IJCAT International Journal of Computing and Technology, Volume 1, Issue 3, April 2014
ISSN : 2348 - 6090
www.IJCAT.org

Multiple Disease Diagnosis by using Fuzzy Pattern Applications on Large Database

[1] Dr.S.R.Gupta, [2] Ms. N. D. Ghuse

[1, 2] Department Of Computer Sci. & Engg, PRMIT,Amravati(M.H)

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. There are systems to predict diseases of the heart, brain and lungs based on past data collected from the patients. In this paper 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 K-means clustering and Differential Diagnosis all integrated into one single approach. Here we used different techniques for pattern recognisation for the correct diagnosis of the disease which show similar symptoms.

Keywords **- Image Enhancement, fuzzy Logic, K-means Clustering, Image segmentation, Pattern Recognisation, fuzzy pattern**

1. Introduction

Due to 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.

In this paper the proposed system uses this vast storage of information so that diagnosis based on these historical data can be made. It focuses 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 the key points of k-means clustering and differential diagnosis all integrated into one single algorithm. When similar symptoms of disease are found then by using differential diagnosis method, the experienced doctors generally classify such diseases. 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. This process of differential diagnosis has been emulated in the system proposed in this paper, thus making this rather tough task a lot easier. 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.

1.1 Objective

1. To diagnose correct disease.
2. Cauterized patterns properly.

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.

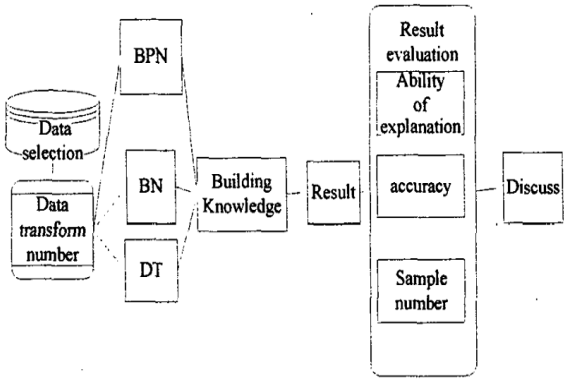

Fig 1: Procedure of Knowledge Discovery

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 guidelinesdevelopment 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.

We know that 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, LAMSTARattempt 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.

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.

3. Proposed Methodology

The Proposed System consist of 2 phases

3.1 Training phase
3.2Testing phase

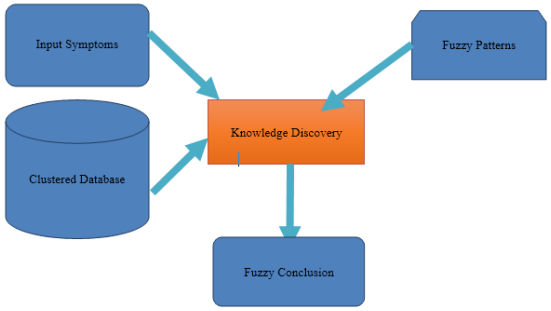

Fig 3.Proposed System Architecture

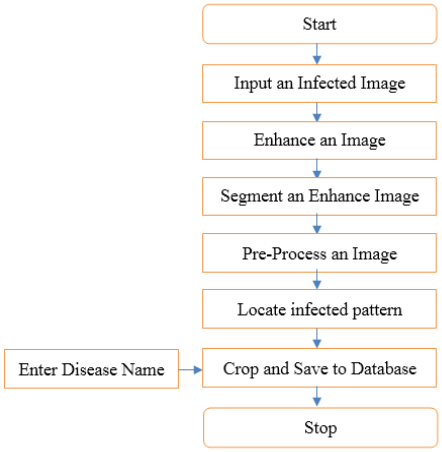

Fig 3. Training Phase

3.1 Training Phase

Training phase is a phase where we create a database by applying fuzzy rules on the various symptoms taken by doctors. Our probable area of working may be cancer, diabetes, Hypertension & some more diseases. The proposed methodologies may contain 50 to 100 symptoms of various diseases.

An Image Enhancement is a process where noisy pixels of an image get filtered.this process caan be achieved with gabor wavelet filter and can be expressed as

$$Fimage = \int_1^{h*w}[(pr)||pg||(pb)] * Ef \quad \text{Eq……..3.1}$$

Where
pr,pg,pb=Image Pixel componants
Ef=Enhancement Factor
h=Height of an Image
W=width of an Image
Fimage=Filter Image

Once an image is filtered and it's noisy pixels are removed,future step is to segment that image either with canny edge detection or sobel edge detection method.Image segmentation have an signification of pattern search and locate.Image Segmentation process can be expressed as
For each pixels line
If
$0< |P_{Back}–P_{forward}|<=10$ Eq……..3.2
Set $P_{forward}$=255(for first hit only).for second, third and onward set P_{Back}=255

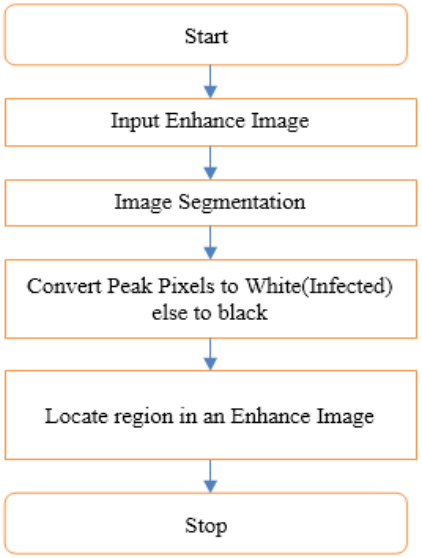

Fig 4.DFD Image Pre-Processing

An image pre-processing in a proposed method is a important step where an infected pattern is located in an enhace image .processing is a step where an extreme level pixels are identified and expressed as

