Decision Support System for Hepatitis Disease Diagnosis using Bayesian Network

Shamshad Lakho[1], Akhtar Hussain Jalbani[1], Muhammad SaleemVighio[1], Imran Ali Memon[1], Saima Siraj Soomro[1], Qamar-un-Nisa Soomro[2]

Abstract

Medical judgments are tough and challenging as the decisions are often based on deficient and ambiguous information. Moreover, the result of decision process has direct effects on human lives. Act of human decision declines in emergency situations due to complication, time limit and high risks. Therefore, provision of medical diagnosis plays a dynamic role, specifically in preliminary stage when a physician has limited diagnosis experience and identifies the directions to be taken for the treatment process. Computerized Decision Support Systems have brought a revolution in the medical diagnosis. These automatic systems support the diagnosticians in the course of diagnosis. The major role of Decision Support Systems is to support the medical personnel in decision making procedures regarding disease diagnosis and treatment recommendation. The proposed system provides easy support in Hepatitis disease recognition. The system is developed using the Bayesian network model. The physician provides the input to the system in the form of symptoms stated by patient. These signs and symptoms match with the casual relationships present in the knowledge model. The Bayesian network infers conclusion from the knowledge model and calculates the probability of occurrence of Hepatitis B, C and D disorders.

Keywords: Decision support system, Diagnosis, Diagnosticians, Probablistic Model, Knowledge Model

1. Introduction

Diagnosis is the process of recognition the cause of a problem and disease. The process of diagnosis is performed on the basis of information acquired from past and check-up of a patient. Since the advancement of Information Technology, there are significant improvements in the development of computerized medical diagnosis systems. These improvements have led to advances in medical aid. The integration of Information Technology (IT) in health care centers is not only limited in administrative applications like patients' registration, billing, record keeping and payroll, but also it plays an important role in the assistance of the physicians in the diagnosis of different diseases. In this regard, decision/assistance support system has evidenced to be a suitable tool that helps physicians in solving complicated medical issues such as disease diagnosis [1] [2]. Medicine, engineering, business and science etc. are the fields that work on diagnosis [3]. The stipulation of decision making plays a chief role in the field of medical science. A good doctor diagnoses a disease by his practice, knowledge and talent on the basis of symptoms reported by a patient. A recent practice is that patients consult specialists for better diagnosis and treatment. Other general physicians may not have sufficient expertise in controlling some high risk diseases. On the other hand, it is very hard to get an appointment from a specialist; it may take some days, week or months as well. Most likely, the disease may have affected the

---

[1] Department of Information Technology, Quaid-e-Awam University of Engineering, Science & Technology Nawabshah, Pakistan

[2] Department of Examination, Quaid-e-Awam University of Engineering, Science & Technology Nawabshah, Pakistan

Corresponding email: Shamshad.lakho@quest.edu.pk

patient most before the patient refers to the specialist for diagnosis [4]. Most of the high-risk disorders could be treated only in the initial stage. Therefore, the computer-based approaches for the diagnosis of the diseases are important. By using the computer-based disease diagnosis systems, the death ratio and the waiting time could be reduced [5].

Recently, many researchers have used different AI techniques to identify the correct disease [6] - [7]. Bayesian Belief Network is also commonly used AI technique in the field of Biomedical Science for the diagnosis of different high-risk diseases [8].

