Volume 3, Issue 6, June 2013 ISSN: 2277 128X
International Journal of Advanced Research in Computer Science and Software Engineering
Research Paper
Available online at: www.ijarcsse.com

Pulmonary Disease Diagnosis Using Bayesian Belief Networks: a Conceptual Framework

Nidhi, AP Gurpinder Kaur
Department of Computer Science,
Swami ViVekanand Institute of Engineering and Technology,
Punjab Technical University, Jalandhar, India

***Abstract:** Considerable population of the world does not have access to even bare minimum healthcare facilities. Majority of these are not able to avail these facilities because of lack of resources/funds, resultantly they suffer. Irony however is that most of the diseases are curable provided diagnosis and treatment is initiated at right time. Information Technology in general and predictive models in particular is being used to facilitate the physicians in diagnosis of diseases. Researchers have used Bayesian Belief Networks in medicine for (i) Diagnosis of various diseases like ovarian cancer [16], breast cancer [17] (ii) analysis of f-MRI [15] (iii) analysis of mammography [19] and (iv) Study of mutations. In the present study attempt is made to develop a conceptual framework for pulmonary disease diagnosis.*

***Keywords:** Bayesian Networks, Pulmonary Disease, Diagnosis, Predictive model*

I. INTRODUCTION

Incorrect diagnosis may have severe implications. The current annual per capita public expenditure on health in India is no more than Rupees Two Hundred, which is among the five lowest in the world. As per National Health Accounts [2], the total health expenditure in India from all the sources was Rupees 1, 33,776 crores, consisting of 4.25 percent of the GDP. Of the total health expenditure, the share of private sector was highest at 78 percent (private individual households 71 percent and private firms 7 percent); public sector at 20 percent (Central Government 7 percent, State Government 12 percent and Local Government 1 percent) and external flows contributed about 2 percent.

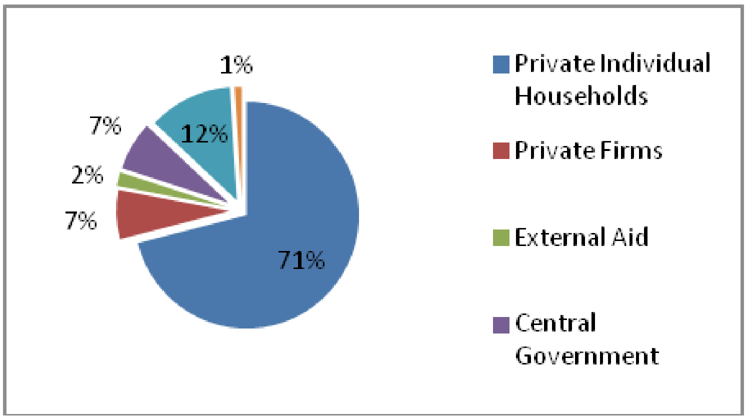

Figure 1 (Health Expenditure India – 2004)
Source: National health accounts 2004-2005

It has been estimated that less than 20 percent of the population, which seek OPD and less than 45 percent of that which seek indoor treatment avail of such services in public hospitals. These is despite the fact that most of these patients do not have the means to make out – of – pocket payment for private health services except at the cost of other essential expenditure for items such as basic nutrition. Stated plainly this means most of the people in India do not have the capacity to pay for the expensive tests. As such, it is very important for the doctors to make a correct diagnosis and if possible with minimum reliance on laboratory tests and more on their expertise. The current review study is targeted to understand how the doctors can be helped making correct computer aided diagnosis.

1.1 Artificial neural networks A neural network consists of an interconnected group of artificial neurons, and it processes information using a connectionist approach to computation. In most cases a neural network is an adaptive system that changes its structure during a learning phase.

1.2 Bayesian Network

A Bayesian network, Bayes network, belief network, Bayes(ian) model or probabilistic directed acyclic graphical model is a probabilistic graphical model (a type of statistical model) that represents a set of random variables and their conditional dependencies via a directed acyclic graph (DAG). For example, a Bayesian network could represent the probabilistic relationships between diseases and symptoms. Given symptoms, the network can be used to compute the probabilities of the presence of various diseases.

