



---

# Computer-Aided Medical Diagnosis Using Bayesian Classifier - Decision Support System for Medical Diagnosis

De Silva NT and Jayamanne DJ

School of Computing, National School of Business Management, Sri Lanka

---

## ABSTRACT

*This study employs a Bayesian framework to construct a Web-based decision support system for medical diagnosis. The purpose is to help users (patients and physicians) with issues pertinent to medical diagnosis decisions and to detect diseases with highest probability through the Bayesian framework. Users could perform a more accurate diagnosis with the prior/conditional probabilities obtained from selected data sets and compute the posterior probability using the Bayes theorem. The proposed system identifies diseases by analyzing symptoms or by analyzing medical test results. Currently the system detects different types of diseases that people suffer in their day-to-day lives (general diseases) with an average detection accuracy of 92.56%. System also detects complex diseases (e.g.: heart disease - 83.67%, breast cancer - 80.98%, liver disorders - 79.43%, lung cancer - 71.00%, primary tumor - 78.02%, etc.) based on the analysis of the medical test results. The proposed system enhances the quality, accuracy and efficiency of decisions in medical diagnosis since the use of Bayesian theorem allows this system to offer more accurate platform than the conventional systems. Other than that this web-based system provides value-added services in conjunction with CAD system, such as; e-Chat & e-Channeling. More importantly, the targeted user group will be able to access the system as a software element freely and quickly. In this way the goal of this study – which is to provide a web-based medical diagnosis system is effectively achieved.*

**KEYWORDS:** *Bayesian framework, medical decision support systems, computer-aided medical diagnosis, probability distributions, diseases*

## **1. INTRODUCTION**

Computer-aided medical diagnosis (CAD) has become far more widespread in the world and provides real world business solutions to users in areas ranging from automated medical diagnosis (Chang, 1998) to the extended applications such as decision support tools, clinical diagnosis, prediction of diseases, etc.

In medical science, Bayes' theorem can be used as the logical process of performing medical diagnosis, particularly in automated medical diagnosis decision support systems (Sahai, 1991). This research also incorporates the theoretical framework of Bayesian classifier to implement a web based medical diagnosis decision support system to perform medical diagnosis and find appropriate recommendations and solutions when encountering medical diagnosis problems. This web-based system also provides value-added services in conjunction with the CAD system, such as; e-Chat & e-Channeling. The proposed system provides users with following facilities.

- i. Effectively access the knowledge and provide solutions when needed from clinical databases.
- ii. Help doctors when decision making.
- iii. Automatic diagnosis.
- iv. Facilitate a communication channel between two peers over a web-browser.
- v. E-channeling.

## **2. BACKGROUND**

### **2.1. Importance & Advancements of Bayes theorem in Automated Medical diagnosis**

When considering medical science, Bayes theorem plays a major role in intelligence systems by modelling the underlying process of medical diagnosis. When considering about clinical medicine, clinical diagnosis is very crucial.

Medical diagnosis is based on several different parameters like symptoms, allergies, signs, etc. Physicians diagnose a disease by making a subset of all the possible diseases depending on the symptoms provided. Similarly, the advancement of mathematics & computer engineering has achieved a greater success in automating computer-aided medical diagnosis (Sahai, 1991). Accuracy of the computer-aided medical diagnosis depends on the wide range of information used to calculate the probability (Sahai, 1991; Chung & Lu, 2009).

The logical process behind the medical diagnosis can be expressed as a mathematical model that can indicate meaningful relationship between disease and a particular set of symptoms or medical test results. The computer based diagnosis requires a suitable and an applicable mathematical model and the use of a computer to store, manipulate and process wide range of data related to disease, symptoms, laboratory results etc. Therefore there are several mathematical techniques that can aid the computer-based diagnosis; which are Bayes' theorem, logistic regression, likelihood ratio, discriminant analysis, and numerical taxonomy (Sahai, 1991).

Medical decision support systems can be traced back to the Dombal's acute abdominal pain diagnosis system and Shortliffe's MYCIN (blood infectious diseases diagnosis system).

'Dombal's acute abdominal pain diagnosis system' was developed in 1970s (Chang, 1998; Silverman, 1992; Shortliffe, 1990) to diagnose acute abdominal pain in an emergency room. The research was done with 304 case studies to successfully identify most of the acute appendicitis patients. However, in 6 cases it generated misjudgments; non-specific abdominal pain for acute appendicitis.

