RESEARCH ARTICLE   OPEN ACCESS

Differential Disease Diagnosis Based On K-Means Symptoms Clustering and Use of Fuzzy Logic in Adaptive Precautionary Decision

Prof Swati S Jayade, Prof. D. T. Ingole

Department Of Electronics & Telecommunication Government Polytechnic Washim
Department Of Electronics & Telecommunication P.R.M. Institute Of Technology And Research, Badnera-Amravat

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. An automated Decision Support System [1] 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. This algorithm can be used in solving a few common problems that are encountered in automated diagnosis these days, which include diagnosis of multiple diseases showing similar symptoms, diagnosis of a person suffering from multiple diseases, receiving faster and more accurate second opinion, and faster identification of trends present in the medical records.

Keywords: Image Enhancement, fuzzy Logic, K-means Clustering, Image segmentation, Pattern Reorganization, fuzzy pattern.

I. 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.

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

1.1 Aims & Objectives
The main aim of our research work is to target the development of algorithms, to be used in differential diagnosis system, with improved performance in terms of decision accuracy. The images and manual symptoms are inputs for differential diagnosis system as a data and hence, Maintaining the Integrity of the Specificationsuse of these differential diagnosis system will have to be investigated as how to overcome the doctors knowledge.

Following objectives will be accomplished throughout this research work:
To diagnose correct disease.
Clusterized patterns properly.

1.2 Scope
In the medical field it can be used by doctors in rural and remote area with a computer. The common man can also used this system at home with a computer.

1.3 Limitations
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. The current system does not give 100% accurate results as not even the doctors can claim to do so; however, its results are promising. It cannot be used as a substitute or a shortcut to diagnosis, but it can definitely complement the doctors' knowledge and could assist them to reach a conclusion.

II. Review of Literature

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 medical 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 computerized, Physicians at the Cosenza General Hospital currently arc using the diagnostic decision support system lo help them with the timely identification of breast cancer in patients through The application of a well-defined set of classification data. Dr. MimmoConforti 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.MimmoCnnforti addresses our attention to cancer early detection on the basis of small amount 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.
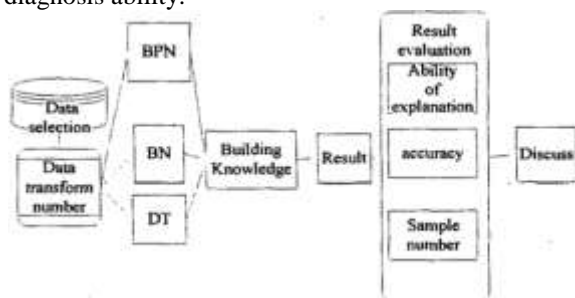

Figure 2.1: Procedure of Knowledge Discovery

Peter Kokol, Petra Povalej, Gregor Stiglic1, DejanDinevski [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, FoscaGiannotti, 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, AnargyrosAkrivos, 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'.

A S Hussein,Wail M Omar, Xue Li, Modafar Al[39] introduces a chronic disease diagnosis recommender system approach based on a hybrid method using multiple classification and filtering. Multiple classification was based on decision tree algorithms. The unified CF method based on learning classification model was used for achieving higher recommendation accuracy. Diabetes disease case study is designed as experiment to show the model.

J.C. Ho, J Ghose, K.P. Unnikrishnan [40] investigates the use of data available in electronic medical records to create a risk prediction model.

They have presented a risk prediction model from EMRs to help address the difficulty of early diagnosis in MS patients. A sparse set of features

were selected to minimize model complexity while maintaining reasonable predictive performance. The results show that it can able to help identify patients at high-risk of developing MS, in spite of a limited sample of patient data. In addition, the models have the ability to generalize to other healthcare systems as they rely only on components commonly found in electronic patient data. The work demonstrates the potential of leveraging EMRs to aid medical professionals with difficult tasks, especially with early disease diagnosis. Future work will focus on incorporating temporal components, such as time of diagnosis, into the model, decreasing the false positive rate, and integrating a larger control population.

