the system proposed in this paper, thus making this rather tough task a lot easier. This method is further modified and enhanced to reduce the huge number of underlying variables to just one by finding the root disease, or the most probable disease, using smart pattern matching involving *k*-NN classification technique [26] and the next probable diseases by performing differential diagnosis, using the Hopfield neural networks theory [33] and Large Memory Storage and Retrieval (LAMSTAR) Networks [17]. Using all these, and by utilizing a database having a comprehensive list of medical history at the disposal of this system, the probability of occurrence of a disease may be calculated, regardless of the various unknown variables. The algorithm will output the disease from the symptoms entered and also gives the next highly probable disease, and thus, the most effective course of action to be performed can be determined. This system, built on service-oriented architecture (SOA) [14], has been implemented on a web server, so that it can be accessed by anyone with an Internet connection. After successful implementation of the system, not only will this be accessible by most doctors, this can also be used by doctors in rural and remote areas, with a computer and Internet, or even a mobile phone. The doctors or any medical personnel have to enter the symptoms of the disease. The system, making use of various techniques mentioned, will in turn display the root disease along with the set of most probable diseases which have similar symptoms. This system will give the doctors the list of diseases that the patient has maximum probability of suffering from. This, in turn, will help the doctors to recommend specific tests corresponding to the diseases in the list, thus reducing the number of non consequential tests and resulting in saving time and money for both the doctor and the patient.

2. Literature Survey

Shamsul I. Chowdhury [15] discusses issues related to the analysis and interpretation of medical data in 1994, thus allowing knowledge discovery in medical databases. He also showed that knowledge can also be effectively extracted from a database of patient observations and from interpretation of those observations. The system shows how retrospectively collected data could be utilized for the purpose of knowledge extraction. The main emphasis was to Study the feasibility of the approach exploring a large patient record system. The analysis was carried out to test the hypothesis of a possible causation between hypertension and diabetes.

Basically, the rnedical diagnosis process can be interpreted as a decision making process, during which the physician induces the diagnosis of a new and unknown case from an available set of clinical data and from his/her clinical experience. At the University of Calabria in Italy, the medical decision making process has been cornputeriized, Physicians at the Cosenza General Hospital currently arc using the diagnostic decision support systern lo help thern with the timely identification of breast cancer in patients through The application of a well-defined set of classification data. Dr. Mirnmo Conforti presented the system before the ITTS-TTAB'99 audience in 1999 & he explained the architecture horn this particular poiul of view, emphasizing the powerful efficiency and effectiveness of Mathematical Programming approaches as the basic tools for the design of the CAMD or Computer Aided Medical Diagnosis system. Mimmo Cnnforti addresses our attention to cancer early detection on the basis of small arnount of clinical information.

Hubert Kordylewski[17] , Daniel Graupe[16] describes the application of a large memory storage and retrieval (LAMSTAR) neural network to Medical diagnosis and medical information retrieval problems in the year 2001. The network also employs features of forgetting and of interpolation and extrapolation, thus being able to handle incomplete data sets. Applications of the network to three specific medical diagnosis problems are described: two from nephrology and one related to an emergency-room drug identification problem.

Jenn-Lung Su, Guo-Zhen Wu [19] introduced the database concept has been widely used in medical information system for processing large volumes of data in 2001. Symbolic and numeric data will define the need for new data analysis techniques and tools for knowledge discovery. Three popular algorithms for data mining which includes Bayesian Network (BN), C4.5 in Decision Tree (DT) , and Back Propagation Neural Network (BPN) were evaluated. The result shows that BN had a good presentation in diagnosis ability.