2. Related work

Decision support systems are communicating and computer based programs that assist users in decision making and judgment [9]. They do not substitute humans but enhance their limited capability to solve complex problems [10]. DSSs are mostly used in military, health care and business areas where complex decision making situations will encounter [11]. The diagnosis system uses Bayesian network in identifying the particular category of Hepatitis disease. A Bayesian belief network is also known as acyclic graphical model. It is a probabilistic model that denotes a set of arbitrary variables and their conditional independencies through a directed acyclic graph [12]. A Bayesian Network could denote the probabilistic connections between disorders and symptoms. The Bayesian Network could calculate the likelihoods of the existence of the numerous diseases on the basis of the signs or symptoms stated by the patient. A Bayesian Network is a bidirectional diagram that permits the stream of information from causes to effects and vice versa. A Bayesian Network can manage with partial and ambiguous data. Artificial intelligence is the art of converting human intelligence in machines and deals with enlarging such intellectual machines which could support and assist humans in problem solving and decision making domains. The manipulation of AI in medical applications has been increasing since the last decades. The application area of AI may be in educational systems, diagnostic systems, expert systems and as well as in machine learning systems. Thus there are significant developments in the field of computer-based diagnosis. AI machine learning techniques play a vital role in the development and support of computer based disease diagnosis systems as diagnosis requires reliable accuracy and performance [13]. Velikova et al [5], have proposed a Bayesian network approach for constructing probabilistic disease progress models built on clinical principles and developed a system for remote pregnancy care. Alonzo et al [7], have proposed a diagnosis system using Bayesian Network to diagnose ENT (Ear, Nose, Throat) diseases based on the specified signs and symptoms. The author claimed that the system is capable of diagnosing a patient with ENT disease and also able to detect a person if he/she is not infected of ENT disorder. Mahesh et al [8], worked on Hepatitis B disease and built a system for the Hepatitis B diagnosis which comprises the neural network. The system is able to classify the patient infected with immune category. Ratnapida and druzdzel [14], have developed an expert system named as Marilyn that is capable of diagnosing and identifying different problems (like computer, heart disease, breast cancer and lymphographic) based on previous cases cured. They constructed a diagnostic model that performs diagnosis based on communication between diagnostician and diagnostic system. Yan et al [15], designed a decision support system for the diagnosis of heart disease. The decision support system is based on multilayer perceptron. They claimed on the basis of assessment that the diagnosis accuracy of the system is greater than 90%. A probabilistic model named as Hepar is developed for the hepatic disease diagnosis. The proposed system is founded on expert knowledge and patients data. The authors claimed that the diagnostic accuracy of the proposed model reaches nearly 80% [13]. Hussain et al [16], proposed a three-layered Bayesian Network model for analyzing and diagnosing the mental states of students from their gestures.

3. Proposed system

Shamshad Lakho (et al.) Decision Support System for Hepatitis Disease Diagnosis Using Bayesian Network   (pp. 11-19)

(pp.4-10)

The Figure 1 shows the design framework of the proposed system that comprises the following components:
1) Graphical User Interface
2) Medical Database
3) Knowledge Model
4) Bayesian Network
5) Ranking
6) Disease Recognition
   - HBV (Hepatitis B Virus)
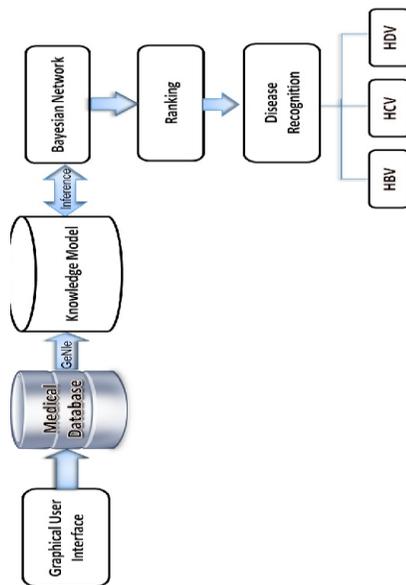   - HCV (Hepatitis C Virus)
   - HDV (Hepatitis D Virus)

Figure 1: Design Framework

The physician selects the symptoms described by the patient through the graphical user interface. The submitted query is stored in the Medical Database that contains patients' cases and disease information. The query is then compared with the casual relationships present in the knowledge model. The knowledge model is created by learning the network.

The Bayesian network applied inference on knowledge model to calculate the posterior probability distribution for a query provided by the user. The Bayesian network has

been used for classification and making decisions to recognize the particular category (i.e. B, C or D) of Hepatitis disease in our case. The categories of hepatitis are classified based on the probability distribution. Finally, the particular category of hepatitis is diagnosed i.e. either the patient is infected of hepatitis B, C or D virus. We use GeNIe/SMILE[3] software package for training and testing the Bayesian network.

