Rule-Based Reasoning For Medical Diagnosis

by

Mohd Safwan Bukhari

Dissertation submitted in partial fulfillment of
the requirements for the
Bachelor of Technology (Hons)
(Business Information System)

JANUARY 2009

Universiti Teknologi PETRONAS
Bandar Seri Iskandar
31750 Tronoh
Perak Darul Ridzuan

CERTIFICATION OF APPROVAL

Rule-Based Reasoning For Medical Diagnosis

by

Mohd Safwan Bin Bukhari

A project dissertation submitted to the
Computer and Information Sciences Department
Universiti Teknologi PETRONAS
In partial fulfilment of the requirement for the
BACHELOR OF TECHNOLOGY (Hons)
BUSINESS INFORMATION SYSTEM

Approved by,

(Ms. Diana Wong Mei Leng)

UNIVERSITI TEKNOLOGI PETRONAS
TRONOH, PERAK
JANUARY 2009

CERTIFICATION OF ORIGINALITY

This is to certify that I am responsible for the work submitted in this project, that the original work is my own except as specified in the references and acknowledgements, and that the original work contained herein have not been undertaken or done by unspecified sources or persons.

MOHD SAFWAN BIN BUKHARI

ABSTRACT

Medical diagnosis is a major part of health care services. Clinicians are always looking for a way to better the services and quality of healthcare. This project purpose is to develop a knowledge based system in order to assist clinicians in medical diagnosis. The use of such system is expected to increase the quality of medical and healthcare. The system main purpose is to act as a reminder system to the clinicians. Ultimately, the clinicians themselves are the one who will be making the diagnose decision. The system is there to assist them, by giving timeliness medical information, and to help refresh their memory and reduce the error that could happen during diagnosis. Methodology used in developing the system is phased development of the Rapid Application Development methodology. Several versions of the system will be developed. The result will be an expert system that takes input, which are clinical symptoms from user and then list several disease that the patient might have suffered according to the symptoms. Besides that, suggested treatment will also be display for each listed disease. The system also acts as repository system where it has patient records functions.

ACKNOWLEDGEMENT

The author would like to take this opportunity to express utmost gratitude to the various individuals involved for their time and effort in assisting the author in completing the project. Without the cooperation of these individuals, no doubt the author would have faced some minor complications throughout the course.

First and foremost, the author would like to express his appreciation and praise to God for His guidance and blessings throughout completing the project.

A special thank to author's supervisor, Ms. Diana Wong Mei Leng, for her ideas, guiding, monitoring and teaching the author. Without her support and assistance, the author might not able to complete the project successfully.

The appreciation would be incomplete without giving credit to UTP, especially Computer Sciences Department who has equipped the author with essential knowledge, theories and skills.

Finally, to author's family and all individuals that has helped the author in any way, but whose name is not mentioned here, the author thank you all for assistance.

TABLE OF CONTENTS

LIST OF ILLUSTRATIONS

Page

LIST OF FIGURES

LIST OF TABLES

CHAPTER 1
INTRODUCTION

1.1 BACKGROUND OF STUDY

Medical practitioners strive to provide quality health care services to patients. In order to achieve that, decisions that are made regarding patient have to be precise and timely.
The usage of Information Communication Technology (ICT) for healthcare can help to overcome issues like junior clinicians making error or afraid of making crucial decision in the absent of their seniors and doctors making mistake. Having a Decision Support System (DSS) in clinical decision making, can introduce standardization practice methods to increase the homogeneity and accountability of healthcare decisions and actions. The usage of DSS can reduce cost and increase the quality of healthcare as well as public opinion satisfaction.

1.2 PROBLEM STATEMENT

Clinical and human errors. Naturally, humans are prone to error. This scenario could happen in the process of decision making. In healthcare, the slightest mistake could prove fatal, so the need to reduce the possibility of error from happening is a major issue.
Knowledge sharing for new medical practitioners. Junior clinicians sometimes are scared or late in making decision, without the advice of their seniors. But situation where senior clinicians are available to them 24/7 is highly improbable. A system that can replace the role of senior clinicians is vital in medicine practice. Also helps to refresh the memory of many medical scenarios.
Timeliness. To make decisions, clinicians have to take into account of factors such as the patient current symptom and the patient medical history. The need for the clinicians to receive all of the information in a timely manner is important so that decisions can be made as soon as possible.

1.3 OBJECTIVE

The main objectives of this research are:

- To study the potential of intelligent systems in assisting decision making in medical diagnosis and prediction.
- To study the medical practice in diagnosing patients illness and identify predicaments/opportunities to enhance the current situation
- To identify the list of related disease/symptoms that are usually hard to diagnose and create rules based on researches to contribute to Clinical Decision Support System.

1.4 SCOPE OF STUDY

The scope of this project is to focus on 5 types of test cases based on disease and symptoms for diagnosis purposes. Apart from that, knowledge on the treatment also needed, so that the system can detail out what treatment needed for a particular type of disease.

The rules based on these 5 test cases will be developed in the knowledge based of the intelligent systems. The project will be implemented 3 phases:

- Phase 1: Study the concept part of artificial intelligence and the application of it in medical practices. Rule-based is choose as the methodology to be use.
- Phase 2: Acquire the medical knowledge from interview and research and design the framework of the system intelligence.
- Phase 3: Using Microsoft Visual Web Developer 2005 to develop the prototype and the real system.

CHAPTER 2

LITERATURE REVIEW

2.1 KNOWLEDEGE BASED SYSTEM TO ASSIST MEDICAL DIAGNOSIS

2.1.1 History And Purpose

In the early years around the 50s diagnostic system can classified into three categories, which are neural and belief networks, probabilistic, and logical or now known as rule-based. The neural and belief networks, works by training it with set of solutions so that they can be able to make decision to a new problem. Probabilistic on the other hand, adopted the Bayes rule, where it looks on the disease in a population and clinical findings to make assumption and calculate the most likely explanation. Its basically calculates the probability using mathematical formula. Logical uses sets of rules similar to the if-else concept to reach to a decision [11].

Eneida A. Mendoca et al. believes that no one in the medical world can abreast all the medical information or knowledge. Considering that almost 20,000 journals are published every year. It also has been stated that in the United States, preventable medical errors may cause up 44 000 to 96 000 deaths a year cost $29 billion. These errors are related to diagnostic evaluation, operative procedure, ward management, or drug prescription [16].

P Ramnarayan et al said that diagnostic happens due to errors of omission, where the clinicians fail to consider all clinically relevant diagnosis during initial workup. They added that, difficulty in extracting relevant information from sources such as textbooks are the additional contributors of diagnosis errors.

As a result, clinicians should employ a Clinical Decision Support System to help physicians to store and process patient's data and other medical detail should be employed to address to correct the diagnosis and reduce the rate of errors in medicine or healthcare.

Dimitris A. Kalogeropoulos et al. say that the purpose of such system is to support the scientific homogeneity and increase the accountability of medical diagnosis and healthcare. The benefits that can be expected are overall reduction in cost, improved quality of healthcare, patient and public opinion satisfaction.

2.1.2 Basic Structure

The basic structure of an Expert System is shown in Figure 2.1. Clinical Decision Support System (CDSS) - a type of an Expert System – has the same basic components.