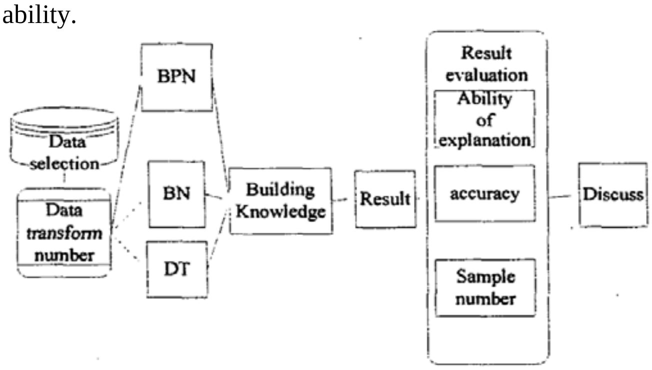

Figure 1: Procedure of Knowledge Discovery

Peter Kokol, Petra Povalej, Gregor Stiglic1, Dejan Dinevski [25] elaborates the use of self organization to integrate different specialist's opinions generated by different intelligent classifier systems with a purpose to increase classification accuracy in 2007. Early and accurate diagnosing of various diseases has proved to be of vital importance in many health care processes. In recent years intelligent systems have been often used for decision support and classification in many scientific and engineering disciplines including health care. However, in many cases the proposed treatment, prediction or diagnose can differ from one intelligent system to another, similar to the real world where different medical specialists may have different opinions The main aim here is to mimic this real world situation in the manner to merge different opinions generated by different intelligent systems using the self organizing abilities of cellular automata.

Michele Berlingerio, Francesco Bonchi, Fosca Giannotti, Franco Turini [24] introduced the concept of Time-Annotated Sequence (TAS) in 2008.Time-annotated sequences (TAS), is a novel mining paradigm that solves this problem. Recently defined in our laboratory together with an efficient algorithm for extracting them, TAS are sequential patterns where each transition between two events is annotated with a typical transition time that is found frequent in the data. The TAS mining paradigm is applied to clinical data regarding a set of patients in the follow-up of a liver transplantation. The aim of the data analysis is that of assessing the effectiveness of the extracorporeal photpheresis (ECP) as a therapy to prevent rejection in solid organ transplantation.

Lishuang Li, Linmei Jing, Degen Huang [26] This paper presents a novel method to extract Protein-Protein Interaction (PPI) information from biomedical literatures based on Support Vector Machine (SVM) and K Nearest Neighbors (KNN) in 2009. A model based on SVM is setup to extract the interaction. To improve the accuracy of SVM classifier, KNN method is introduced. Furthermore, to fit the unbalanced data distribution, a modified SVM-KNN classifier is proposed. The two protein names, words between two proteins, words surrounding two proteins, keyword between or among the surrounding words of two protein names, Exp Distance based on word distance of two proteins, Pro Distance between two proteins in a protein pair are extracted as the features of the vectors. Experimental results show that this approach can achieve higher F-score in extracting PPI information than sole SVM classifier and original SVM-KNN classifier, and the model especially fits the unbalanced data distribution.

Demosthenes Akoumianakis, Giannis Milolidakis, Anargyros Akrivos, Zacharias [23] elaborates on the concept of transformable boundary artifacts and their role in fostering knowledge-based work in cross-organization virtual communities of practice in 2010. The domain of investigation is clinical practice guidelines development for cancer. The distinctive characteristic of the approach presented earlier is that it fosters a computer-mediated practice for clinical guideline development. This practice inherits engineering and social properties required to facilitate guideline development through cycles of 'conception–elaboration– negotiation – reconstruction'.

Rebeck Carvalho, Rahul Isola, Amiya Kumar Tripath [29]introduced the concept of Medi-Query in 2011. Traditionally the enormous quantities of medical data are utilized only for clinical and short term use. Medi-Query puts to use this vast storage of information. so that diagnosis based on this historical data can be made. There are systems to predict diseases of the heart, brain and lungs based on past data collected from the patients. We focus on computing the probability of occurrence of a particular ailment from the medical data by mining it using a unique algorithm which increases accuracy of such diagnosis by combining Neural Networks, Bayesian Classification and Differential Diagnosis all integrated into one single approach. It will also help the medical fraternity in the long run by helping them in getting accurate diagnosis and sharing of medical practices which will facilitate faster research and save many lives.

Medical data are an ever-growing source of information generated from the hospitals in the form of patient records. When mined properly, the information hidden in these records is a huge resource bank for medical research. As of now, these data are mostly used only for clinical work. Rahul Isola, Rebeck Carvalho Amiya Kumar Tripathy [27] introduce a system which uses Hopfield networks, LAMSTAR, and *k*-NN, an attempt has been made to assist the doctors to perform differential diagnosis. The system proposes an innovative utilization of the misdiagnosis factor for differentia diagnosis along with a possible method of implementation using the SOA technique in 2012. The possibility of usage of vastly available EHR data for the purpose allows latest and continuously updated medical data available to the system. In the field of medical diagnosis, there is always the scope for uncertainty. This system has been built on the experience of doctors only, so there will always be a scope for ambiguous or uncertain diagnosis.