II. Related Work

Extensive research has been carried out on diagnosis of different types of diseases using Bayesian belief networks. Bayesian networks can be used to generate posterior probabilities using the conditional probability of particular events occurring. In relation to medical expert systems this means they can state the probability that a patient has a disease based upon personal information, existing symptoms, and the results from tests. The ability to generate causal relationships between these types of information is desirable because it allows doctors to understand how the diagnosis was made and promotes trust in the system. Other approaches used to build medical expert systems, such as ANN have a notable disadvantage due to their lack of explain ability. Madsen. A. [4] carried out a study on the usage of Bayesian network in breast cancer diagnosis, ovarian cancer diagnosis, and tongue diagnosis. Results show that Bayesian networks are a viable approach in these fields. Nikovski D [5] has discussed several knowledge engineering techniques for the construction of Bayesian networks for medical diagnostics when the available numerical probabilistic information is incomplete or partially correct. This may happen when epidemiological studies publish only indirect statistics and when significant un-modeled conditional dependence exists in the problem domain. Nikovski suggested that while nothing can replace precise and complete probabilistic information, still a useful diagnostic system can be built with imperfect data by introducing domain-dependent constraints and proposed a solution to the problem of determining the combined influences of several diseases on a single test result from specificity and sensitivity data for individual diseases. Two techniques for dealing with unmodeled conditional dependencies in a diagnostic network were also demonstrated in the context of an effort to design a portable device for cardiac diagnosis and monitoring from multimodal signals. Xiang *et al* [6] developed a prototype neuromuscular diagnostic system (PAINULIM) that diagnoses painful or impaired upper limbs based on Bayesian networks. They discuss non-mathematically the major knowledge representation issues that arose in the development of PAINULIM. Motivated by the computational overhead of large application domains, and the desire to provide a user with an interface that gives a focused display of a sub-domain of current interest, they built PAINULIM using the idea of multiply sectioned Bayesian networks. A preliminary evaluation of PAINULIM with 76 patients demonstrated good clinical performance. Haddawy *et al* [7] constructed a Bayesian belief network in the domain of hepatobiliary disease. The network model's nodes represented diagnoses, physical findings, laboratory test results, and imaging study findings. The connections between nodes incorporated conditional probabilities, such as sensitivity and specificity, to represent probabilistic influences. They extracted Statistical data from peer-reviewed journal articles on hepatobiliary disease, and created a network to reflect the data. The network successfully determined the a priori probabilities of various diseases, and incorporated laboratory and imaging results to calculate the a posteriori probabilities. The most informative examination was identified, that is, the laboratory study or imaging procedure that led to the greatest diagnostic certainty. The study concluded that Bayesian networks represent a very promising technique for decision support in radiology: they can assist physicians in formulating diagnoses and in selecting imaging procedures. To capture, conserve and disseminate such valuable expert knowledge remains a key challenge to the application of knowledge-based systems in veterinary medicine. McKendrick *et al* [8] explore the use of a Bayesian belief network to quantify expert opinion with a view to estimating the likelihood of various diseases in the presence and absence of certain signs. They elicited information from a panel of 44 experienced veterinarians to provide the response matrix of 27 signs associated with 20 commonly occurring diseases in sub-Saharan cattle. Bayesian belief networks besides being used in disease diagnosis have also been used in trauma cases. Ogunyemi. O [9] describes a method for diagnostic reasoning under uncertainty that is used in Trauma SCAN, a computer-based system for assessing penetrating trauma. Uncertainty in assessing penetrating injuries arises from two different sources: the actual extent of damage associated with a particular injury mechanism may not be easily discernable, and there may be incomplete information about patient findings (signs, symptoms and test results) which provide clues about the extent of the injury. Bayesian networks are used in Trauma SCAN for diagnostic reasoning because they provide a mathematically sound means of making probabilistic inferences about the injury in the face of uncertainty. Gadewadikar *et al* [10] explored the implementation of a Bayesian Belief Network for an automated breast cancer detection support tool. They were of the opinion that it is intuitive that Bayesian networks are employed as one viable option for computer-aided detection by representing the relationships between diagnoses, physical findings, laboratory test results and imaging study findings. Their work brings important entities such as Radiologists, Image Processing Scientists, Data Base Specialists and Applied Mathematicians on a common platform. They concluded that encoding of independencies in the network topology admits the design of efficient procedures for performing computations over the network. Further, for the application of computer-aided detection in mammography, the researchers designed an interface between the project's Bayesian network learning algorithm and the radiologists, so that the radiologists could have interaction with the system by labeling only a small number of informative images presented by the active learning algorithm. In a study conducted by Lu *et al* [11] used Bayesian Belief Network for predicting Coronary Artery Disease (CAD) in Traditional Chinese Medicine. They used Bayesian Network to construct a high-confidence syndrome predictor based on the optimum subset, which was collected by Support Vector Machine (SVM) feature selection and compared this subset with Markov blanket feature select using ROC. They also designed Naıve Bayes, C4.5 Logistic, and Radial basis function (RBF) network compared with Bayesian network. It was found that Bayesian network method based on the optimum symptoms shows a practical method to predict six syndromes of CAD in Traditional Chinese Medicine (TCM). Bayesian networks have been used in other medical areas as well as like emotion recognition [12], ventilator associated pneumonia (13), comparing risk of alternative medical diagnosis [14], analysis of f-MRI \[15], ovarian cancer [16], breast lesion classification [17], breast cancer diagnosis [18], analysis of mammography [19], cancer diagnosis on the basis of bio-markers [20] and study of mutations [21].