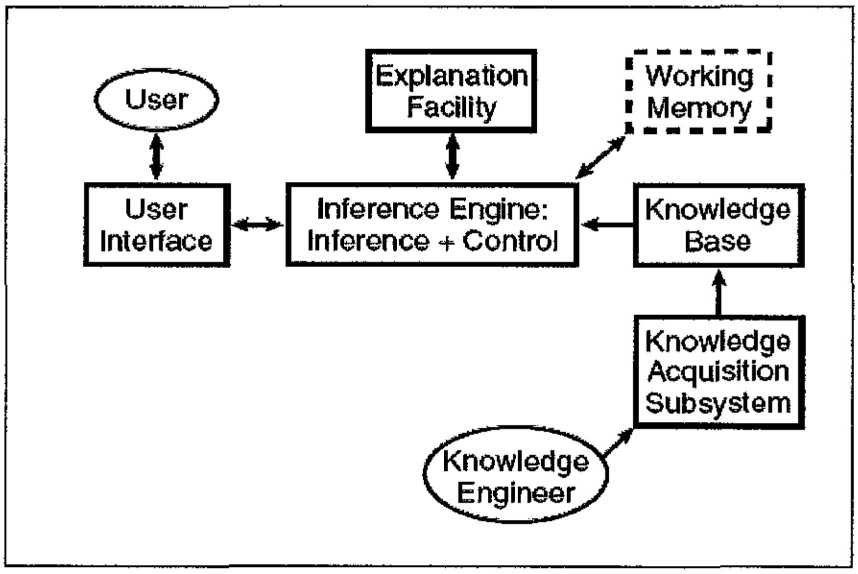


Figure 2.1: Basic Structure of an Expert System

Eneida A. Mendonca et al. (2003) is very definitive, that the Inference Engine (IE) is the core of the system. According to Nils J. Nilson

> The Inference Engine consist of all the process that manipulate the knowledge base to deduce information requested by the user – resolution or forward or backward, for example (p. 282)

The interference engine makes use of the medical knowledge that is kept in the knowledge based and the patients' data which might include sex or date of birth. It uses all these knowledge and information to provide decisions or conclusions. Further more, it can also perform control features such as giving alerts and reminders.

Specifically for CDSS, the IE applies the knowledge of clinical medicine that is represented in the knowledge base with the patients' database. This database is known as 'working memory'. IE also performs control features of the system, such as giving reminders base on a patient's condition [11]. Knowledge engineer is the person who works with expert(s) in clinical field, to represent the relevant knowledge of the expert(s) to be apply in the knowledge base. Usually the person is a software engineer or a system analyst.

2.1.3 Isabel Diagnosis Software

Isabel is web-based diagnosis software that is introduced by Isabel Healthcare Inc, USA. It works by accepting key findings or a whole-text entry and uses a novel search strategy to identify possible diagnose from the input and clinical findings. Mark L.Graber, MD et al. tested the Isabel with 50 Internal Medicine case records that were published in the *New England Journal of Medicine*. The test indicated that Isabel has a 96% success rate in suggesting the correct diagnosis with key findings entry and 74% success rate if the entire case history was inputted into it.

Isabel allows user to enter as many text input as they would like, although result are best when only few important finding or input are entered. Isabel uses thesaurus and natural language process to indicate and match the terms with the rules and algorithm. If user were to paste the entire histories, it allowed an even faster entry but it will reduce the system sensitivity due to the negative findings of the terms to match with the rules [22].

It offers two main functions: Isabel Diagnosis Reminder System (IDRS), and Isabel Knowledge Mobilizing System (IKMS). The IDRS provides a list of possible diagnosis, drugs based on the input that the user give, which is the patient clinical features. The system will list the possible diagnoses to the user, but it will not be ranked by order of importance or probability because it has been assumed that there will always be information that are not entered into the system. Ultimately the clinicians are the one who will have the responsibility to do the judgment. The system is there as a reminder system only [23].

In conclusion, Isabel is quick in making decisions and suggesting the correct diagnosis. However it has to be tested on the natural environment to test its real potential in slimming down diagnosis error [22].

Isabel uses *Autonomy* as its interference engine. This software utilizes the Bayesian interference and Shannon's principles of information theory. It works by collating text related to one specific diagnosis under a single diagnostic label within a diagnostic tree [16]. Appendix B shows the interface of the Isabel.

2.1.4 Intelligent Medical System for Diagnosis Of Bone Disease: XBONE

The system comprise of three major parts: a user interface (UI), a database management system (DBMS) and an expert system (ES). The DBMS function is to manipulate the patient's data. Scintigrams will be acquired by a camera and store in the patient's database via the Data & Image Manager. The main part of the system is its ES, where it will perform the diagnosis [27]. Figure 2.2 shows the system architecture.

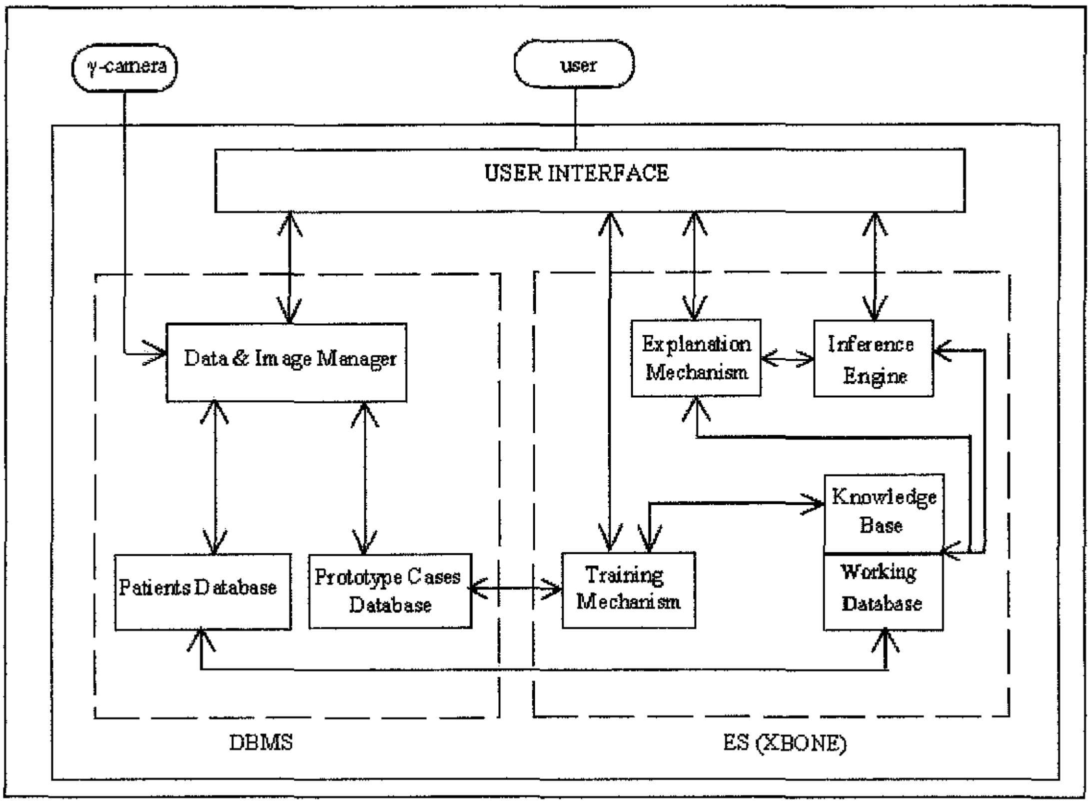


Figure 2.2: Architecture of XBONE

According to Hatzilygeroudis et al., the system uses an integrated knowledge representation formalism based on production rules for the Knowledge Based. It is important to note that in diagnosing a disease, not all of the symptoms have the same significance. In order to represent this scenario in the formalism rules, a factor is to be assigned to each condition of a rule. Adaline unit is adopted to calculate the value of all the factors. Each input, *Ci* is assigned with a number *SFi,* called *Significance Factor.* Even the rules are assigned with a *SFo*, called the *bias factor*, that correspond to the weight of the bias input, *Co* = 1. Different to the conventional rule-based systems, it allows distinguishing between the falsity and the absence of a condition. The output, *y,* is calculated as the weighted sum [27]. Figure 2.3 shows the general syntax of the rule, while Figure 2.4 shows the Adaline unit.