Table 1: Comparison of diagnosis techniques

Sr.	*Year*	*Author*	*Advantages*
1	1994	Shamsul I. Chowdhury Gustavsson R.	It introduce a issue related to analysis & interpretation of medical data.
2	1999	Dr. Mirnmo Conforti	For cancer early detection
3	2001	Hubert Kordylewski, Daniel Graupe	LAMSTAR is used for info retrival & also provides interpolation & extrapolation of input data based on stored info.
4	2001	Jenn-Lung Su, Guo-Zhen Wu	It uses Bayesian n/w tech. for knowledge discovery.it gives better accuracy in diagnosis of breast & tumor.
5	2007	Michele Berlingerio, Francesco Bonchi, Fosca Giannotti, Franco Turini	It uses TAS Tech.It prevents the rejection in solidorgan transplantation
6	2008	Peter Kokol, Petra Povalej, Gregor Stiglic1, Dejan Dinevski	It uses intelligent system to increase classification accuracy
7	2009	Lishuang Li, Linmei Jing, Degen Huang	They presents a new methods to extract Protein-Protein Interaction (PPI) information frombiomedical literatures based on Support Vector Machine(SVM) and K Nearest Neighbors (KNN).
8	2010	Demosthenes Akoumianakis, Giannis Milolidakis, Anargyros Akrivos, Zacharias	It gives us guideline management information system.
9	2011	Carvalho, Rahul Isola, Amiya Kumar Tripathy	It introduced a MediQuery, which help the medical fraternity in the long run by helping them in getting accurate diagnosis
10	2012	Carvalho, Rahul Isola, Amiya Kumar Tripathy	They introduce the new methods to differential diagnosis.

3. Data Mining In Medical Database

The medical database is a massive database which maintains different types of medical data types such as treatment records, patient history, medication profiles, pathology report, radiology reports, signal and images. Medical care data are characterized by their complexity and heterogeneity with respect to the data type. These data are highly dimensional, uncertain, distributed and may be noisy with incorrect, invalid or missing values [18]. Uncover patterns in such type of data using the traditional statistics methods is difficult if not unattainable [6]. Hence, data mining techniques in medical data provide more effective analyzing methods to improve diagnostic capabilities and so patient care. Data mining in medical records relates to the idea that there is more knowledge hidden in these data than readily accessible. To discover this knowledge, groups of techniques instead of a single technique are commonly applied on medical data for different purposes and to answer critical questions such as: given a record of cancer patient, what can be done to improve the treatment of this patient. Furthermore, data mining tasks in medical database can be categorized to tasks of description and prediction data mining aims to find human-interpretable patterns and associations after considering the entire data and constructing a predictive model seeks to portend a response of interest. Although goals of the two categories may overlap (the models generated by predictive tasks may show some interesting patterns); the main different is that prediction requires the data to include a special categorical or numerical response variable [30]. In this paper, the focus is on the applications of predictive data mining techniques to support clinical decision making. Predictive data mining techniques maybe applied to construct decision model or system for clinical procedures such as prediction, diagnosis and treatment plans. Generally, predictive data mining techniques are classified into three categories: association rule; classification and clustering. Association refers to the discovery of associations or relationships among item sets or objects. For medical data example, it may discovers that a set of indications or symptoms frequently occur together with another set of symptoms. The A priori algorithm is an example of association techniques. Classification maps data items into one of several pre-defined classes. For example, classification rules about a disease can be extracted from previous known cases and then used to diagnose new patients of this disease based on their symptoms. Decision Trees (DT), Bayesian Networks (BN), Naïve Bayes (NB) and Support Vector Machine (SVM) are examples of classification approaches. Clustering recognizes the class or cluster for a set of unclassified objects according to their attributes. For example, a set of diseases can be grouped into several clusters based on the similarities in their symptoms, and the common symptoms of the diseases in a cluster can be used to describe or predict that group of diseases. The K-nearest neighbor (K-nn) is the most popular method of clustering.