III. Conclusion

Although numerous studies have been carried out for the diagnosis of diseases and one even trauma using Bayesian belief network, very limited work has been carried out on diseases related to lungs. Some work has been reported on diagnosis on lung cancer using Bayesian belief network, literature survey however indicates no comprehensive work on pulmonary diseases as such. Present study aims to use Bayesian belief networks for diagnosis of various pulmonary diseases.

Pulmonary Disease Diagnosis – A Conceptual Framework

3.1DataCollection: For this study initial data bank of symptoms will be collected on the basis of a survey questionnaire to be administered to doctors seeking their opinion and rating of symptoms based upon the strength to correctly predict pulmonary disease.

An initial data bank will thus be developed for further classification and study.

This will be followed by prioritizing the symptoms for accurate disease prediction and diagnosis.

3.2 Symptom Selection Phase **-** Symptom selection was regarded as the problem of feature selection. Symptoms are essential to diagnose a disease. Therefore, a strong predicting model of syndrome is based on key symptoms. In this phase, we investigated which symptoms influence the predicted syndromes most. We propose SVM feature selection methods to discover critical symptoms.

Support vector machines (SVMs) have been a promising tool for data classification. Its basic idea is to map data into a high

dimensional space and find a separating hyperplane with the maximal margin. Given training vectors $x_k \in R^n$, k = 1, m in two classes, and a vector of labels $y \in R^m$ such that YK ∈ {1, −1}, SVM solves a quadratic optimization problem [22, 23]

Subject to $YK\ (\omega^T \varphi\ (xk) + b) \geq 1 - \xi k\ \xi k \geq 0,\ k = 1, \ldots, m,$

$$\boxed{\min_{\omega, b, \varepsilon} \ \frac{1}{2}\, \omega^t \ + C \sum_{k=1}^{m} \xi\, t}$$

Eqn. 1

Where training data are mapped to a higher dimensional space by the function Φ and C is a penalty parameter on the training error. For any testing instance x, the decision function (predictor) is

$f(x) = \mathrm{sgn}(\omega^t \ \varphi(x) + b)$

Practically, we need only $k(x, x') = \varnothing(x)^T \ \varnothing(x')$ the kernel function, to train the SVM. The RBF kernel is used in our experiments:

$k(x, x') = \exp(-\gamma \| x - x' \|^2)$

Several existing strategies have been combined with SVM for feature selection. Given training vectors x_k, k =1, 2, . . . ,m, if the positive and negative instances are $n+$ and $n-$, respectively, then the F-score of the ith feature is defined as:

Formula no. 3 to be used from base paper

We selected features with high *F*-scores and then applied SVM for training/prediction. The procedure was as follows [24]. (1) Calculate *F*-score of every feature.

