An Application of Data Analysis and Data Mining for Multiple Disease Prediction through the Use of Fuzzy Logic

Miss. Shilpa Shaligram Redekar
M.E in Computer engineering
Padm. Dr. V. B. K. College of engineering
Malkapur
e-mail: shilparedekar3@gmail.com

Dr. Manoj Kathane
Padm. Dr. V. B. K. College of engineering
Malkapur
e-mail: manojkathane@gmail.com

Abstract—Medical systems use text mining which have providing facility of health condition monitoring and disease modeling. Medical system work on the PHI of the patient or user. PHI means personal health information. System helps to store data at one place like disease precaution and healthcare. The successful application of data mining found in highly visible fields such as e-business, commerce and trade has led to its application in other firm.

Keywords-Data Mining, Fuzzy Logic.

_____*****_____

1. INTRODUCTION

Data mining is a process of extracting data from database. Generally data mining is used to shows knowledge out of data and represent it in the form of that easily understand to common person.

Disease prediction plays an important role in life of common person. It possible from data mining. Data mining use in medical field for medicine to predict disease such as heart disease, lung cancer, breast cancer etc.

Normally experience doctors classify disease based on various test or diagnosis method [3]. If they got similar symptoms then they are done using their knowledge and experience. It also confirm by various tests. So, we are trying to predict disease by easier way by using fuzzy logic in data mining.

The idea of fuzzy logic was first implement by Dr. Lotfi Zadech of the University of California at Berkeley in 1960's.

Fuzzy logic is an approach base on "true or false' (0 or 1) Boolean logic which better understand to computer. We can say that medical diagnosis process is based on decision making process.

2. LITERATURE SURVEY

2.1 History

Hubert Kordylewski [7], Deniel Grape [6] in 2001 show the application and principal of LAMSTAR. LAMSTAR is large memory storage and retrieval neural network. The LAMSTAR is fast and it can also shrink without any using reprogramming.

Dejan Denevski, Peter Kokol Gregor Stiglic, Petra Povalej elaborates to combine different specialist's opinions of self-organization with purpose to shows accuracy, early and accurate diagnosing of different have proved of vital important in health care.

Jenn-Lung Su,Guo-Zheun Wu [9] introduce the concept of data mining that is database has used in medical information system for processing large volume of data. It means that both type of data, symbolic and numeric will define the need for new data analysis technique and tools for knowledge discovery. They used popular algorithm for data mining which includes decision tree (DT), Bayesian network (BN) and back propagation neural (BPN) network were evaluated.

Using medical text mining and image feature extraction, medical system have been increasingly facilitating health condition monitoring. Author given a PPDM. That means Privacy-Preserving dynamic medical text mining and image feature extracting scheme.
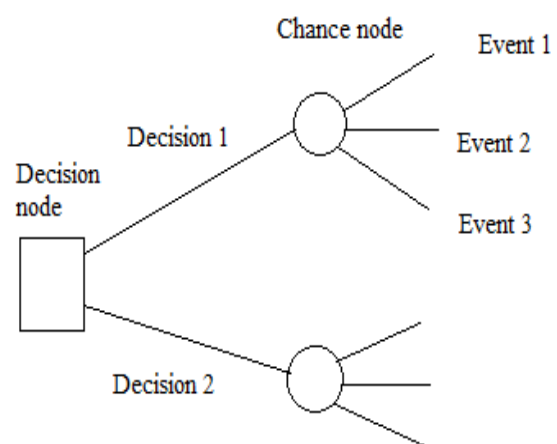

Figure 2.1: Decision Tree.

Decision Tree is a simple is a simple and fast way for classification. Decision Tree consists of nodes where attributes are tested. At the end of each branch, the final nodes represent different classes or events.
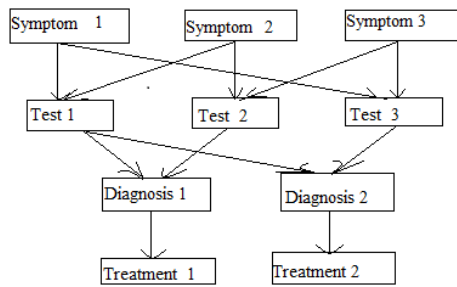

Figure 2.2: Bayesian Network

Bayesian Networks are used more expertise because it provide less complicated diagnosis models. It provide a probability flow chart, where each node represents a possibility and each are connecting the nodes to represents a probability function.