4. Computational Techniques Used In Medical Database

There are different computational techniques used in medical system for data mining.

A) LAMSTAR

LAMSTAR is the application of a large memory storage and retrieval (LAMSTAR) neural network to several medical diagnosis problems. The neural network discussed in this paper is a network specifically designed for large-scale memory storage and retrieval of information. The LAMSTAR network attempts to store and retrieve patterns in a computationally efficient manner, using tools of neural networks, especially self-organizing map (SOM)-based network modules, combined with statistical decision tool .The input word is coded in terms of a real vector given by
$\underline{X}=[\underline{x}1^T,\underline{x}2^T,\ldots.\underline{x}^T_N]$
where denotes transposition.

All neurons of an SOM module are checked, in the LAMSTAR network, only a finite group of p neurons is checked at a time due to the huge number of neurons involved (the large memory involved). The final set of p neurons is determined by the weights. A winning neuron is determined for each input based on the similarity between the input and a weight vector (stored information). For an input sub word, the winning neuron is determined by minimization of a distance norm ||*|| given by
$||X_i-W_{i,m}||=\min||X_i-W_{i,k}||$ VK€<l, l+p>;l~$\{N_{i,j}\}$

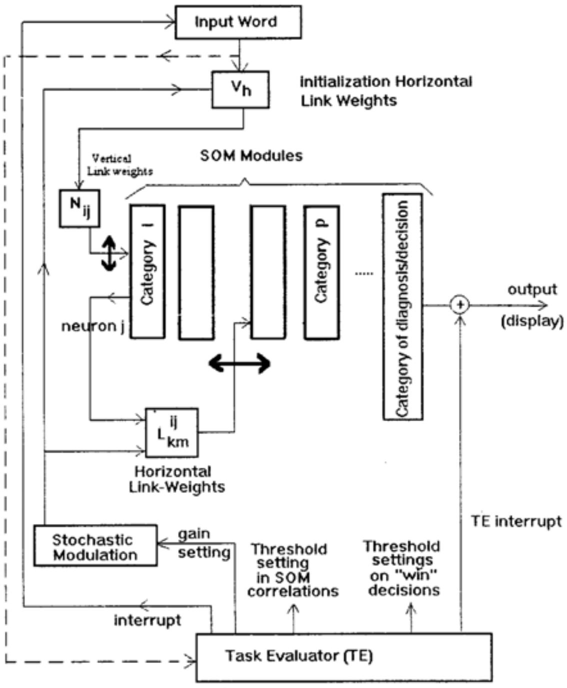

Figure 2: General block diagram LAMSTAR network

B)Support Vector Machine

The Support Vector Machine (SVM) is a classification. Algorithm in statistical learning theory [34]. It can provide accurate models because it can capture nonlinearity in the data. The classification tasks are performed by maximizing the margin separating both classes and minimizing the classification errors [34]. The training of SVM involves the optimization of a convex cost function where the learning process is not complicated by local minima [35]. The testing used the support vectors to classify a test dataset and the performance is based on error rate determination [35]. For a training set of l samples, the learning procedure is as the follows:

$$\min_{\alpha} : \frac{1}{2}\sum_{i=1}^{l}\sum_{j=1}^{l} y_i y_j \alpha_i \alpha_j K(x_i, x_j) - \sum_{j=1}^{l} \alpha_j .$$

$$\text{s.t.} \quad 0 \le \alpha_i \le C, i = 1, \cdots, l .$$

$$\sum_{i=1}^{l} \alpha_i y_i = 0 .$$

$$f(x) = \operatorname{sgn}\left(\sum_{i=1}^{n_s} y_i \alpha_i^* K(x_i, x) + b^* \right) .$$

The y_i is the label of the i^{th} sample x_i [33]. The ai is the Langrangian multiplier of x_i. The C is the upper bound of ai and K (x_i, x_j) is the kernel. The samples with a > 0 are called support vectors [12]. The decision function is as follow, where ns are the number of support vectors [33]:

$$f(x) = \operatorname{sgn}\left(\sum_{i=1}^{n_s} y_i \alpha_i^* K(x_i, x) + b^* \right) .$$

In this section, we have introduced four famous data mining techniques. In the subsequent section, the importance of data mining in medical systems is briefly discussed.