3.1. Diagnostic interface

Figure 2 shows the screenshot of the user interface that is used to interact with the Hepatitis Diagnostic system.
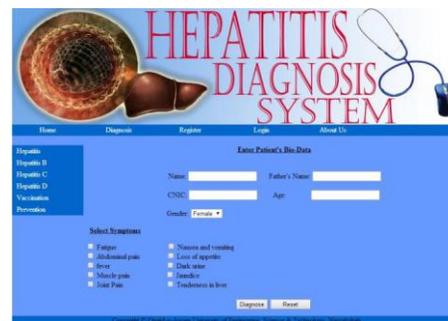

Figure 2: User Interface to interact with the Diagnostic System

Figure 3 shows a snapshot of the backend interface of the system. The right hand side of the window comprises a list of all possible signs and symptoms present in the Bayesian network model. The upper right part of the window holds a list of those possible observations that have not yet been instantiated by the user. Those symptoms that have been observed are taken over to the lowest part of the window.
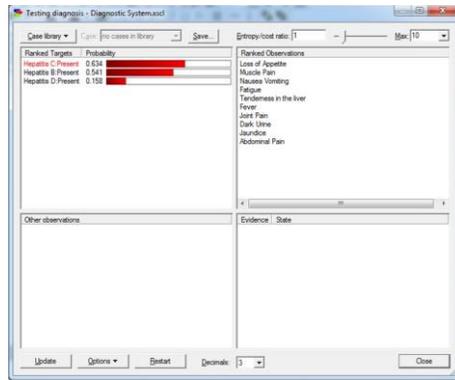
---

[3]available at http://genie.sis.pitt.edu

Figure 3: Diagnostic Interface

3.2. Data description

The data was collected time by time from different online and offline resources like mayoclinic.com and Peoples Medical College & Hospital Nawabshah for the training purpose and interviews were also conducted with hepatologists. The collected data contained one year record of Hepatitis patients consisting of 275 cases. It is difficult to collect medical records from hospital due to their privacy however despite of that difficulty, we tried our best to collect 275 records of hepatitis patients from well-known hospital and train our system to drive the results. The system´s decision making capability is based on the physical exam of the patient. The patient´s cases belonged to Hepatitis B, C & D categories. We have used 200 records for training of the Bayesian Network and 75 records for testing of the system.The statistics of trained data set is that among 200 cases, 79 cases are of HBV patients, 95 cases are of HCV patients and 26 cases are of HDV patients as shown in Figure 4.
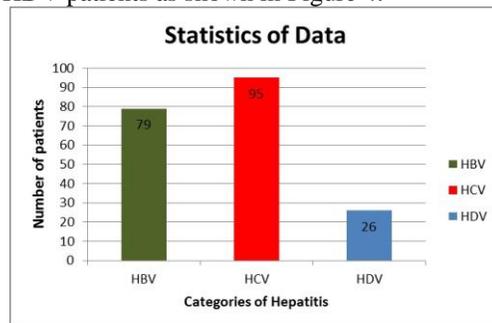