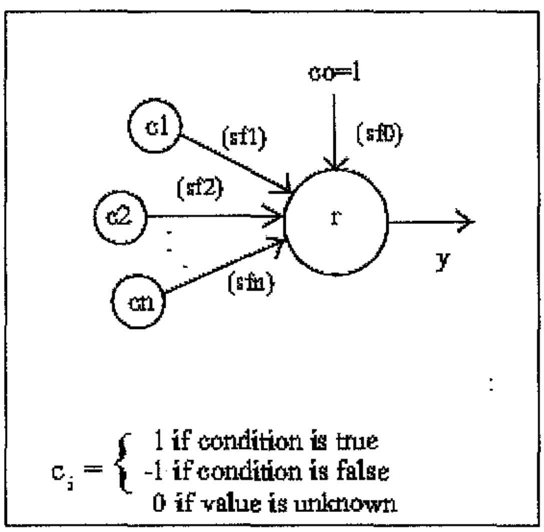


Figure 2.3: Adaline Unit

<rule> ::= [(<bias-factor>)] **if** <conditions> **then** <conclusions>

<conditions> ::= <condition> {**and** <condition>}

<conclusions> ::= <conclusion> {**and** <conclusion>}

<condition> ::= <object> <l-operator> <value> [(<significance-factor>)]

<conclusion> ::= <object> <r-operator> <value>.

Figure 2.4: General Syntax of XBONE Rules

2.2 METHODOLOGIES FOR MANAGING KNOWLEDGE

2.2.1 Rule-based Reasoning

Rule-Based reasoning (RBR) has been largely use in expert system since the first one was develop. Programs that are called expert systems are programs that use the knowledge of a human expert as its knowledge base.

Montani et al.(2002) explained, to design RBR, all the knowledge must be define in a base of production rules, each one containing a piece of information. It works by triggering the rules and using the data to make decision.

There are 3 elements in RBR:

1. Sets of rules
2. Interference engine – mechanism to trigger the rule, usually the program
3. Repository on information that contains the available data for reasoning.

It is said that there are two types of knowledge that can be found in the rule-base's knowledge based: facts and rules. Facts can be defined as true about the domain while rules describe the relation in the domain. It is represented as if-then statements and defines the relationship between the concepts and the problem in the domain of the application [20]. Basically it works by comparing all the facts and inputs from the user with the set of rules that have been predefined to reach a conclusion.

2.2.1.1 Rule-Based in Aachen Aphasia Test

Aphasia is damaged in human brain which causes disturbance of comprehension and formulation of language. They are six-subtypes of test and each test has each own code that signifies the test (e.g.: N5). The rule based, are design according to these unique code .They use the rule-based reasoning to detect which type of aphasia a patient is suffering. By using two types of models, crisp rule-based and fuzzy rule based as the systems knowledge based method [21].

For the crisp rule based, example of the rule that they use are as follows:

a) Rule#1 : If N5 >= 23 Then *Anomic Aphasia*
b) Rule#2 : If P5 >= P3 Then *Broca Aphasia*
c) Rule#3 : If P5 <= 1 Then *Global Aphasia*

Below are the examples of the fuzzy rule-based decision methodology:

(a) If B2 is R40 and B1 is R40 and V0 is R60 and C3 is R0 then Broca aphasia
(b) If B4 is R0 and D B1 is R40 and C1 is R0 then Broca aphasia
(c) If B4 is R0 and C1 is R0 and C1 is R0 and B2 is R50 then Broca aphasia
(d) If B1 is R40 and C1 is R0 then Wernicke aphasia
(e) If B4 is R0 and B2 is R40 and N0 is R100 then Wernicke aphasia
(f) If V0 is R60 and V0 is R60 and B2 is R50 then Wernicke aphasia
(g) If N2 is R40 and B2 is R50 and V0 is R60 and B2 is R40 and P1 is medium and C1 is R0 and V0 is R40 and C1 is R0 then Wernicke aphasia
(h) If B2 is R40 and V0 is R40 then Wernicke aphasia
(i) If C0 is R0 and P1 is Medium and N2 is R40 then Global aphasia
(j) If N0 is R100 then Anomic aphasia
(k) If B2 is R40 and B1 is R40 and C1 is R0 then Anomic aphasia

Figure 2.5: Fuzzy Rule-Based For Aphasia Test

2.2.1.2 Rule-Based in Pap-Smear Test

The purpose of the smear is to diagnose pre-malignant cell changes before these cells progress into cancer. They are 7 different classes of pap-smear cases. Using the crisp rule-based, Tsakonas et al. construct a flowchart (see Appendix C) for the decision methodology to distinguish the classes. There are seven rules defined in the flowchart. These rules are used to distinguish the input to five different classes. Several experiments are conducted and accuracy between 94.6 to 99.8% was recorded. The important aspects of developing the rules are the sequence of execution. Because if a particular rule is not true, it would not proceed to the next step. [21]

From the two tests it has been found out that in order to produce good and credible rules, developer must consult and work closely with the experts. The most challenging part is to convert the facts into rules. Rules are developed in the format of *if-else.* In some particular cases, the rules are sequential in nature, where a rule will lead to another rules instead of results or outcome. For this sort of scenario, a flow chart of rules, like being used in the pap-smear test, must be developed in order to organize the rules sequence and its outcome.

2.2.2 Case-based Reasoning

Cased-Based reasoning (CBR) is a powerful tool for computer reasoning. The concept of CBR is to use past similar problem or knowledge, to solve new problem. The success of CBR highly depends on the quality and quantity of it's predefine knowledge [3].
There are four processes in CBR cycle:

1. RETRIEVE
2. REUSE
3. REVISE
4. RETAIN

When a new problem is given, it will be define as a new case. The system the will RETRIEVE cases or cases from its knowledge collection has the most similarity. The cases will be REUSE to solve the new case. Then the retrieved case is combined with the new case and it will REVISE a proposed solution. The current case is RETAIN into the library for future problem solving.

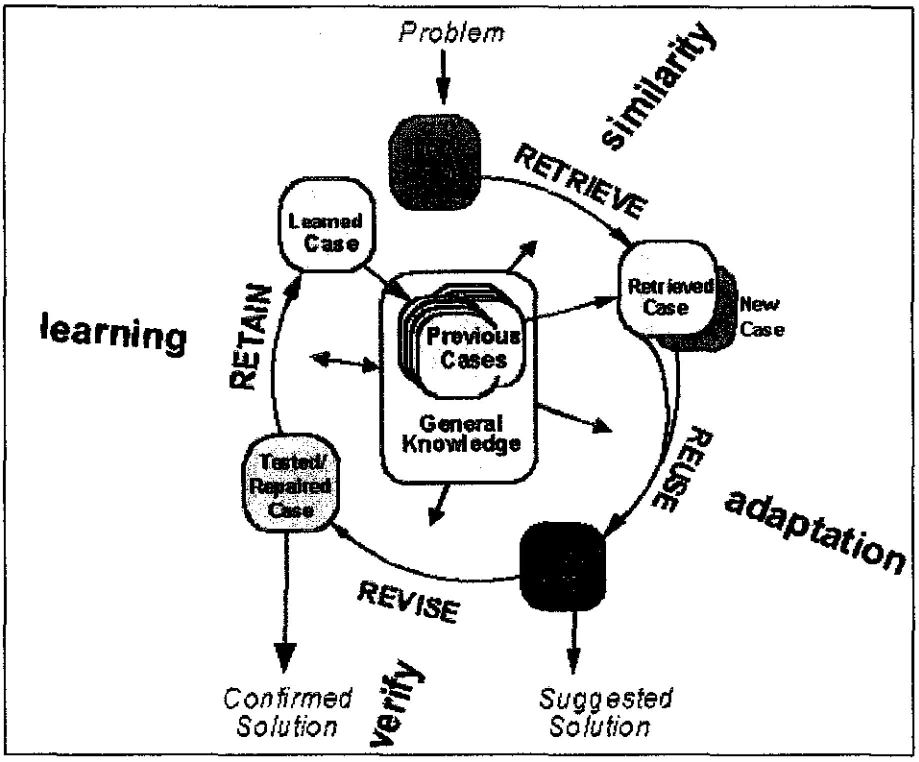


Figure 2.6: CBR Life-cycle

2.2.2.1 Case-Based Diagnosis With Multiple Faults

This study focuses on using case-based to handle multiple faults diagnosis. The focus domain is sonography, where the diagnoses consider several parts of disjunctive subdomains. For an instance liver, that could results in multiple faults or multiple diagnoses. Martin Atzmueller et al. said that the main approaches for this methodology are:

1. Dividing complex cases into several simpler cases
2. Finding solutions for the simpler cases
3. Combining all these simpler cases solution to produce the main solution.