C .Bayesian Classifiers

The Bayesian classifiers have a structural model and a set of conditional probabilities [16]. The structural model is represented as a directed graph where the nodes represent attributes and arcs represent attribute dependency. A representation of a Bayesian Classifier Structure [36].

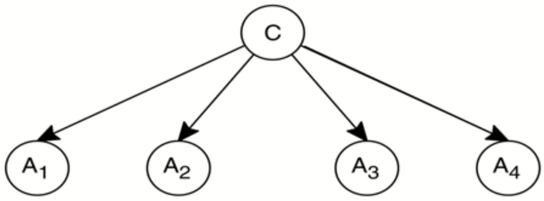

Figure 3: A Representation of a Bayesian Classifier Structure

For classifications, the Bayesian networks are used to construct classifiers from a given set of training examples with class labels. The classifier of a general Bayesian network:

$$c(E) = \arg\max_{c \in C} P(c)P(a_1, a_2, \ldots, a_n|c).$$

The a1 to an refers to the attributes or nodes in the Bayesian network [34]. The variable C represents the class variable that refers to the class node in a Bayesian network. The variable c represents the value of C and c(E) denotes the class of E. For Naïve Bayes (NB), all the attributes have to be assumed as independent. Therefore, the definition of Naïve Bayes is as follows:

$$c(E) = \arg\max_{c \in C} P(c) \prod_{i=1}^{n} P(a_i|c).$$

The attributes a are independent attributes [34]. For example, X is a child whose class is to be determined. Then, H is the class that child X going to be predicted. The formula to calculate P (H|X) using Bayesian networks is as follows:

$$P(H|X) = \frac{P(X|H)P(H)}{P(X)}.$$

P(H) is the probability of class H, P(X|H) is the posterior probability that X is conditioned on H, and P(X) is the probability of X [12]. These values can be obtained from the data used for training. For Naïve Bayes classifiers, the steps to predict probability for data X with the assumption that all the attributes are independent of each other is as follows:

$P(X \backslash C_i) = \prod_{k=1}^{n} p(Xk, Ci)$

P(X1|Ci), P(X2|Ci),…,P(Xn|Ci) can be calculated from the training samples [33].

5. Summary & Discussion

Predictive models were generated by applying various predictive mining methods and statistical techniques on historical medical data. They were constructed in a single mode, hybrid mode and ensemble-based mode. Hybrid models aim to improve the performance of individual technique and to overcome the weaknesses of any single based-model. Ensemble-based models aim to increase the accuracy the overall classification accuracy by reducing the variance of estimation errors and avoiding a biased decision. Generating effective predictive models are faced by several problems that mainly are the lack of input data, limitations of the construction method and drawbacks of the combination methods. The construction of effective models is constrained by the characteristics and size of datasets to train models the model's construction is also constrained by the capabilities of the technique used for model construction as the model inherits the weaknesses and limitations of its construction method.

6. Conclusion & Future scope

Knowledge is one of the most significant assets of any organization and especially in healthcare environment. Healthcare environment is rich of information; however, creating knowledge out of this information is still a serious challenge. Practical use of healthcare database systems and knowledge discovery and management technologies like data mining can enormously contribute to improve decision making in healthcare. Converting massive, complex and heterogeneous healthcare data into knowledge can help in controlling cost and maintaining high quality of patient care. A variety of data mining techniques have increasingly applied to tackle various problems and challenges of knowledge discovery in administrative and clinical facets of healthcare. In respect to clinical decisions, intelligent data mining tools can contribute effectively to enhance effectiveness of disease treatment and preventions as in the case of heart diseases. With the support of various medicinal practitioners and hospitals, higher probability of getting the diagnosis right can be obtained, compared to what individual doctors can do alone. The system does not give 100% accurate results, which even the doctors cannot do. So it cannot be used as a substitute or a shortcut to diagnosis, as each patient is different. But it can definitely complement the doctor's knowledge and assist them to reach a conclusion. The doctor always has the upper hand to decide whether to use the diagnosis given by the algorithm or not. After sufficient self-learning, with an extensive database of medical records to mine from, this can be used to build formidable medical assistance software that can be of great use to all doctors, and specially the new practitioners and students. It will also help the medical fraternity in the long run by helping them in getting accurate diagnosis and sharing of medical practices which will facilitate faster research and save many lives.