Figure 4: Statistic of trained data set

3.3. Two layered Bayesian network model

Figure 5 shows the Bayesian network that reveals the arrangement of problem domain. The network models 13 variables to diagnose hepatitis disease including 10 symptoms and 3 disorder nodes. The arcs depict the direct probabilistic associations between couple of nodes, so the arc between Jaundice and Hepatitis B node denotes the fact that the existence of Jaundice increases the chance of Hepatitis B disorder. The arrangement of the model is the illustration of the relationships between components of diagnosis process i.e. cause and effect. We have created the structure based on medical fiction and discussions with hepatologists and the numerical factors of the model i.e. prior and conditional probability distributions are mined from a medical database containing cases of Hepatitis patients.
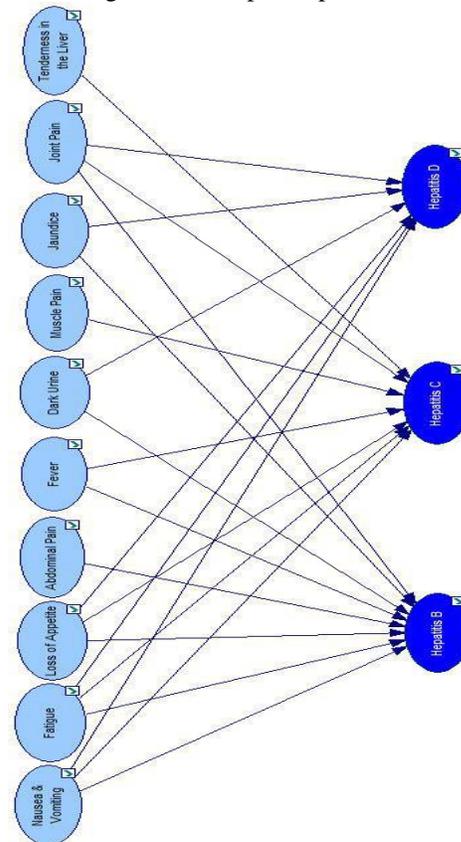

Figure 5: Two Layered Bayesian Network Model

4. Results & Discussion

Calisir and Dogantekin [17], proposed PCA-LSSVM classifier for recognition of hepatitis disorder. The authors claimed that the proposed technique produced valuable results with that the proposed technique produced valuable results with high accuracy rate of 96.12%. Kaya and Uyar [18], proposed RS and ELM based decision support system for identifying the hepatitis disease. The authors claimed that the RS-ELM method is better than other classification methods. Bascil and Temurtas [19], proposed hepatitis disease diagnostic network and LevenbergMarquardt training algorithm and compared the derived results with the other studies focusing on the similar UCI database. The medical diagnosis systems are gaining popularity in disease diagnosis and providing timely medical aid. The diagnostic accuracy of such systems is rather promising and reliable. This research study also shows that the developments of decision-support systems of high-risk diseases using AI techniques proved beneficial in providing inexpensive and timely diagnosis. Chen et al. [20], offered the LFDASVM method for the hepatitis disease diagnosis. The proposed method examined the condition of the liver and distinguished the live liver from the dead one. Other researchers worked on image based features to diagnose the hepatitis disease while our proposed system works on the symptoms stated by patient. However if both image features and physical features are combined together the system will provide good results that help physicians to take appropriate decisions.

4.1. Diagnosis result of trained data set

The Figure 6 shows the diagnosis result of the proposed system performed on trained data set. Selecting the possible values of the symptoms, the associated probabilities of the disorders are updated spontaneously and a new well-organized list of the disorders is presented. In the following result Hepatitis B is diagnosed with the probability value of approximately 95%.
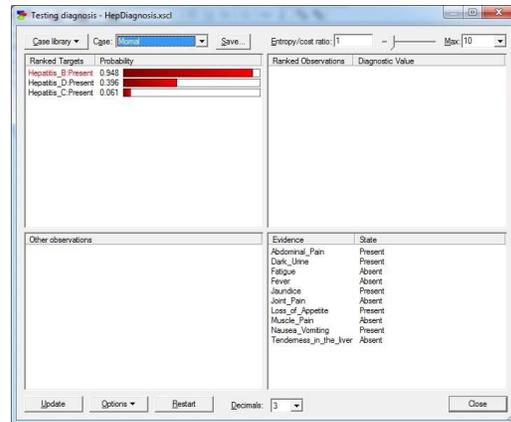

Figure 6: Diagnosis of Hepatitis B Disease

4.2 Diagnosis rate of trained data set

The diagnosis rate of the trained data is calculated with the help of hit and trial method. Twenty five trained cases of each category of Hepatitis are tested on the system respectively. Out of 25 cases of Hepatitis B, system diagnosed only 21 cases correctly, remaining 4 cases are diagnosed incorrectly, 2 cases are identified as HCV and 2 as HDV. Similarly out of 25 cases of HCV, system diagnosed only 23 cases accurately and the remaining 2 cases were diagnosed inaccurately as HBV cases. Likewise, out of 25 cases of HDV, system recognized only 18 cases properly and the remaining 7 cases were identified improperly. Five cases are diagnosed as HBV and 2 cases are diagnosed as HCV.