Figure 2.7 shows the process model for case-based diagnosis handling multiple faults.

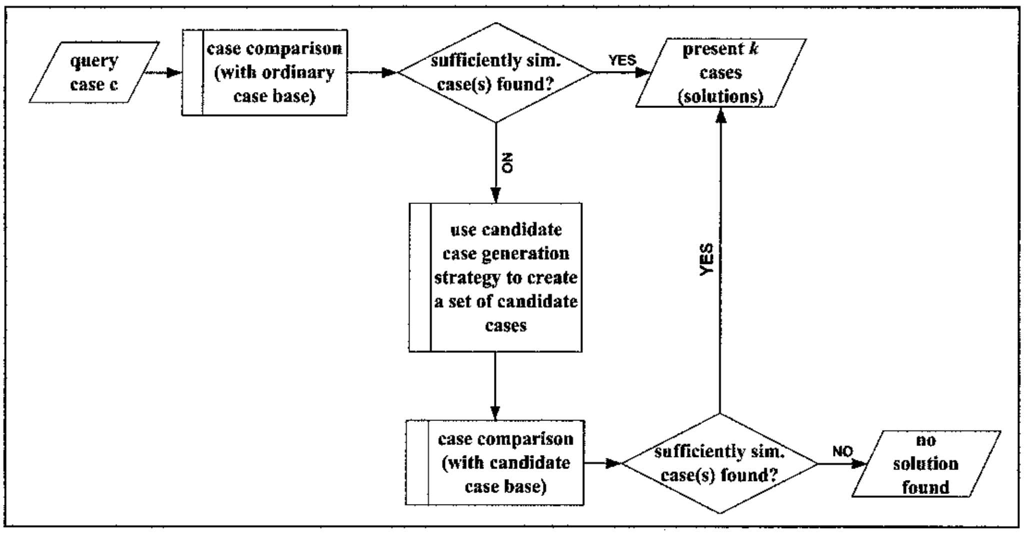


Figure 2.7: Process Model For Case-Based Diagnosis Handling Multiple Faults

The crucial part in the model is to generate the candidate case once no similar cases was found. A candidate case is created from a set of subcases which are merged to one case. The algorithm use to create the candidate cases are as follow:

1. Combining sets of solution contained in the subcase.
2. Combining problem descriptions of the subcases.
3. Conflict resolution. If the combine subcases contains more than one finding for an attribute.

2.2.3 Potential Of Multi Modal Reasoning For Medical Diagnosis

Case-based reasoning (CBR) and rule-based reasoning (RBR) have been use in numbers of system. Both have their own strength and weaknesses. CBR is said to be suited for managing knowledge of the operative type, while RBR is very successful dealing with formalized knowledge [2]. The concerns are, for CBR, little or no common rules can be generalized based on the experience it had about the problem domain. RBR is considered effective only when the theory on the underlying problem domain can be well defined [20].

To be able to take advantages of both methods, and reduce the inefficiency, newly method, Multi-Modal Reasoning (MMR) is proposed. The simple concept is first, using RBR to handle standard problems. If it fails to propose a solution, CBR is then applied to retrieve similar cases from a library of peculiar and non-standard situations [2]. The method provides a better way to solve problem by combining inductive – CBR and deductive – RBR methods.

Domain Knowledge (Rich ↑ / ↓ Poor)		
Rich	Rule-Based Systems	Any one of the three systems
Poor	Integrated Systems	Case-based System
	Poor ← Domain Experience	→ Rich

Figure 2.8: The Categories of Different Types of Systems For Different Problem Domains

T. H. Chi et al is adamant that the integrated system is unique compare to the other methods:

- *Improved problem solving capability*: The integrated system provides an explanation mechanism by which an old case can be extended and explained. The explanation mechanism can generalize the old cases stored in the case-base and consequently make the system applicable to a larger problem domain.

- *Greater flexibility*: The rule-based and case-based knowledge are stored separately in the knowledge base and are independent of each other. If later on, for some reason, the rules (policies) stored in the rule-base part of the system need to be changed, we can easily modify the rule-base portion of the system. This advantage also gives the system we proposed more adaptive power to survive in a dynamic environment.

2.2.3.1 MMR for diabetes care

Montani and Bellazi (2002), in their article, talks about the system that they develop using the MMR methods. The system is about *type 1 diabetes mellitus*. They say that, the decision support of the system is provided by exploiting case-based retrieval results for specializing the rule-base reasoning system behavior, and tailoring it into the individual patient's needs. They mapped the concept of case to the one of the periodical control visit. Each time a patient is visited, a new case is automatically stored in the case library. Each case in the library is linked to the previous one and following ones, using two chains of pointer. This case organization has allowed the system to interpret the patients' clinical condition, using past experiences, in light of the current condition that the patient's is experiencing.

2.3 CREATING TAXONOMY FOR MEDICAL CONDITIONS

2.3.1 Introduction

Taxonomy, according to Oxford English Dictionary is word origin from Greek that brought the meaning arrangement. The branch of science concerned with classification. American Nurses Association in 1999 defines taxonomy as classification according to presumed natural relationships among types and their subtypes [28]. McGregor define

> Taxonomy as hierarchical arrangements of terms that describe a particular branch of science or field of knowledge.

It is hard to organize a particular medical terms into one category. Diseases, "Lymphomas" for instance, falls in three MEDLINE's Medical Subject Heading (MeSH) hierarchies under "Neoplasms", "Hemic and Lymphatic disease Diseases" and "Immunulogic Disease" [25].

2.3.2 Basic presentation Class

According Starren et al, presentations objects can be divided into five major classes, which are list, table, graph, icon, and generated text. They further added

> Two primary criteria were used in creating the classes. First, different classes represent different methods of organizing information in the presentation. A list is organized as a sequence. The rows and columns of a table provide two orthogonal groups of categories into which items are placed. Graphs rely on spatial organization. Icons, intrinsically, are discrete symbols representing discrete concepts. Generated text relies on syntactic rules for organization. The second criterion was that the classes mapped closely to common software objects used for data presentation. In this way, the taxonomy would more closely reflect the actual design choices made by system developers.

2.3.3 MEDLINE's Medical Subject Heading (MeSH) Tree Structured

A French-centric online health resource, Catalogue et Index des Sites Medicaux Franchophones (CISMeF), adopted the MeSH Thesaurus for its organization. Inexperienced user will find it hard to use the MeSH Thesaurus because of its complexity and it needs user with vast experience to use it efficiently. This certainly does not applicable to the novice user. For example, in virology, novice user wouldn't know how to search "virus disease" or "antiviral agents". In order to overcome this, a concept called "meta-term" is introduced. This concept indicates a medical specialty or a division of the biological science. If we take the cardiology field for instance, it will be organize as shown in table 1 below [29].