References

[1] Berner, Eta S., ed. Clinical Decision Support Systems. New York, NY: Springer, 2007

[2] Kensaku Kawamoto, Caitlin A Houlihan, E Andrew Balas and David F Lobach , "Improving clinical practice using clinical decision support systems: a systematic review of trials to identify features critical to success", BMJ 330 : 765 doi: 10.1136/bmj.38398.500764.8F (Published 14 March 2005)

[3] Randolph A Miller, "Medical Diagnostic Decision Support Systems—Past, Present, And Future – A Threaded Bibliography and Brief Commentary", JAMIA 1994;1:8-27 doi:10.1136/jamia.1994.95236141

[4] Bell, Michael (2010). SOA Modeling Patterns for Service-Oriented Discovery and Analysis. Wiley & Sons. pp. 390. ISBN 978-0470481974.

[5] Wasan, S., Bhatnagar, V., Kaur, H.m "The impact of data mining techniques on medical diagnostics," Data Science Journal 5, 119–126, 2006.
[6] Cios KJ, Moore GW, "Uniqueness of medical data mining" Intell Med; 26:1—24. 2002
[7] JavaEE at a glance. [Online]. Available: http://java.sun.com/j2ee
[8] P. Herzum, "Web services and service-oriented architectures," CutterDistributed Enterprise Architecture Advisory Service, Executive Report, 2002.
[9] Jiawei Han, Micheline Kamber, Data Mining Conceptsand Techniques, 2011 edition, Morgan KaufmannPublications.
[10] Misdiagnosis: Symptom and heath diagnosis checker. Available at: http://www.misdiagnosis.com.(4th Feb 2011,7.30pm GMT)
[11] R. A. Miller, "Medical diagnostic decision support systems—Past, present, and future—A threaded bibliography and brief commentary," J. Amer. Med. Inf. Assoc., vol. 1, pp. 8–27, 1994.
[12] W. Siegenthaler, Differential Diagnosis in Internal Medicine: From Symptomto Diagnosis. New York: Thieme Medical Publishers, 2011.
[13] J. Han and M. Kamber, Data Mining Concepts and Techniques. SanMateo, CA: Morgan Kaufmann, 2011.
[14] M.Bell, SOAModeling Patterns for Service-Oriented Discovery and Analysis.New York: Wiley, 2010, p. 390.
[15] S. I. Chowdhury.: Statistical Expert Systems - A Special Application areafor Knowledge-based Computer Methodology. Linkoping Studies in Scienceand Technology, Thesis No 104, Department of Computer and Information Science, University of Linkoping, Sweden.
[16] D. Graupe and H. Kordylewski, "A large memory storage and retrieval neural network for adaptive retrieval and diagnosis," Int. J. Software Eng. Knowledge Eng., vol. 8, no. 1, pp. 115–138, 1998.
[17] H. Kordylewski and D. Graupe, "Applications of the LAMSTAR neural network to medical and engineering diagnosis/fault detection," in Proc7th Artificial Neural Networks in Eng. Conf., St. Louis, MO, 1997.
[18] H. Kordylewski, D. Graupe, and K. Liu, "Medical diagnosis applications of the LAMSTAR neural network," in Proc. Biol. Signal Interpretation Conf., Chicago, IL, 1999
[19] G.Z. Wu, "The application of data mining for medical database", Master Thesis of Department of Biomedical Engineering, Chung Yuan University, Taiwan, Chung Li, 2000.
[20] R. Tunkel, B. J. Hartman, S. L. Kaplan, B. A. Kaufman, K. L. Roos, W. M. Scheld, and R. J. Whitley, "Practice guidelines for the management of bacterial meningitis," Clin. Infectious Dis., vol. 39, no. 9, pp. 1267– 1284, Nov. 2004.
[21] E. Davies, P. J. McKenzie, Preparing for opening night: temporal boundary objects in textually-mediated professional practice Available at http://InformationR.net/ir/10-1/paper211.html
[22] Star, S. L. & J. Griesemer, Institutional ecology, 'translations' and boundary objects: Amateurs and professionals in Berkeley's museum of vertebrate zoology, Social Studies of Science, 19, 1989, pp. 387-420.