AnantMadabhushi, Ajay Basavanhally, Scott Doyle, Shannon Agner, George Lee[41] have highlighted some of the projects in computer-aided prognosis(CAP).While the highlighted projects were focused on prostate and breast cancer outcome prediction, one can easily conceive of similar models being applied to survival prediction for other diseases. Other groups are applying similar techniques to predicting survival outcome in the context of follicularlymphomas. Further developments in this area will only come about by close and dedicated interactions between computer and imaging scientists, clinicians, on cologists, radiologists, and pathologists.Asha.T., Dr. S. Natarajan, Dr. K.N.B. Murth[42] have apply ensemble classification techniques, Adaboost, Bagging and Random forest for classifying tuberculosis (TB). The data is obtained from the state hospital which mainly includes twelve preliminary symptoms (attributes).The data is classified into two categories namely pulmonary tuberculosis (PTB) and retroviral PTB i.e. TB along with AIDS. Evaluation measures such as sensitivity, specificity and accuracy are used for comparison. Random forest is found to be weak with 93% accuracy against 97% that of Bagging and 96% of Adaboost.

Yu-Hsiang Wu, Jhu-Yun Huang, Shyi-Chyi Cheng, Chen-Kuei Yang1, Chih-Lang Lin, Chang Gung[43] have presented the method to construct an automatic system for liver disease prediction. The contributions of the paper include: (1) a preprocessing method is proposed to extract effective features from the selected ROIs of the input image; (2) using a genetic algorithm, a training procedure is proposed to optimize the selection of ROIs and to obtain the stable parameters for classification; (3) an AdaBoost based classification approach is presented to predict the possibility of liver cirrhosis for a patient based on the patient database and the input ultrasound image. As compared with past approaches, they have span a new way to unify the processing steps in a clinical application using the evolutionary algorithms

for interpreting ultrasound images. Future work would be on increasing the size of patient database and generalized the proposed method to deal with ultrasound image sequences. Both of them would improve the prediction accuracy of the system.

K. Kosasih, U.R. Abeyratne, V. Swarnkar[44] have shown that there is useful cough sound information in much higher frequency range than is traditionally used in cough sound analysis. Based on available data, there is unique information in frequencies higher than is traditionally used in clinical settings, especially in frequencies beyond 15 kHz. When it comes to diagnosis of the disease groups considered, however, the results of this study should be interpreted with due care as the number of subjects was limited four.

Imran A. Khan[45] The Personalized Electronic Health Record System for Monitoring Patients with Chronic Disease (PEHRS-MPCD) is innovative as it is designed to continually monitor the patient as he can provide feedback whenever he wishes to. Furthermore it is interactive, flexible and provides a comprehensive solution to the meet the needs of persons with multiple chronic diseases. The goal of the application is to improve the efficiency of the diagnosis and the monitoring of information and ensure quality of care. Future work on the application would focus on the following 3 main areas: Fully developing and modifying the functionalities of the app Performing usability and robustness testing for both bugs and efficiency of algorithm of the app.

Usha a Eswaran, Vivek Eswaran, Sudharshan. V.B[46] gives Automatic drug delivery system based on bio sensors which functions like a human. An experimental setup is developed to validate the performances of Matlab model against Nanohub model. It is proved that the mathematical model developed in this work exhibits similar behaviour as compared to existing nanosensor models with less than 2% deviation. Three different sensors are analyzed for its performances. From the simulation results obtained, it is found that nanowire sensor is suitable for biosensing applications. Developing mathematical models and validating the performances of the developed model is another major contribution of this research work.

Anjali Joshi, and VidhyacharanBhaskar [47] presents Multi-resolution approach using genetic algorithm tool to achieve spatial and functional mapping between two medical images. Noise is added in the images while acquisition. If the images are pre-processed to remove this noise and then registered, they can get the exact mapping of the required information in the images. This helps to detect the minute injuries, very small tumors and any unwanted information in images. Image registration

using multi- resolution approach will be helpful for perfect diagnosis and treatment planning.

RebeckCarvalho, Rahul Isola, Amiya Kumar Tripathy [1] 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 [32] introduce a system which uses Hopfield networks, LAMSTAR 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.