(2) Pick possible thresholds as cutoffs for *F*-scores. (3) For each threshold, complete the following:

(a) Drop features with *F*-scores below this threshold,

(b) Randomly split the training data into Xtrain and Xvalid,

(c) Let Xtrain be the new training data. Use the SVM procedure to obtain a predictor; use the predictor to predict Xvalid,

(d) Repeat the steps above five times and then calculate the average validation error.

(4) Choose the threshold with the lowest average validation error.

(5) Drop features with *F*-scores below the selected threshold. Then apply the SVM procedure. Finally, the features with efficient prediction power were selected.

Disease Prediction. Disease prediction is important for doctors. In this study we will be developing a Bayesian network framework to construct a high-confidence disease predictor by integrating a comprehensive list of mingling Symptoms. Bayesian network, which is one of the most effective classification methods for graphically representing and processing feature interdependencies, represents a joint probability distribution over a dataset [25, 26]. Bayesian network is directed acyclic graphs (DAG) that allow for efficient and effective representation of joint probability distributions. In this paper, we will be constructing a Bayesian network structure to simulate the data model based information obtained through research survey. Based upon the results obtained a suitable model will be developed to predict the prevalence of pulmonary disease.

IV. Future Scope

Bayesian networks have been extensively used in developing computer aided diagnosis of various diseases. The purpose of using these computer based systems in healthcare is to reduce the chances of faulty diagnosis and prevent cost and time over runs. In future study the use of Bayesian belief networks for the diagnosis of pulmonary diseases can be explored.

Acknowledgement

I am thankful to AP GURPINDER KAUR, ASST.PROF, SWAMI VIVEKANAND INSTITUTE OF ENGG. TECHNOLOGY, BANUR, for providing constent guidance and encouragement for this research work.

References

[1] Cooper, C.L., Sloan, S.L. and Williams, S.L.(1988),*Occupational Stress Indicator Management Guide*, NFER- NELSON, Windsor

[2] http://planningcommission.nic.in/reports/genrep/health/National_Health_Account_04_05.pdf. Assessed on January 22, 2013

[3] R Montironi, W F Whimster, Y Collan, P W Hamilton, D Thomson, P H Bartels (1996). How to develop and use Bayesian Belief Network. Downloaded from jcp.bmj.com on January 16, 2013.

[4] http://www.cs.rit.edu/~rlaz/ai2010/research_summaries/bayesian_diagnosis.pdf. Assessed on January 16, 2013

[5] Nikovski D (2000). *Constructing Bayesian networks for medical diagnosis from incomplete and partially correct data*. IEEE transaction on knowledge and data engineering. Vol. 12 (4), 509-516.

[6] Ynag Xiang, B Pant, A. Eisen, M P Beddoes, D Poole (1993). *Multiply sectioned Bayesian networks for neuromuscular diagnosis*. Artificial Intelligence in medicine. Vol. 5 (4), 293-314.

[7] P Haddawy, C E Kahn Jr., M Butarbutar (1994). *A Bayesian network for radiological diagnosis and procedure selection: work-up of suspected gallbladder disease*. Medical Physics. Vol. 21 (7), 1185 – 1192.

[8] http://www.ncbi.nlm.nih.gov/pubmed/11058776. Assessed on January 18, 2013.

[9] Ogunyemi O, Clarke J R, Webber B (2000). Using Bayesian networks for diagnostic reasoning in penetrating injury assessment. Computer based medical systems, 2000. Proceedings, 13th IEEE symposium.