MYCIN is another decision support system that offers recommendations about the treatments and type of dosage (Shortliffe, 1990; Shabot & Gardner, 1994). This is a system that separates

the problem-solving rules and inference engine of domain knowledge, using rule-oriented syntax through program code. Because of the birth of MYCIN, two other expert systems were derived from it: CLOT (for diagnosing abnormal bleeding) and PUFF (examining lung functions). This study employs a Bayesian framework to construct a web-based decision support system for medical diagnosis because the Bayes theorem has been frequently used in many studies and has performed remarkably well in clinical applications against the independent assumptions.

## 2.2. Computer Technology and Computer-aided Medical diagnosis decision support systems

Computer-aided Medical decision support systems could be classified into Probability systems and Knowledge based systems (Sahai, 1991). Due to the advancement of computer-aided Medical decision support systems, it could assist medical practitioners basically in two areas: medical diagnosis and data interpretation (Chung & Lu, 2009). Now we could implement computer-based or web-based programs to generate more accurate output via computation of various mathematical formulae and help medical practitioners to extract vital information to arrive at better diagnosis, minimizing misjudgments and effectively use the computational analysis in their medical operations (Chung & Lu, 2009). Features of these systems are listed as follows:

- i. Access the knowledge and provide solutions when needed from clinical databases effectively.
- ii. Store patient's historical medical records.
- iii. Help doctors with decision analysis.
- iv. Automatic diagnosis.
- v. Determine suitable medications.

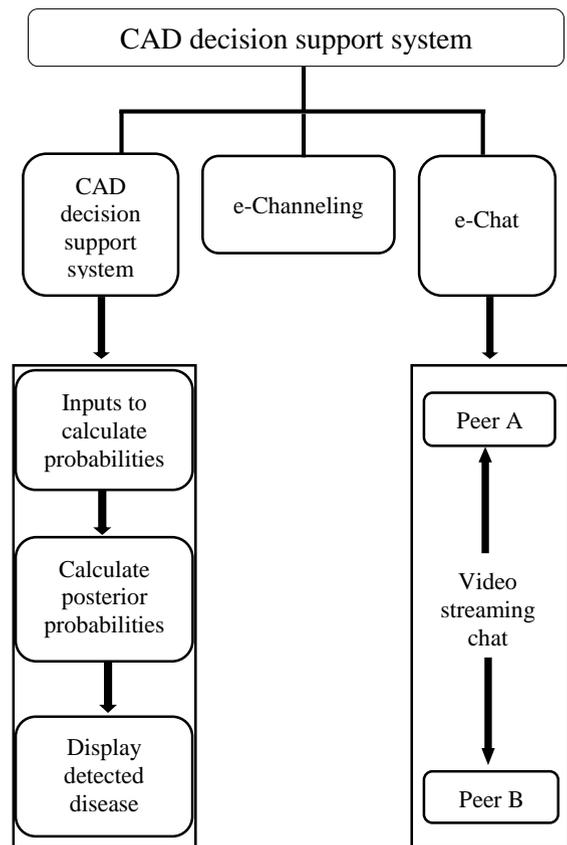
In our work, we use the Bayesian Classifier for accurate medical diagnosis and provide value-added services like e-Chat & e-Channeling in conjunction with CAD system.

## 3. METHODOLOGY

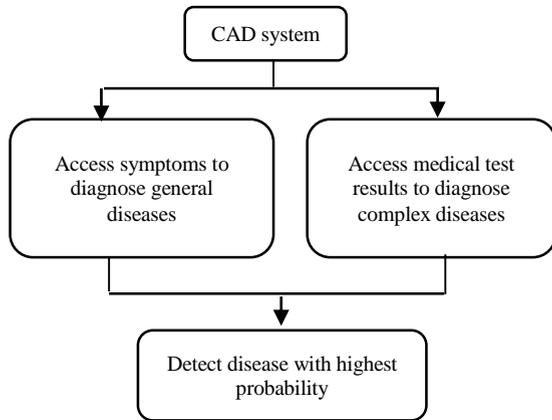
### 3.1. Bayes theorem and System Architecture of CAD system

The proposed system uses probability distributions of symptoms/medical tests of twenty-five (25) diseases and uses the Bayesian classifier to predict the presence of a disease. It has the capacity to predict whether a disease is positive or not for a new set of measurements by using different measures obtained from conducting various tests per disease.

Figure 1 and Figure 2 represent the prototype implementation of the CAD system along with e-Chat and e-Channeling functionalities and the data that the system can access when performing the computer-aided diagnosis respectively.