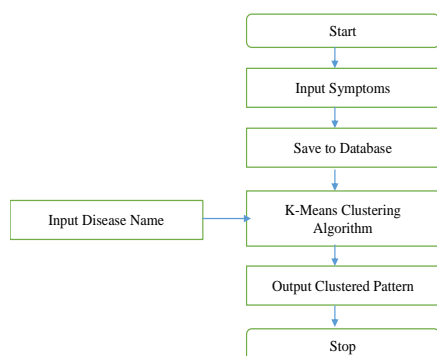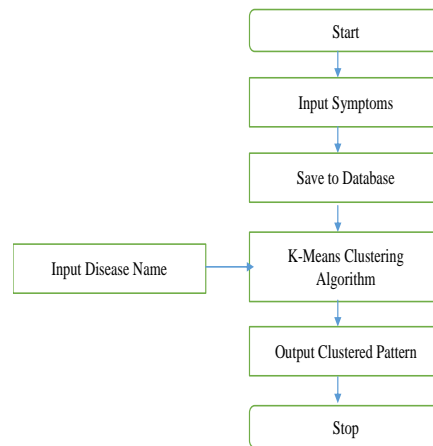Figure3.2 Training Phase for image input

Figure3.3 Training Phase for manual input

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 can be achieved with gabor wavelet filter and can be expressed as

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

Where
pr,pg,pb=Image Pixel components
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
$$0< |P_{Back} - P_{forward}| <= 10 \quad\quad \text{Eq}.......3.2$$
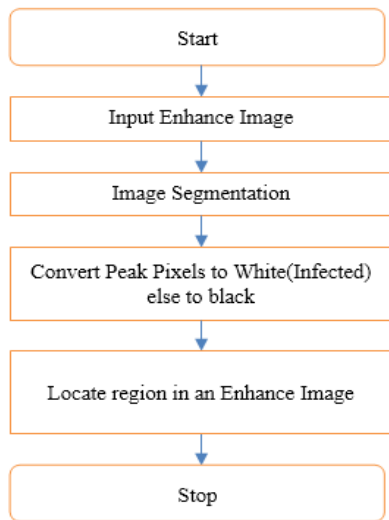Set $P_{forward}$=255(for first hit only). For second, third and onward set P_{Back}=255

Figure 3.4 DFD Image Pre-Processing

An image pre-processing in a proposed method is a important step where an infected pattern is located in an enhance 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

$(0 \leq pg \leq 10) \,||\, (245 \leq pg \leq 255) \&\&(pr||pb == 0)$
Eq……..3.4

$(0 \leq pb \leq 10) \,||\, (245 \leq pb \leq 255) \&\&(pr||pg == 0)$
Eq……..3.5

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 255 means white color.
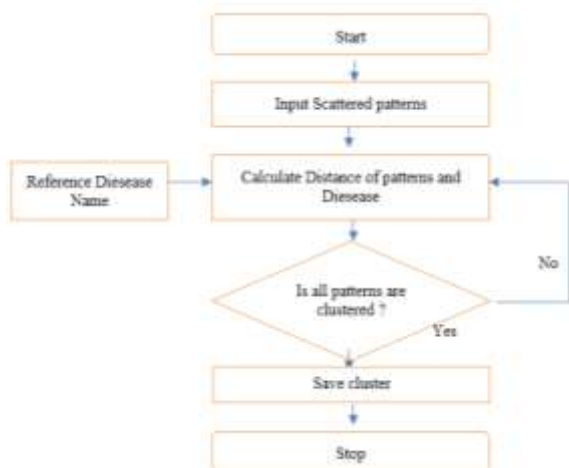

Figure 3.5. DFD k-Means Clustering

k-Means clustering algorithm is used to cluster disease patterns where all patterns are gathered 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

$Ed = |d - p|$ 　　　　　　　　Eq……….3.7

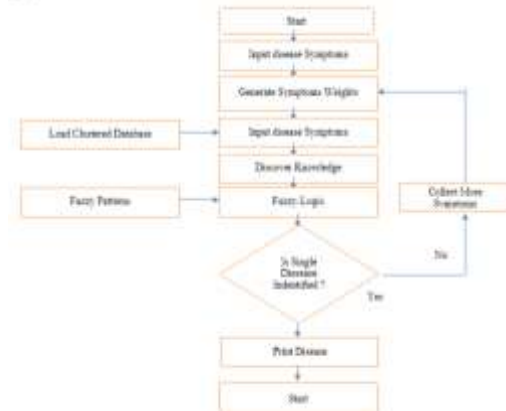A patterns may fall into multiple clusters as per Ecludian distance measures