2.3 Motivation

Now on internet, there is a vast amount of medical knowledge available. The problem is that you need to know what you are looking for. If you don't know then it will not help very much. Diagnosis system solves this problem to search medical knowledge with the help of pattern of symptoms and images.

2.4 Problem Definition

On historical medical data apply various predictive mining method. They were constructed by different mode such as hybrid mode, single mode and ensemble-based mode. Hybrid models aim to increase accuracy by reducing the variance of estimate errors and avoiding a biased decision. Generating effective predictive. Models are faced by lack of input data, limitation of constriction and drawback of combination method.

2.5 Objectives

- ❖ To diagnose correct disease.
- ❖ To assist doctors for various diseases associated with symptoms i.e. to be a home assistant.
- ❖ To assist medical students working as in pathological labs.
- ❖ To help nurses, nursing student.

2.6 Parameter

- ❖ Time for disease prediction.
- ❖ Accuracy for database clustering.
- ❖ Accuracy for fuzzy rules.
- ❖ No. of expertise use for database generation.

3.   SYSTEM DEVELOPMENT

3.1 Propose Methodology

It consist of two phases.
1. Training Phase
2. Testing phase

1. Training Phase

In training phase, we create a database by applying fuzzy rule on the various symptoms taken by doctors. Database contains 50 to 100 symptoms of various disease.
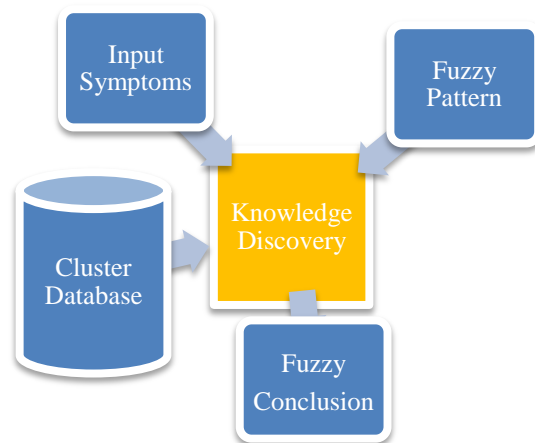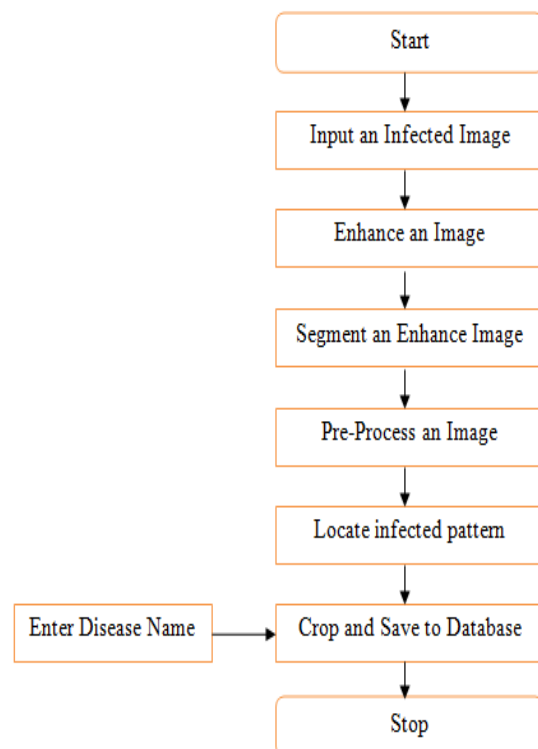

Figure 3.1: Proposed System Architecture

Figure 3.2: Disease Pattern collection from an image (Training Phase)

2.Testing Phase

In testing phase we apply fuzzy rules on symptoms taken by the doctors, Pre-process it and take out a conclusion. Testing phase work with the following steps:

Collect symptoms from patients.
Pre-process symptoms.