**Figure 1.** The prototype implementation of the CAD system along with e-Chat and e-Channeling functionalities.



**Figure 2.** The data that the system can access when performing the computer-aided diagnosis

Our system analyses any given number of symptoms ( $S_n$ ) as follows: Let the probability of event  $P$  describes set of symptoms;  $S_1, S_2, S_3 \dots S_n$ . Then by assuming that the symptoms are independent, it follows that;

$$P(S_1, S_2, S_3 \dots S_n) = P(S_1)P(S_2)P(S_3) \dots P(S_n)$$

With the use of Bayes theorem, given the symptoms, the posterior probability of a disease being positive could be computed (Sahai, 1991).

Therefore, we calculate the probability of a particular disease as follows:

$$P(+|S_1, S_2, S_3, S_4 \dots S_n) = \frac{P(S_1, S_2, S_3, S_4 \dots S_n | +)P(+)}{P(S_1, S_2, S_3, S_4 \dots S_n | +)P(+)+P(S_1, S_2, S_3, S_4 \dots S_n | -)P(-)}$$

where  $P(+)$  and  $P(-)$ , are the prior probability distributions of a particular disease. When generating the probability distributions, the prior probability was assumed as 0.5 as there are two (2) possible outcomes in our situation; either the presence of a disease or not.

### 3.2. Peer-to-peer communication over web browser

As an additional feature of the system, when implementing the video streaming chat, web real-time communication (WebRTC) technology is used as the theoretical basis as it provides high accuracy level, flexible support in

different browsers/platforms and real-time communication capabilities over JavaScript APIs'. WebRTC enables browser to browser communication starting from small scale group up to multipart communication and is very inexpensive.

Peer-to-peer live video streaming chat through the web browser is one of the most suitable option that provides a live communication channel for patients and doctors to communicate and establish a high-level of relationship.

WebRTC provides 3 types of APIs' for developers (David, 2014);

- Media Stream (Get User Media): identifies and captures the end-user camera and microphone for use in video chat.
- RTC Peer Connection (Peer Connection): enables audio/video call setup.
- RTC Data Channel (Data Channels): provides peer-to-peer (P2P) data/file-sharing between Web browsers.

Although Real-Time Communications (RTC) provide such services like call control, call handling and present modifications to enable browser-based video communication (David, 2014), WebRTC enables users such services without installing any plugin.

WebRTC provides several features along with the above mentioned APIs'; such as, effective end-to-end performance tests, compatible with most of the platforms, flexible implementation process and accurate issue identification & active feedback (David, 2014).

WebRTC technology also provides service benefits like; implement a solution for a problem as expected, observe customer experience by increasing customer satisfaction, reduce issues regarding time and cost, achieve customer requirements and ready the product with a confident (David, 2014).

### 3.3. e-Channeling functionality over the web browser window

To increase the accuracy and the usability of the proposed system, we have implemented a traditional e-Channeling gateway along with the CAD system.

The channeling sub-section provides users with several facilities like channeling physicians, getting appointments for medical check-ups, generating reminders to keep track of patient's medications along with SMS and email generating facilities, and another four (4) interactive health tools such as BMI (Body Mass Index) calculation, pregnancy calculation, calculate paracetamol dosage for kids and first-aid help over several video tutorials.

## 4. RESULTS & DISCUSSION

In this experiment, we use the Bayesian Classifier for accurate medical diagnosis. We generated probability distributions for selected twenty-five (25) diseases by collecting data and using online databases. Selected twenty-five (25) diseases were categorized into two (2) sectors; general diseases (diseases that people suffer in their day-to-day lives) and complex diseases (diseases that cause long-term harmful effects to the human body).

Data used for the general diseases category were collected over several medical centers in western province, Sri Lanka with relevant permissions from doctors and only the patients who were willing to contribute to this research were examined. Data used for the complex diseases category were collected over an online source 'UCI Machine Learning Repository: Data Sets' (Asuncion & Newman, 2007).

Independency of all the possible inputs/data/measurements were checked mathematically by obtaining the reduced row echelon form of the data matrix as Bayes theorem requires independent measurements.

### 4.1. Results analysis - Generated by Bayes theorem

For testing the performance of the system, considerable amount of data-samples were used as test data and some of the validation techniques were used to reduce the error rate of the system. The system diagnoses a particular disease from the given symptoms by computing the posterior probability for each disease and choosing the disease with the highest probability. The results of general and complex diseases are tabulated in Table 1 and Table 2 respectively.