Figure 3.8 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

III. Fuzzy Patterns rules

Fuzzy patterns are useful in a situation where there is difficult to conclude when multiple patterns are input which may fall into more than one disease, decision will taken by fuzzy rules

Disease Pattern	Malaria	Calera	Cancer	Typhoid	Fuzzy Conclusion
Fever	√	√		√	MDFC
Omitting	√				Malaria
Skin etching			√		Cancer
Head ach		√		√	MDFC
Stomach ach		√			MDFC
Nose running	√			√	MDFC
Cold	√	√			MDFC

Patterns. MDFC means multiple disease fuzzy conclusion. In the previous case, if multiple diseases are found with similar ranking, it becomes difficult to pinpoint to one of them, when no more symptom is unique to any single disease affecting its ranking. In such cases, recent medical historical data stored in the database of the proposed system is used, with a time period of three months to rank the diseases on basis of the probability of their occurrence in the review period. Time frame of three months provided accurate diagnosis with less error. The duration depends mainly on the seasonal cycles in the location where the system is to be implemented. If the interpreted diagnosis is still vague, then the system proceeds to differential diagnosis

For the differential diagnosis, weights are assign to each disease whenever it is entered in the system. According to the weights the decision can be taken out.

The following hardware and software would be used during developments of methods:

- PC with 1GB RAM
- MATLAB software
- C#
- Database MS-ACCESS 2007
- SQL

IV.   IMPLICATION (PROBABLE)

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. 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]   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.

[2]   K. Kawamoto, C. A. Houlihan, E. A. Balas, and D. F. Lobach, "Improving clinical practice using clinical decision support systems: A systematic reviewoftrialstoidentifyfeaturescriticaltosucc ess,"Br.Med.J.,vol.330,p. 765, 2005.

[3]   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.

[4]   W. Siegenthaler, Differential Diagnosis in Internal Medicine: From Symp- tom to Diagnosis. New York: Thieme Medical Publishers, 2011.

[5]   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.

[6]   J. Han and M. Kamber, Data Mining Concepts and Techniques. San Mateo, CA: Morgan Kaufmann, 2011.

[7]   R. Rojas, Neural networks: A Systematic Introduction. Berlin, Germany: Springer-Verlag, 1996, pp. 337–370.

[8]   D. Graupe, H. Kordylewsky, and N. Schneider, Principles of Artificial Neural Networks, 2nd ed. Singapore: World Scientific, 2010.

[9]   M. Bell, SOA Modeling Patterns for Service-Oriented Discovery and Analysis. New York: Wiley, 2010, p. 390.

[10]   (Jun. 10, 2012). WebMD: Better Information, Better Health. [Online]. Available at http://symptoms.webmd.com/ sym ptom checker

[11]   L. Li, L. Jing, and D. Huang, "Protein-protein interaction extraction from biomedical literatures based on modified SVM-KNN," in Nat. Lang. Process. Knowl.Engineer., 2009, pp. 1–7.

[12]   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.

[13] H.R. Warner and O. Bouhaddou, "Innovation review: Iliad—A medical diagnostic support program," Top Health Inf. Manage., vol. 14, no. 4, pp. 51–58, 1994.

[14] Department of Medicine Massachusetts Hospital, Boston, DXplain System. (2011). Available at :http://dxplain.org/dxpdemopp/ dxpdemo- brief_files/frame.htm

[15] E. Coiera, The Guide to Health Informatics, 2nd ed. London, U.K.: Arnold, Oct. 2003, pp. 101–123.

[16] K.A. Mohan. (Jun. 2012). National Electronic Health Record Models.Tata Consultancy Services (TCS) Whitepaper. [Online]. Available at: http://www.tcs.com/ resources/white_papers/Pages/NationalElect ronic HealthRecordModels.aspx

[17] D.S. Rizzuto and M.J. Kahana, "An autoassociative neural network model of paired-associate learning," Neural Comput., vol. 13, pp. 2075– 2092, 2001.

[18] S.M. Polyn and M.J. Kahana, "Memory search and the neural represen- tation of context," Trends Cognitive Sci., vol. 12, pp. 24–30, 2008.

[19] Hubert Kordylewski and Daniel Graupe, "A novel large-memory neural network as an aid in medical diagnosis applications," IEEE Trans. Inf. Technol. Biomed., vol. 5, no. 3, pp. 202–209, Sep. 2001.