Table 2.1: Example of MeSH Tree Structure Organization

Categories
• Cardiovascular agents • Cardiovascular disease • Cardiovascular physiology • Cardiovascular procedures • Cardiovascular system • Diagnostic Techniques, cardiovascular
Mesh Associated Terms
• Cardiac care, facilities • Cardiology service, hospital • Heart valve prosthesis

CHAPTER 3

METHODOLOGY

3.1 PROJECT WORKFLOW

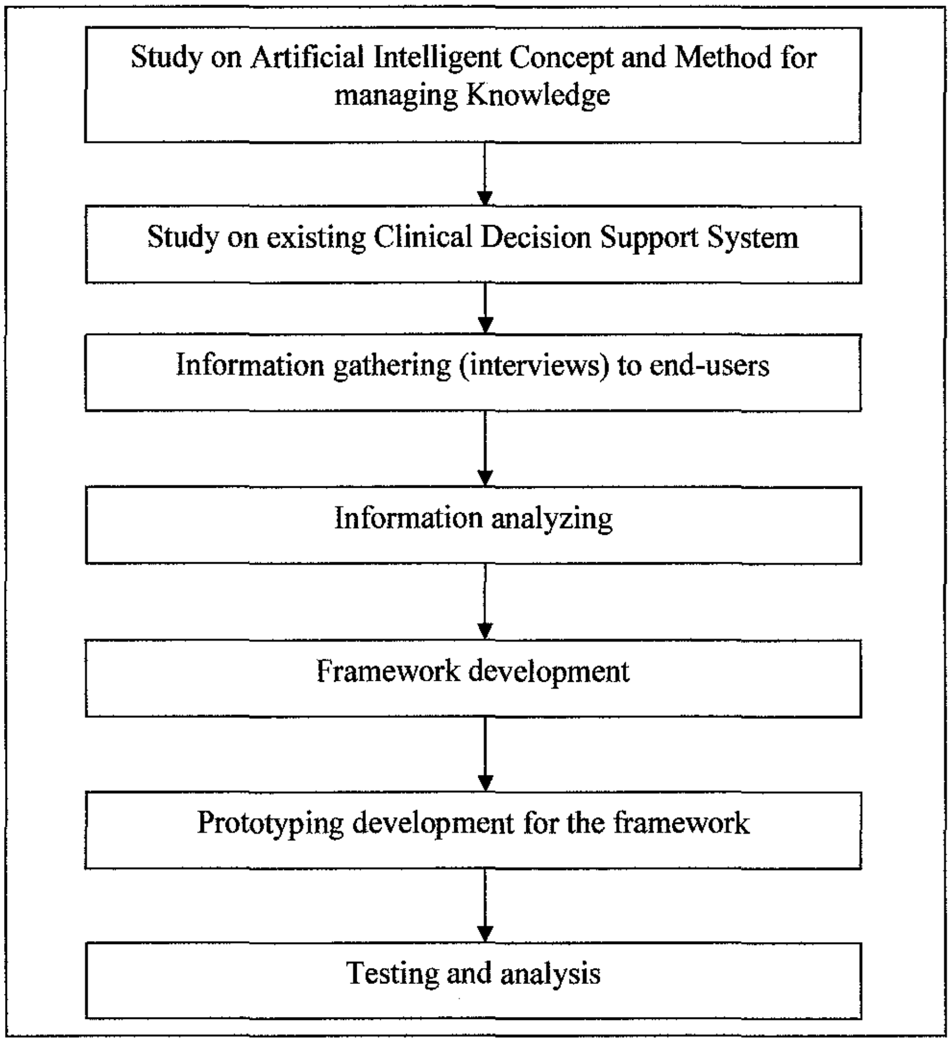


Figure 3.1: Project Workflow Diagram

Diagram above shows the workflow for the study the project. Studies on Artificial Intelligence are a high priority in this project which is why it is being done first. Then work continues on studying existing Clinical Decision Support System. Information gathering is done through interview and research. Information analyzing comes next and by analyzing the information, a framework on the disease detection intelligence is developed. A system prototype will be develop based on the framework, and it will be tested and analyze.

3.2 SYSTEM DEVELOPMENT METHODOLOGY

The methodology that that is adopted by this project is the Rapid Application Development (RAD) Model. The type of RAD that is use is the *phased development.* This methodology allows developer to produce a useful system to be use in short period of time. 3 version of the system will be created throughout the course. The first version will not have full functionality, but the value will grow as the system is near its completion. By using this methodology I can focus on completing a portion of the system first, test, and then move on to another portion. This gives the flexibility as information might not be able to be compiled to complete the system in one shot. Fig. 3.2 shows the workflow of the methodology.

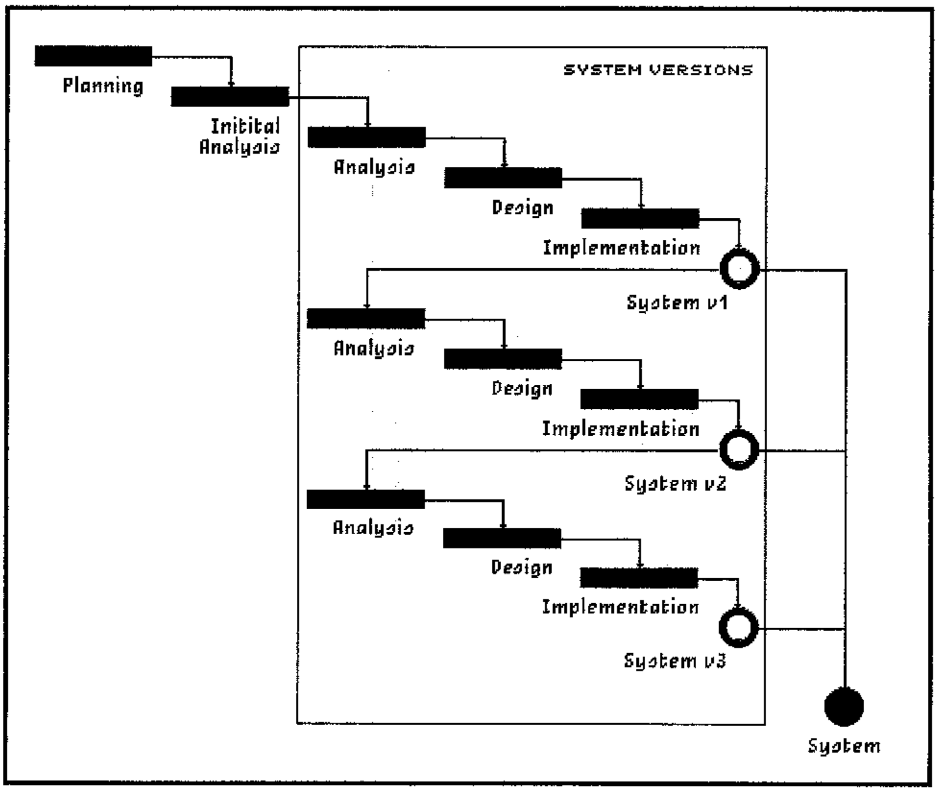


Figure 3.2: Phased Development Methodology

3.2.1 Planning

The activities that will be conducted in this phase are, firstly planning what to do in order to develop the system and set important milestone. Then will be looking at the problem and the feasibility of developing the system. The draft of a system flow and the architecture will be done. To improve knowledge, researches on existing clinical decision support system are to be done. The objective is to identify the various knowledge based methods that developers use in developing such system. Lastly an interview session will be conducted with a specialist of the chosen disease to acquire the medical knowledge.