For RGB Image
$(0 \leq pr \leq 10) \;||(245 \leq pr \leq 255)\&\&(pg||pb == 0)$
Eq…3.3
$(0 \leq pg \leq 10) \;||(245 \leq pg \leq 255)\&\&(pr||pb == 0)$
Eq3.4
$(0 \leq pb \leq 10) \;||(245 \leq pb \leq 255)\&\&(pr||pg == 0)$
Eq3.5
For gray scale image

$(0 \leq (pr + pg + pb)/3 \leq 10)||(245 \leq (pr + pg + pb)/3 \leq 255)$
eq……..3.6
In proposed method, we set all those pixels which satisfy these condition to 255means white color.

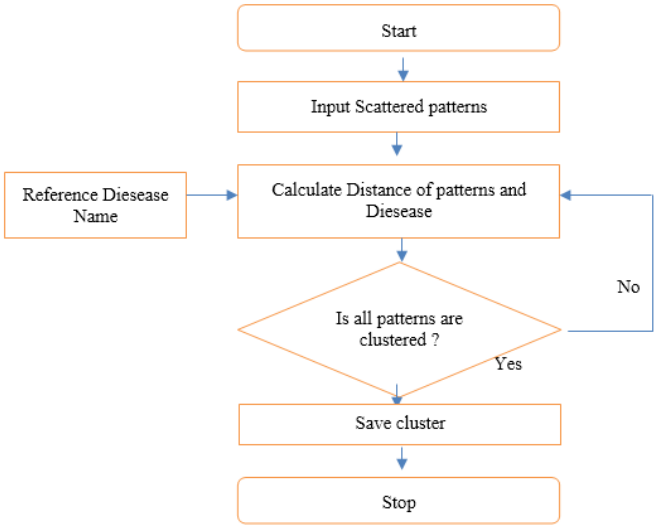

Fig 5.DFD k-Means Clustering

k-means clustering algorithm is used to cluster disease patterns where all patterns are gathred around it's reference disease name.All patterns are clustered as per their Ecludian distance which is a distance between disease and patterns and can be expressed as

$$Ed = |d - p| \qquad \text{Eq…….3.7}$$

A patterns may fall into multiple clusters as per ecludian disstance measures.

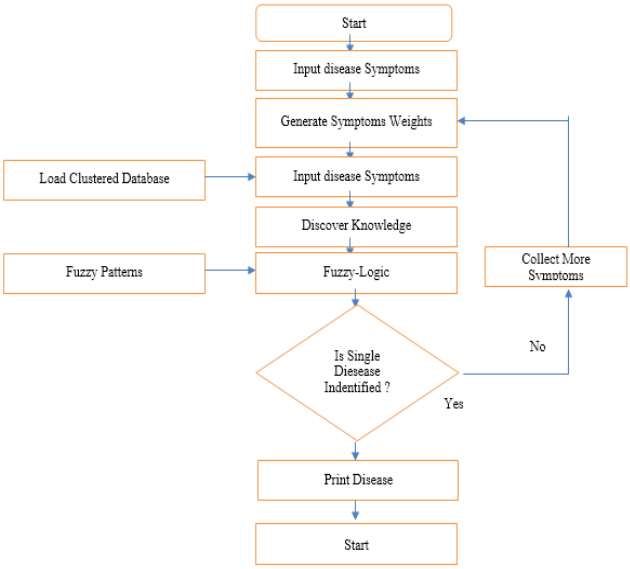

Fig 6. DFD Testing Phase

3.2 Testing Phase

Testing phase is a phase where we apply fuzzy rules on symptoms taken by the doctors, Pre-process it & out a one conclusion .Testing phase may work with the following steps:

1) ***Collect symptoms from patients.***
2) ***Pre-process symptoms.***
3) ***Extract features from database, apply fuzzy rules & out one proper conclusion.***
4) ***If conclusion out is about more than one diseases then proposed methods needs more symptoms & processed from step (2).***
5) ***This method will be iterative from step (2) to (4) until it does not output a single and accurate disease***

3.3 Fuzzy Patterns Rules

Table2. Fuzzy Rule Result

Disease / Pattern	Maleria	Calera	Cancer	Typhoid	Fuzzy Conclusion
fever	√	√		√	Maleria
vomiting	√				Maleria
Skin etching			√		Cancer
Head ach		√		√	Typhoid
Stomach ach		√			Calera
Nose running	√			√	Maleria
cold	√	√			Maleria

Table 3 .Result Analysis

Patterns			Result	Accuracy
Fever	Cold		Maleria	90
Bachache	Deformation of join		Dengue	88
MCV and MCHC are reduced	Increase Swelling		Deficiency Anemia	96
vomiting	fever		Calera	95

4. Conclusion

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.

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 Kaufmann Publications.

[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 InterpretatioConf., 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] A. 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] I. 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.

Author Profile

1)

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

2)

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