[20] T. Kohonen and T. Honkela.(2007). Kohonen network.[Online]. Available: Scholarpedia at: http://www.scholarpedia. org/article/Kohonen_network (Accessed on 4th Feb 2011,7.30 pm IST).

[21] M.H. Valipour, B.A. Zafari, K.N. Maleki, and N. Daneshpour, "A brief survey of software architecture concepts and service-oriented architecture," in Proc. 2nd IEEE Int. Conf. Comput. Sci. Inf. Technol., Aug. 8–11 2009, pp. 34–38.

[22] 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.

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

[24] Wasan, S., Bhatnagar, V., Kaur, H.m "The impact of data mining techniques on medical diagnostics," *Data Science Journal* 5, 119–126, 2006.

[25] 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

[26] P. Herzum, "Web services and service-oriented architectures," Cutter Distributed Enterprise Architecture Advisory Service, Executive Report, 2002.

[27] Jiawei Han, MichelineKamber, Data Mining Concepts and Techniques, 2011 edition, Morgan Kaufmann Publications.

[28] Misdiagnosis: Symptom and heath diagnosis checker. Available at: http://www.misdiagnosis.com.(4th Feb 2011,7.30pm GMT)

[29] S.I.Chowdhury.: Statistical Expert Systems - A Special Application area for Knowledge-based Computer Methodology. Linkoping Studies in Science and Technology, Thesis No 104., Department of Computer and Information Science, University of Linkoping, Sweden.

[30] 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.

[31] 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.

[32] R. Isola, R. Carvalho, A. Tripathy ,"Knowledge Discovery in Medical Systems Using Differential Diagnosis, LAMSTAR, and K-NN," in IEEE transaction on information technology in Biomedicine, Vol no 16,no 6,November 2012, pp.1287-1295.

[33] 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.

[34] 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

[35] 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.

[36] 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.

[37] 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.

[38] 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.

[39] Asmaa S. Hussein, Wail M. Omar, XueLi"Efficient Chronic Disease Diagnosis Prediction and Recommendation System, "IEEE EMBS International Conference on Biomedical Engineering and Sciences I Langkawi I 17th - 19th December 2012 pp.209-214

[40] Joyce C. Ho, "Risk Prediction of a Multiple Sclerosis Diagnosis,"2013 IEEE International Conference on Healthcare Informatics, pp-175-183

[41] AnantMadabhushi, Ajay Basavanhally, Scott Doyle, Shannon Agner, George Lee, "Computer-Aided Prognosis: Predicting Patient And Disease Outcome Via Multi-Modal Image Analysis," IEEE paper 2010,pp1415-1418

[42] Asha.T, Dr. S. Natarajan, Dr. K.N.B. Murth, "Diagnosis of Tuberculosis using Ensemble methods," IEEE 2010,pp.409-412

[43] Yu-Hsiang Wu, Jhu-Yun Huang, Shyi-Chyi Cheng, "Evolutionary Feature Construction for Ultrasound Image Processing and Its Application to Automatic Liver Disease Diagnosis,"2011 International Conference on Complex, Intelligent, and Software Intensive Systems, pp.565-570

[44] K. Kosasih, U.R. Abeyratne, V. Swarnkar, "High Frequency Analysis of Cough Sounds in Pediatric Patients with Respiratory Diseases,"34th Annual International Conference of the IEEE EMBS San Diego, California USA, 28 August - 1 September, 2012,pp.5654-5657

[45] Imran A. Khan, "Personalized Electronic Health Record System for Monitoring Patients with Chronic Disease," Proceedings of the 2013 IEEE Systems and Information Engineering Design Symposium, University of Virginia, Charlottesville, VA, USA, April 26, 2013pp.121-126

[46] UshaaEswaran, VivekEswaran, Sudharshan. V.B, "Human like Biosensor Disease Simulator, Disease Analyzer and Drug delivery system," Proceedings of 2013 IEEE Conference on Information and Communication Technologies (ICT 2013),pp-1033-1038

[47] Anjali Joshi, and VidhyacharanBhaskar, "Multi-resolution Approach for Medical Image Registration using Genetic Algorithm," IEEE International conference on Communication and Signal Processing, April 3-5, 2013, pp-612-616.