Table-I. Diagnosis Rate of Trained Data Set

	HBV	HCV	HDV	Producer Accuracy
HBV	21	2	2	84%
HCV	2	23	0	92%
HDV	5	2	18	72%
Overall Accuracy				82.667%

Therefore, on trained data set, the success rate of HBV diagnosis is 84%, the success rate of HCV diagnosis is 92% and the success rate of HDV diagnosis is 72% as shown in the Figure 7.
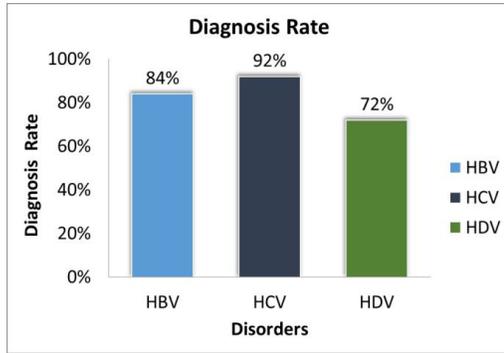
Figure 7. Diagnosis Rate of Trained Data Set.

4.3. Diagnosis result of test data set

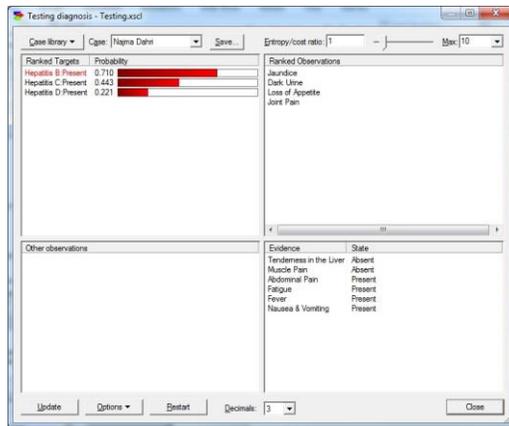Figure 8 shows that in this particular case Hepatitis B is diagnosed with the probability value of 71%.

Figure 8. Diagnosis of Hepatitis B Disease.

4.4. Diagnosis rate of test data set

The diagnosis rate of the test data set is also calculated with the help of hit and trial method. Twenty five new cases of each category of Hepatitis are tested on the system respectively. Out of 25 cases of Hepatitis B, system diagnosed only 20 cases correctly, remaining 5 cases are diagnosed incorrectly, 2 cases are identified as HCV and 3 as HDV. Similarly, out of 25 cases of HCV, system diagnosed only 22 cases accurately and the remaining 3 cases were diagnosed inaccurately, 2 cases as HCV cases and 1 case as HDV. Likewise, out of 25 cases of HDV, system recognized only 13 cases properly and the

remaining 12 cases were identified improperly. Seven cases are diagnosed as HBV and 5 cases are diagnosed as HCV.

Table II. Diagnosis Rate of Test Data

	HBV	**HCV**	**HDV**	**Producer Accuracy**
HBV	20	2	3	80%
HCV	2	22	1	88%
HDV	7	5	13	52%
Overall Accuracy		73.33%		

Therefore, on test data set, the success rate of HBV diagnosis is 80%, the success rate of HCV diagnosis is 88% and the success rate of HDV diagnosis is 52% as shown in the Figure 9.
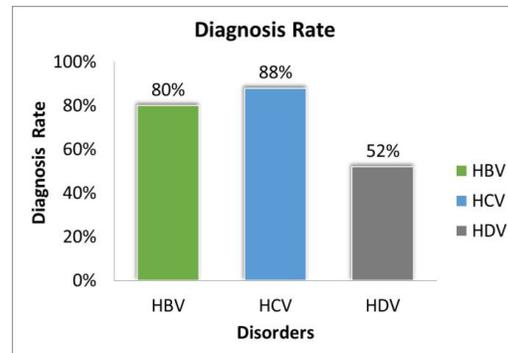

Figure 9. Diagnosis Rate of Test Data Set.