1) Extract feature from database, apply fuzzy rules and out one proper conclusion.
2) If conclusion out is about more than one disease then proposed method needs more symptoms and process.
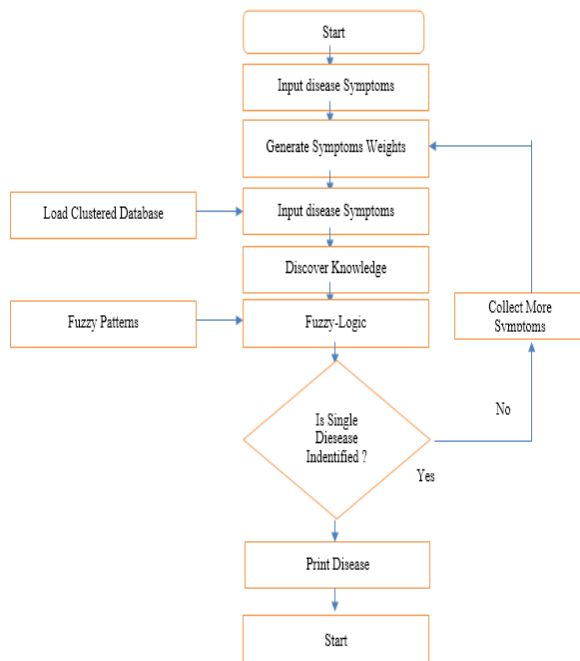3) This method will be iterative until it does not output a single and accurate disease.

Figure 3.3: DFD Testing Phase

❖ Firstly common person gives symptoms which are suffered from disease.
❖ After that input disease symptoms load with clustered database.
❖ Fuzzy logic match with fuzzy pattern.
❖ If it will single disease identified then it will print disease otherwise it collect more symptoms.

4. TECHNOLOGY USE

4.1 Software Requirement

Operating System: Window 8 and above
Programming Language: Matlab, .net, MS office.

4.2 Hardware Requirement
Processor: 1GHz
Memory (RAM): 1GB

Graphics Card: DirectX 9 g1 device with EDI or higher driver
HDD free space: 16GB

5. CONCLUSION

Above proposed concept it is possible for common person to predict his/her disease without visit to expert (doctor). This is pattern and fuzzy logic applied on it. It will provide more accurate result which are based on decision given by expert (doctor). This concept will provide better accuracy and easy interface to common person than any other existing application available in market.

6. FUTURE SCOPE

For the future work, the improvement of image classification technique will provide more accuracy for computer aided diagnosis and more robust method are being developed.

REFERENCES

[1] M.A.Nishara Banu,B Gomathy,"Desease Predicting System Using Data Mining Techinque"2013.

[2] 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.

[3] W. Siegenthaler, Differential Diagnosis in Internal Medicine: From Symptomto Diagnosis. New York: Thieme Medical Publishers, 2011.

[4] 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.

[5] 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.

[6] 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.

[7] 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.

[8] 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

[9] 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.

[10] 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

[11] Shucheng Yu, Cong Wang, Kui Ren, Wenjing Lou in their paper "Attribute based data sharing with attribute revocation"

[12] C.Y. Hsu, C.S. Lu and S.C. Pei,Image Feature Extraction in Encrypted Domain with Privacy preserving SIFT, IEEE Trans. on Image Processing, 21(11): 4593-4607, 2012.

[13] Jun Zhou, Zhenfu Cao, Xiaolei Dong, Xiaodong Lin "PPDM: Privacy-preserving Protocol for Dynamic Medical Text Mining and Image Feature Extraction from Secure Data Aggregation in Cloud-assisted e-Healthcare Systems," IEEE journal of selected topics in signal processing

[14] Olawuni, Omotayo, Adegoke, Olarinoye ,"Medical image Feature Extraction: A Survey," International Journal of Electronics Communication and Computer Technology (IJECCT) Volume 3 Issue 5 (September 2013)

[15] Peter L. Stanchev, David Green Jr., Boyan Dimitrov, "High Level Colour Similarity Retrievl," 28th international conference ICT & P 2003, Varna, Bulgaria

[16] Jun Zhou, Zhenfu Cao, Xiaolei Dond, "Securing M-Healthcare Social Networks: Challenges, Countermeasures and Future Dirctions," IEEE Wireless Communications • August 2013