3.2.2 Analysis

In this phase all the data and knowledge that is acquire in planning will be gather, and analyzed them. Firstly, analyzing will be done to the knowledge based methods that are covered in the previous phase and figure out the advantage and disadvantage of each method. The objective is to choose the best method to be use in developing the system. The collection of the system requirement by questionnaire and interview to the user – clinicians will be done in order to know what the user expect from the system and how they want it to behave. Requirement that is collected is listed and ordered by priority. Requirements with high priority are going to be included in the earlier version as oppose to requirements with lower priority. Lastly, the database for the patient data will be developed according to the medical data and knowledge acquire.

3.2.3 Design

In this phase, works that will be done are, starting on the coding and programming of the system and the database. The coding will be done based on the versions requirements. Each particular version will have different requirements. Testing will be done on each version. After a version has been completed and tested, than only work will continue on the next version.

3.2.4 Implementation

On this phase, system will be presented as per requirement from the university.

3.3 GANTT CHART

The Gantt chart shows the timeline and the important milestone of the project. Refer Appendix D.

3.4 TOOLS AND TECHNOLOGY

3.4.1 Programming Language

Although initially VB.Net was choose as the only language to be used to develop the system, the language now will be develop using ASP.Net and VB.Net as its code behind. The reason to convert the system interface into web-based is for easier navigation between pages and forms. The language is also really compatible with the chosen database. AJAX is also used.

3.4.2 Database

For the database, I've decided to use Microsoft SQL Server 2005, and as the front-end, I will use SQL Server Management Studio Express.

3.4.3 Platform

The system is developed to be run on Windows platform operating system and is to be use with the web browser Internet Explorer, Mozilla Firefox or Google Chrome.

3.4.4 Methodology for Managing Knowledge

The artificial intelligence of the system will use the rule-based reasoning concept. A framework is developed for the rule-based and it will be implemented in the Visual Basic programming using the 'If-Else' decision concept. This section will be discussed more on the next report after more interviews has been conducted to gain the medical knowledge.

3.5 SYSTEM ARCHITECTURE

3.5.1 Logical Architecture

There are three tiers in the proposed logical architecture for the system: Data tier, Application tier and Presentation tier. The Presentation tier is used to handle the display of the management the graphical user interface (GUI) relationship with the modules in the Application tier. The Application tier consists of three modules, the prediction, treatment and record. Lastly the Data tier provides access to underlying SQL data storage of the whole system.

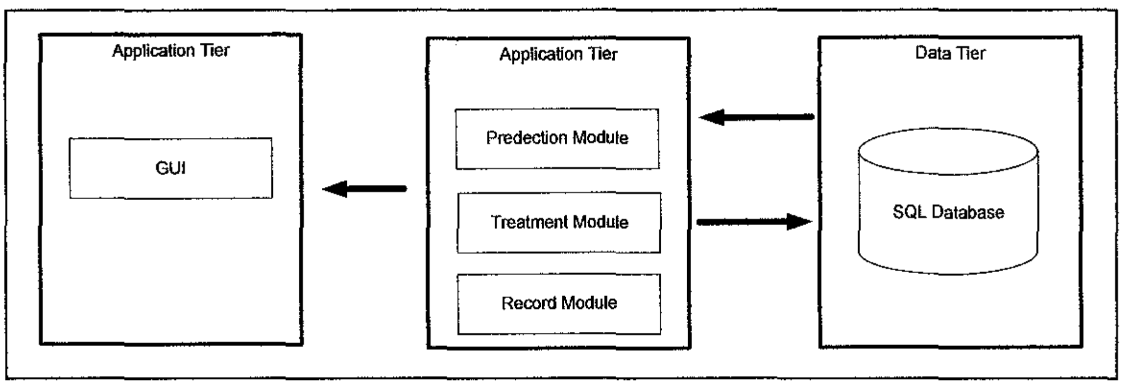


Figure 3.3: System Architecture

Figure 3.4 focuses on the application tier. The Record Module gets the input from the user and save it into the Patient Record section in the database. The Prediction Module would take the data from both user input and the Patient Record in order to make predictions. The treatment module gets the result of the Prediction module and the data from the Treatment and Drug Record from the database and use produce results.

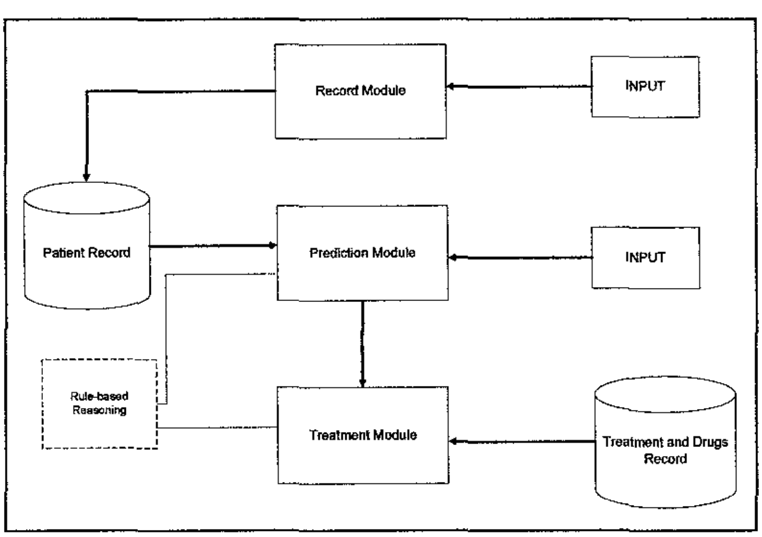


Figure 3.4: Application Tier

3.5.2 Physical Architecture

Since the system is developed using ASP.Net technology, it is a web-based system. However, it is only uploaded in the workstation's server, and to be use internally only on the particular workstation. The type of server to be use is the Microsoft Internet Information Services (IIS) 7. Figure 3.5 shows the illustration of the system's physical architecture.