[10] Gadewadikar J, Kuljaca O, Agyepong K, Sarigul E, Zheng Y, Zhang P (2010). *Exploring Bayesian networks for medical decision support in breast cancer detection*. African Journal of Mathematics and Computer Science Research. Vol. 3 (10), 225 – 231.

[11] P .Lu, J.Chen, H.Zhao, Y. Gao, L. Luo, X. Zuo, Q.Shi, Y.Yang, J.Yi and W.Wang. "*In Silico* Syndrome Prediction for Coronary Artery Disease in Traditional Chinese Medicine". Evidence Based Complementary and Alternate Medicine. Vol. 2012 (2012), article id 142584.

[12] Ke Ko, H C Yang, K B Sim. "*Emotion Recognition Using EEG Signals With Relative Power Values and Bayesian Networks*". International Journal of Control Automation and Systems 2009, vol. 7(5), page 865-870

[13] T.Charitos, L.C.V.Gaag, S Visscher. "*A Dynamic Bayesian Network for Diagnosing Ventilator Associated Pneumonia in ICU Patients*". Expert Systems with Applications 2009, vol. 36 (2), page 1249-1258.

[14] N.Fenton, M.Neil. "*Comparing Risks of Alternative Medicine Diagnosis using Bayesian Arguments*". Journal of Biomedical Informatics 2010, vol. 43 (4), page 485-495.

[15] J. Burge, T.Lane, H. Link, S.Qiu, V.P.Clark. "*Discrete Dynamic Bayesian Network Analysis of f-MRI Data*". Human Brain Mapping 2009, vol. 30(1), page 122-137.

[16] P.Antal, H.Verrelst, D.Timmerman. "*Bayesian Networks in Ovarian Cancer Diagnosis: Potentials and Limitations*". 13th IEEE Symposium on Computer Based Medical Systems 2000, page 103-108.

[17] E.A.Fischer, J.Y. Lo, M.K.Markey. "*Bayesian Networks of BI-RADS Descriptors for Breast Lesion Classification*". 26th Annual International Conference of IEEE on Engineering in Medical and Biology Society 2004, page 3031-3034.

[18] E.S.Burnside, Daniel, L.Rubin, R.D.Shacter. *"Using a Bayesian Network to Predict the Probability and Type of Breast Cancer"*. Med info. 2004, Amsterdam, IOS Press.

[19] E.S.Burnside, Daniel, L.Rubin, R.D.Shacter. "*A Bayesian Network for Mammography*". AMIA Annual Symposium Proceedings Archive 2000, page 106- 110.

[20] X.Deng, H.Geng, H.H.Ali. "*Cross Platform Analysis of Cancer Bio-markers: A Bayesian Network Approach to Incorporating Mass Spectrometry and Microarray Data*". Cancer Informatics 2007, vol. 3, page 183-202.

[21] Z.Cai, E.F.Tsung, V.C.Marinescu, M.F.Ramoni, A.Riva, I.S.Kohane. "*Bayesian Approach to Discovering Pathogenic SNPs in Conserved Protein Domains*". Human Mutation 2004, vol. 24(2), page 178-184.

[22] J. C. Burges, "A tutorial on support vector machines for pattern recognition," *Data Mining and Knowledge Discovery*, vol. 2, no. 2, pp. 121–167, 1998.

[23] C. Chang and C. Lin, "LIBSVM: a library for support vector machines," *ACMTransactions on Intelligent Systems and Technology* , vol. 2, pp. 1–27, 2011.

[24] Y. W. Chen and C. J. Lin, "Combining SVMs with various feature selection strategies," *Taiwan University*, vol.

207, pp. 315–324, 2006.

[25] N. Friedman, D. Geiger, and M. Goldszmidt, "Bayesian network classifiers," *Machine Learning*, vol. 29, no. 2-3, pp.131–163, 1997.

[26] Ouali, A. Ramdane Cherif, and M. O. Krebs, "Data mining based Bayesian networks for best classification," *Computational Statistics and Data Analysis*, vol. 51, no. 2, pp. 1278– 1292, 2006.