[23] D. Akoumianakis, N. Vidakis, G. Vellis, D. Kotsalis, G. Milolidakis, A. Plemenos, A. Akrivos and D. Stefanakis, Transformable Boundary Artifacts for Knowledge-based Work in Cross-organization Virtual Communities Spaces, Journal of Intelligent Decision Technologies Vol. 5 (1), 2011, in press.
[24] M. Berlingerio, F. B. F. Giannotti, and F. Turini, "Mining clinical data with a temporal dimension: A case study," in Proc. IEEE Int. Conf. Bioinf Biomed., Nov. 2–4, 2007, pp. 429–436.
[25] Kokol P, Povalej, P., Lenič, M, Štiglic, G.: Building classifier cellular automata. 6th international conference on cellular automata for research and industry, ACRI 2004, Amsterdam, The Netherlands, October 25-27, 2004. (Lecture notes in computer science, 3305). Berlin: Springer, 2004, pp. 823-830.
[26] L. Li, L. Jing, and D. Huang, "Protein-protein interaction extraction from biomedical literatures based on modified SVM-KNN," in Nat. Lang. Process. Know. Engineer., 2009, pp. 1–7.
[27] R. Carvalho, R. Isola, and A. Tripathy, "MediQuery—An automated decision support system," in Proc. 24th Int. Symp. Comput.-Based Med Syst., Jun. 27–30, 2011, pp. 1–6.
[28] Tan, G. & Cbye H., "Data mining applications in healthcare," Journal of Healthcare Information Management. Vol. 19, No.2,2004
[29] R. Carvalho, R. Isola, and A. Tripathy, "MediQuery—An automated decision support system," in Proc. 24th Int. Symp. Comput.-Based Med. Syst., Jun. 27–30, 2011, pp. 1–6.
[30] K. Kawamoto, C. A. Houlihan, E. A. Balas, and D. F. Lobach, "Improving clinical practice using clinical decision support systems: A systematic reviewof trials to identify features critical to success," Br.Med. J., vol. 330, p. 765, 2005.
[31] R. A. Miller, "Medical diagnostic decision support systems—Past, present, and future—A threaded bibliography and brief commentary," J. Amer. Med. Inf. Assoc., vol. 1, pp. 8–27, 1994.
[32] S. F.Murray and S. C. Pearson, "Maternity referral systems in developing countries :Current knowledge and future research needs, "Social Sci. Med., vol. 62, no. 9, pp. 2205–2215, May 2006.
[33] R. Rojas, Neural networks: A Systematic Introduction. Berlin, Germany: Springer-Verlag, 1996, pp. 337–370.
[34] S. Zhang, et al., "Comparing data mining methods with logistic regression in childhood obesity prediction," Information Systems Frontiers, vol. 11, p. 51, 2009.
[35] J. Chen, et al. (2007). A comparison of four data mining models: bayes, neural network, SVM and decision trees in identifying syndromes in coronary heart disease.4491/2007.
[36] Maglogiannis, et al., "An intelligent system for automated Breast cancer diagnosis and prognosis using SVM based classifiers," Applied intelligence, vol. 30, 2007.
[37] L. Jiang, et al., "A novel bayes model: hidden naive bayes IEEE Trans. on Knowl and Data Eng., vol. 21, pp. 1361- 1371, 2009.

[38] T. Kohonen, Self-Organizing and Associative Memory, 2nd ed. Berlin,Germany: Springer-Verlag, 1988.
[39] Li R., Ye S. W., Shi Z., "SVM-KNN Classifie New Method of Improving the Accuracy of SVM Classifier", Acta Electronica Sinica, Vol 30, No 5, pp, 745-748, August 2002.
[40] Temkin J. M., Gilder M. R., "Extraction of protein Interaction information from unstructured text using a context-free grammar", Bioinformatics; Vol 19, No. 16, pp. 2046-2053, June 2003.

Author Profile

Ms. N. D. Ghuse received the B.E. degrees in Computer Science & Engineering from Sipna College of Engineering & management in 2005. Now she is pursing ME (CSE) from Prof. Ram Meghe College of Engineering & management.

Dr. S. R. Gupta received the M.E (CSE), PhD (CSE) degrees in Computer Science & Engineering from Prof. Ram Meghe College of Engineering & management in 2013.