4.5. Accuracy of trained and test data sets

Consequently the overall diagnosis accuracy of trained data is 83% and test data is 73% as shown in the figure 10, is calculated on the system.
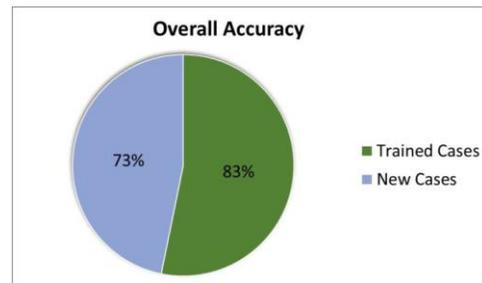

Figure 10. Overall Accuracy of the System.

5. Conclusion &   Future work

Medical diagnosis is one of the challenging tasks. Computerized decision support systems are becoming familiar and useful in assisting the physicians in the diagnosis process. The proposed system is the first step towards the diagnosis of hepatitis disorder; it's not the full fledge system with complete functionality. Yet the system is especially designed for the emergency medical care in ruler areas where there is no any facility of laboratory testing and other necessary examinations for the diagnosis of hepatitis due to that death ratio is very high in those areas. The main objective is to facilitate the ruler areas people as a first aid to save their lives in emergency situations. The proposed model is the prototype of the Hepatitis disease recognition system which may be helpful in the future for practical implementation in real situations in different areas. The model is trained by using the information of prior probabilities, extracted from the collected data of different patients. Bayesian Network Model has become an effective method for avoiding the over fitting of data and collected samples can be "smoothed" so that all available data can be used properly without preprocessing of the information. Bayesian Network model provides the rich facility to handle the uncertain situation. Therefore, we proposed to use it for the recognition of Hepatitis disease; however many researchers proposed the different machine learning techniques to classify the Hepatitis, but in our best knowledge there is no any work has been proposed yet to use Bayesian Network for Hepatitis disease recognition. In the future, the system can be enhanced with different modalities for the recognition of the disease. In this research a diagnosis system has been developed using the decision theoretic approaches to diagnose the Hepatitis B, C & D diseases and tested on a data set of new cases. The system works well for calculating the influence of various symptoms on the probability of different categories of Hepatitis and assists in the process of diagnosis and classification of various categories of Hepatitis. The system could be used as a training tool for the practice of trainee physicians as well. The main application of the system is in helping, serving and supporting in the diagnosis of Hepatitis. System is capable of maintaining a depository of patient cases. The diagnosis accuracy of the system is reasonable. The diagnosis accuracy of those disorders that are greater in number in the medical database of cases ranges almost more than 80%. The overall accuracy of the system seems to be much better and satisfactory than that of trainee diagnosticians. The diagnosis support systems would improve the diagnosis power of the physicians.

In the future, system would be enhanced:

- To increase the number of patients and incorporate more laboratory test results and risk factors in the domain model to improve the accuracy of the system.

- To diagnose all the categories of Hepatitis i.e. A to E.

- To diagnose different Hepatic disorders i.e. PBC (Primary Biliary Cirrhosis), Liver Cancer and Alcoholic Liver disease etc.

REFERENCES

[1]   S. Andreassen, D. Karbing, U. Pielmeier, S.Rees, A. Zalounina, L. Sanden, M. Paul, and L. Leibovici, "Model-based medical decision support–a road to improved diagnosis and treatment?" in 15th Nordic-Baltic Conference on Biomedical Engineering and Medical Physics (NBC 2011). Springer, 2011, pp. 257–260.

[2]   E. S. Berner, Clinical Decision Support Systems. Springer, 2007.

[3]   G. S. Uttreshwar and A. Ghatol, "Hepatitis b diagnosis using logical inference and generalized regression neural networks," in Advance Computing Conference, 2009. IACC 2009.IEEE International. IEEE, 2009, pp. 1587–1595.