**Table 1.** Detection Rates of General Diseases

Disease	Training Samples	Test Samples	Detection Rates
Back Pain	250	100	92.85%
Broken Ankle	200	100	88.01%
Bronchitis	450	100	99.99%
Burns	200	100	90.26%
Diarrhea	350	100	99.75%
Ear Infection	200	100	99.84%
Food Allergy	350	100	94.90%
Gastritis	300	100	92.64%
Knee Pain	150	100	96.55%
Neck Pain	250	100	71.11%

**Table 2.** Detection Rates of Complex Diseases

Disease	Training Samples	Test Samples	Detection Rates
Heart Disease	350	100	83.67%
Breast Cancer	500	100	80.98%
Liver Disorders	450	100	79.43%
Lung Cancer	100	100	71.00%
Primary Tumor	450	100	78.02%

### 4.2. Discussion

- a) Independency of the symptoms and medical test results were used after checking them mathematically by obtaining the reduced row echelon form of the data matrix.

- b) Generated good detection rates in both categories: General diseases and Complex diseases.
- c) Larger number of training data samples were used in order to increase the accuracy of the detection rates.
- d) Data used for the general diseases category were collected over several medical centers in western province, Sri Lanka with relevant permissions from doctors. Data used for the complex diseases category were collected over an online source 'UCI Machine Learning Repository: Data Sets' (Asuncion & Newman, 2007).
- e) To provide an ethical clearance of the data, we did not collect any personal details of patients; such as (NIC numbers, Names, Gender, etc.) other than the required symptoms/measurements for probability calculations and testing.
- f) We used questionnaires to gather data for general diseases by attending several medical centers after obtaining the prior approval from doctors and only the patients who were willing to contribute on this questionnaire were examined.

#### **4.3. Limitations of medical diagnostic support programs**

Some users may face difficulties when interacting with the web based systems due to lack of computer literacy.

Sometimes it might be very hard for the doctors to convey the complex understanding of the patient to a computer program efficiently.

#### **4.4. Advantages of e-Chat and e-Channeling**

Users can perform another two important functionalities such as chatting with physicians over a web-based communication channel and e-

Channeling functionality over a one single window along with the CAD system.

All the activities done by the user via the system will be notified/informed either by an email or a SMS.

Implemented functionalities will be time saving and very convenient for users.

### **5. CONCLUSION**

Usage of the Bayesian theorem as the theoretical basis using prior and conditional probabilities to determine the posterior probability helps users to perform accurate medical diagnosis. Implemented CAD system verifies the value of the Bayesian theorem in medical decision support systems.

Integration of the theories above with web technology provide a quick and efficient way of providing treatment for the users. Physicians can use this system to make better decisions in medical diagnosis and the users have the opportunity to use these functionalities (CAD system, e-Chat and e-Channeling) quickly and efficiently over a one single browser window.

### **REFERENCES**

- ASUNCION A & NEWMAN DJ. UCI Machine Learning Repository [http://www.ics.uci.edu/~mllearn/MLRepository.html]. Irvine, CA: University of California, School of Information and Computer Science. 2007.
- CHANG CC. A Computer-Aided Medical Decision Support Tool for Treatment Risk Analysis. The Journal of Taiwan Association for Medical Informatics. 1998; 4(7): 14-25.
- CHUNG CC & LU HM. A Soa-Based Medical Diagnosis Decision Support System Using the Bayesian Theorem and Web Service

Technology. Journal of the Chinese Institute of Engineers. 2009; 32(7); 923-930. DOI: 10.1080/02533839.2009.9671579.

DAVID G. Video on its head. 2014. Available in <http://searchunifiedcommunications.techtarget.com/feature/How-WebRTC-technology-could-turn-enterprise-video-on-its-head> (Accessed on 4<sup>th</sup> March 2016).

SAHAI H. Bayes' theorem: some examples in computer-aided medical diagnosis, Int. J. Math. Educ. Sci. Technol. 1992; 23(2) 11(2): 257-266.

SHABOT MM & GARDNER RM. Decision Support Systems in Critical Care, Springer Journal of the Chinese Institute of Engineers. Verlag, Berlin, Heidelberg, Germany. 1994; 32(7); 2009.

SHORTLIFFE EH & PERREAULT LE. Medical Informatics. Addison-Wesley, Reading, MA, USA. 1990.

SILVERMAN BG. Survey of Expert Critiquing Systems: Practical and Theoretical Frontiers, Communications of the ACM. 1992; 35(4): 106-127.