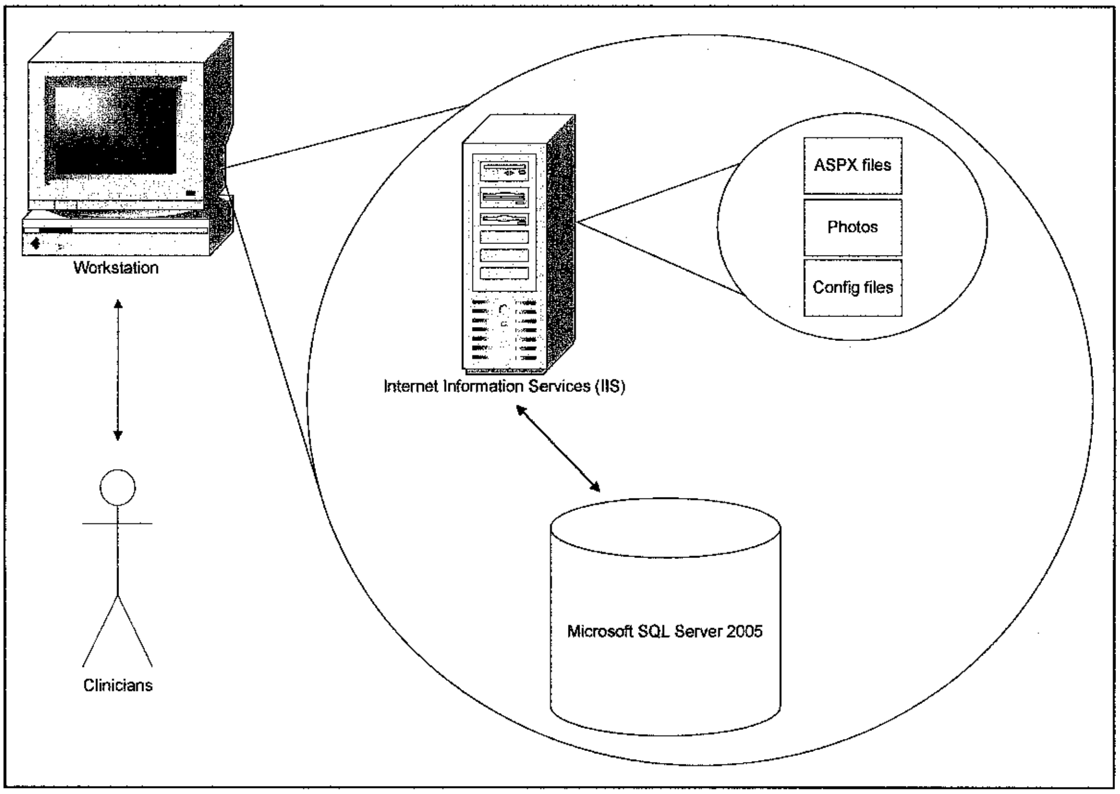


Figure 3.5: Physical Architecture

3.6 SYSTEM MODELLING

3.6.1 Flow Chart

From the flowchart of the system (Figure 3.6), we could see that first; the user will have a choice of either creating new patient record or predict disease. If the user wish to create new patient record, the user will be directed to the create new patient record screen. Here the users are able to save the patient's record in the database. After saving the patient's record, the user will have a choice to continue with disease prediction, or exiting the system.

If the user chooses to predict disease, the user will be directed to the insert symptom screen. Here is where the user insert the symptom and the system use the input to make prediction. The system will also get the patient's medical history from the database. The user input and the data from the database are both treat as an input in the system's artificial intelligence. The artificial intelligence will predict the result and display it.

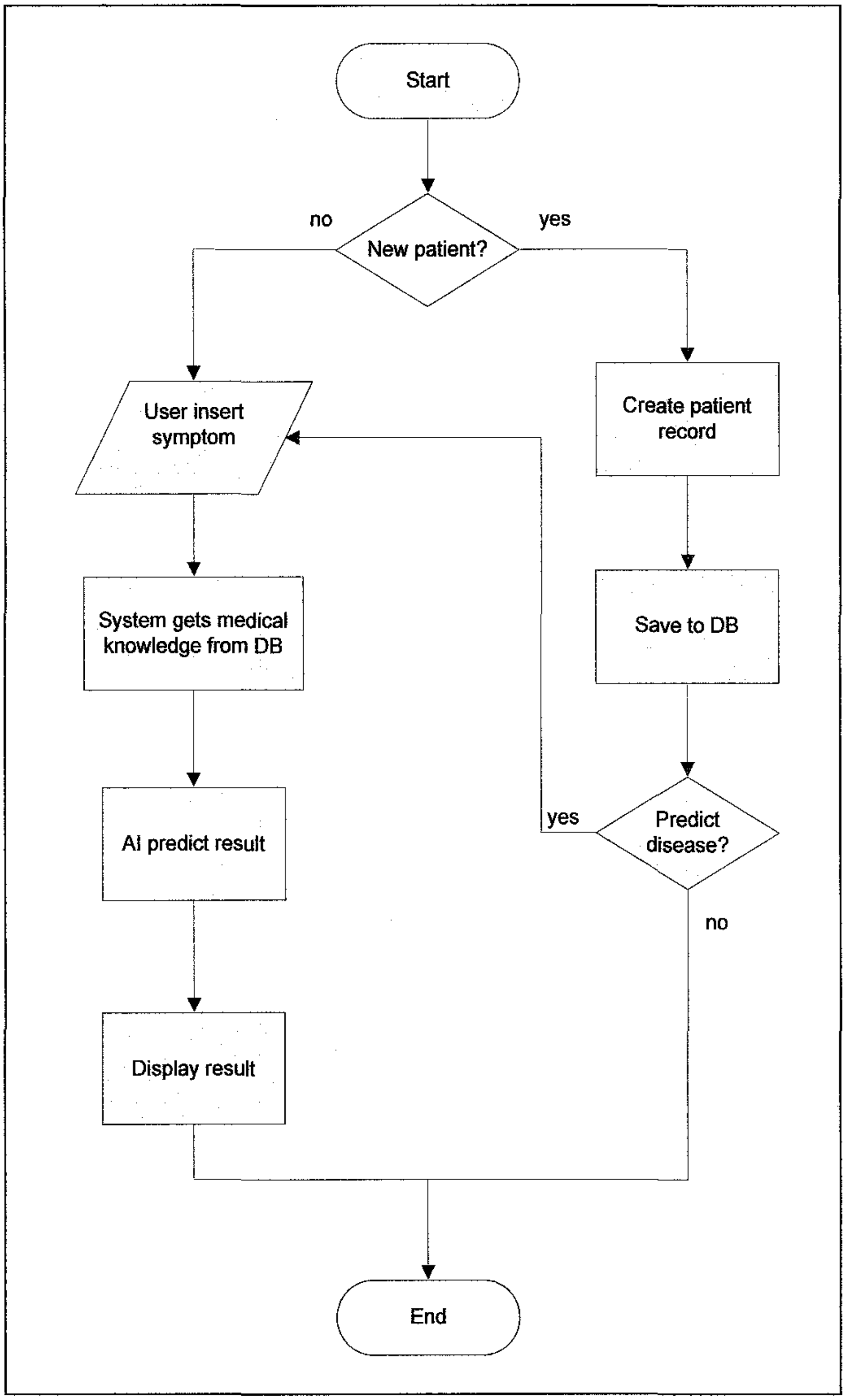


Figure 3.6: System Flow-chart

3.6.2 Use-Case Diagram

The Use-Case diagram (Figure 3.7) is used to describe the interaction between the user, who is the clinician as well as the other entities inside the system, the database and the artificial intelligence. From the diagram, we can see that there are only one types of user in the diagram. The clinicians will be the one who will be creating and editing the patient record, insert the patient's symptoms into the system as well as editing the rules which will be use by the artificial intelligent. The database will provide the patient's medical history, rules to be use by the system and the treatment. The artificial intelligence will use the input from the doctor and the database to make predictions.