[4]   M. Neshat, M. Sargolzaei, A. NadjaranToosi, and A. Masoumi, "Hepatitis disease diagnosis using hybrid case based reasoning and particle swarm optimization," ISRN Artificial Intelligence, vol. 2012, 2012.

[5]   A. B. Mrad, V. Delcroix, S. Piechowiak, and P. Leicester, "From information to evidence in a bayesian network," in Probabilistic Graphical Models. Springer, 2014, pp. 33–48.

[6]   A. Oni´sko and M. J. Druzdzel, "Impact of bayesian network model structure on the accuracy of medical diagnostic systems," in Artificial Intelligence and Soft Computing. Springer, 2014, pp. 167–178.

[7]   A. L. D. Alonzo, J. J. M. Campos, L. L. M. Layco, C. A. Maratas, and R. A. Sagum, "Entdex: Ent diagnosis expert system using bayesian networks," Journal of Advances in Computer Networks, vol. 2, no. 3, 2014.

[8]   C. Mahesh, V. Suresh, and M. Babu, "Diagnosing hepatitis b using artificial neural network based expert system," infection, vol. 3, no. 6, 2013.

[9]   N. M. Sharef and H. Madzin, "Ims: An improved medical retrieval model via medical-context aware query expansion and comprehensive ranking," in Information Retrieval & Knowledge Management (CAMP), 2012 International Conference on. IEEE, 2012, pp. 214–218.

[10]   D. Dinh and L. Tamine, "Towards a context sensitive approach to searching information based on domain specific knowledge sources," Web Semantics: Science, Services and Agents on the World Wide Web, vol. 12, pp. 41–52, 2012.

[11]   I. Niˇzeti´c, K. Fertalj, and B. Milaˇsinovi´c, "An overview of decision support system concepts," in Proceedings of the 18th International Conference on Information and Intelligent Systems/Boris Aurer and MiroslavBaˇca (ur.). Varaˇzdin, 2007, pp. 251–256.

[12]   A. Oni´sko and M. J. Druzdzel, "Impact of precision of Bayesian network parameters on accuracy of medical diagnostic systems," Artificial intelligence in medicine, vol. 57, no. 3, pp. 197–206, 2013.

[13]   H. Wasyluk, A. Onisko, and M. Druzdzel, "Support of diagnosis of liver disorders based on a causal bayesian network model," Medical Science Monitor, vol. 7, no. 1; SUPP, pp. 327–332, 2001.

[14]   P. Ratnapinda and M. J. Druzdzel, "Does query-based diagnostics work?" 2011.

[15]   H. Yan, Y. Jiang, J. Zheng, C. Peng, and Q. Li, "A multilayer perceptronbased medical decision support system for heart disease diagnosis," Expert Systems with Applications, vol. 30, no. 2, pp. 272–281, 2006.

[16]   A. Hussain, A. R. Abbasi, and N. Afzulpurkar, "Detecting & interpreting self-manipulating hand movements for students affect prediction," Human-Centric Computing and Information Sciences, vol. 2, no. 1, pp.1–18, 2012.

[17]   D. C¸ alis¸ir and E. Dogantekin, "A new intelligent hepatitis diagnosis system: Pca–lssvm," Expert Systems with Applications, vol. 38, no. 8, pp. 10 705–10 708, 2011.

[18]   Y. Kaya and M. Uyar, "A hybrid decision support system based on rough set and extreme learning machine for diagnosis of hepatitis disease," Applied Soft Computing, vol. 13, no. 8, pp. 3429–3438, 2013.

[19]   M. S. Bascil and F. Temurtas, "A study on hepatitis disease diagnosis using multilayer neural network with levenbergmarquardt training algorithm," Journal of Medical Systems, vol. 35, no. 3, pp. 433–436, 2011.

[20]   H.-L. Chen, D.-Y. Liu, B. Yang, J. Liu, and G. Wang, "A new hybrid method based on local fisher discriminant analysis and support vector machines for hepatitis disease diagnosis," Expert Systems with Applications, vol. 30, no. 2, pp. 272–281, 2006.