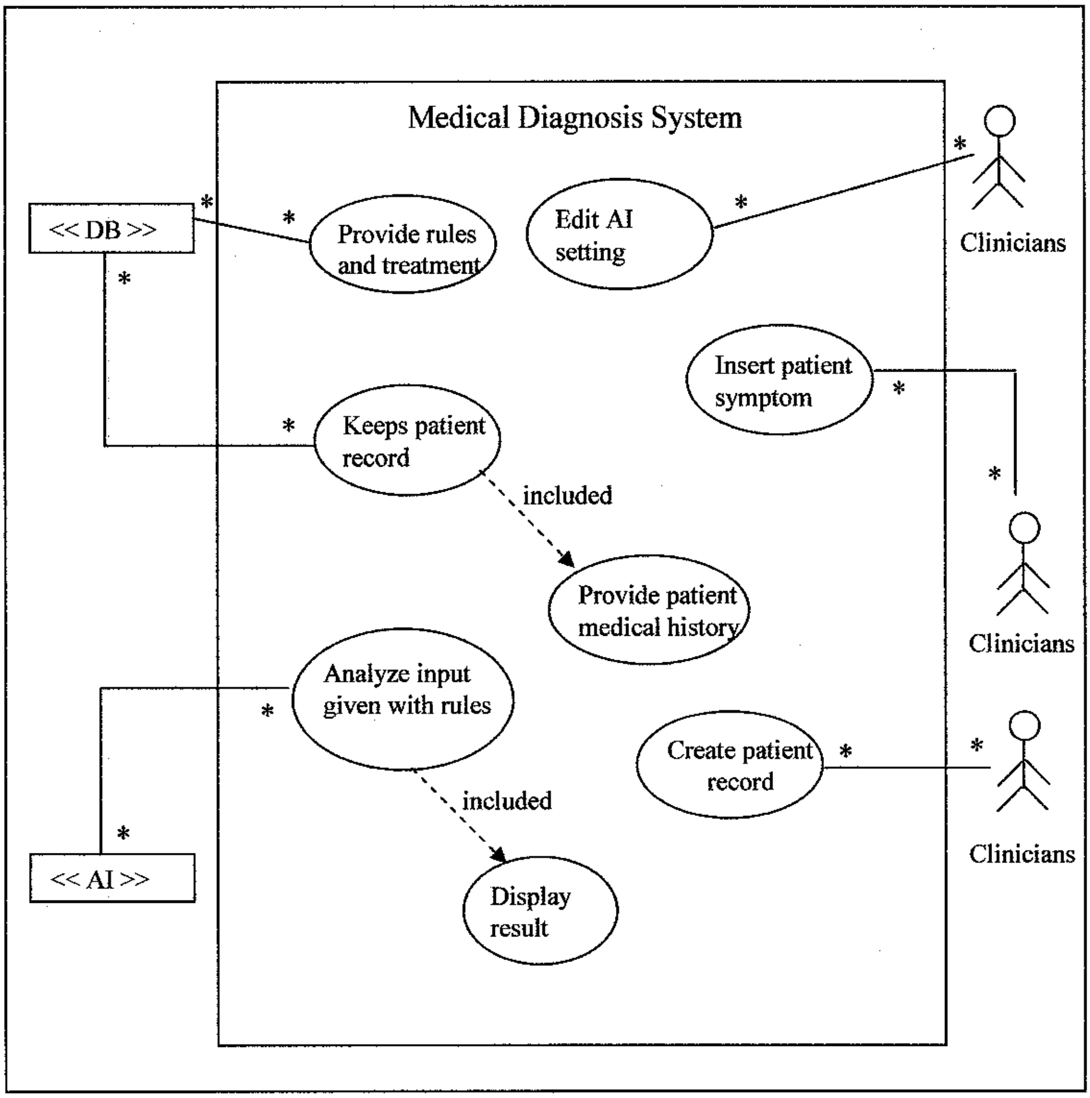


Figure 3.7: System Use-Case Diagram

3.6.3 Class Diagram

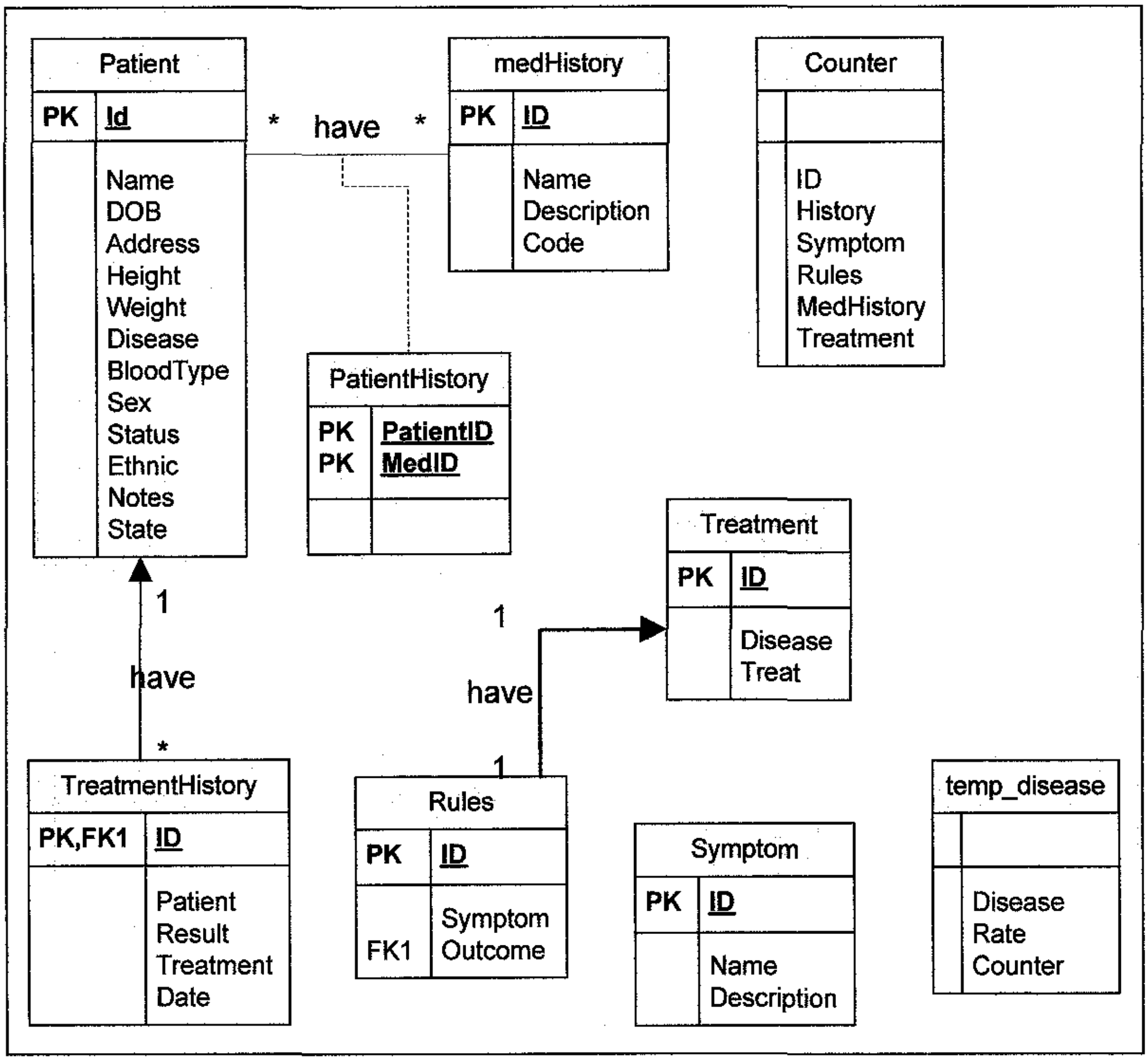


Figure 3.8: System Class Diagram

The system's database consists of 9 tables which are *patient*, *medhistory*, *patientHistory*, *treatmenthistory*, *treatment*, *symptom*, *rules*, *temp_disease* and *counter*. The *patient* tables stores all the information of the patients. The *medhistory* stores the options of the medical history that a patients can have. The relationship between *patient* table and *medhistory* table is, many patients have many medical history. Because of this, a new table is derived from the relationship between those two tables, *patienthistory*. This table has two foreign keys that are taken from patient and medHistory, which are patientID and medID. The *treatmenthistory* stores the diagnosis results that have been done to a patient. The relationship between *treatmenthistory* table and patient table is one patient has many treatment histories.

Symptom table stores all the information about the options of the symptoms that the system can cater. *Rules* table on the other hand, stores the data of the rule that is used by the system's artificial intelligent to make decision. The *temp_disease* table is used to store temporary result that is produced by the system. This table will be clear every time the system has finished its prediction process.

The *counter* table is used to as a counter for all of the other tables' primary key data.

3.6.4 Framework

For the system intelligence framework, the rule-based reasoning methodology will be adopted. There are two part of the framework. The first is the prediction module and the second part is the treatment module.

For the prediction module, the objective is to detect the type and the severity of the disease. Each disease will have a set of symptoms that will lead to it. The system will get the symptom from the user and compare the symptoms in each rule. The framework should detect multiple possibilities and list them to the user and show the matching symptoms for each listed disease.

For the treatment module rule, it will take the result of the prediction module and use it as an input to suggest the treatment for the patient. It will use the same similar 'If-Else' concept as the prediction module. Figure 3.9 shows the illustration of the framework.

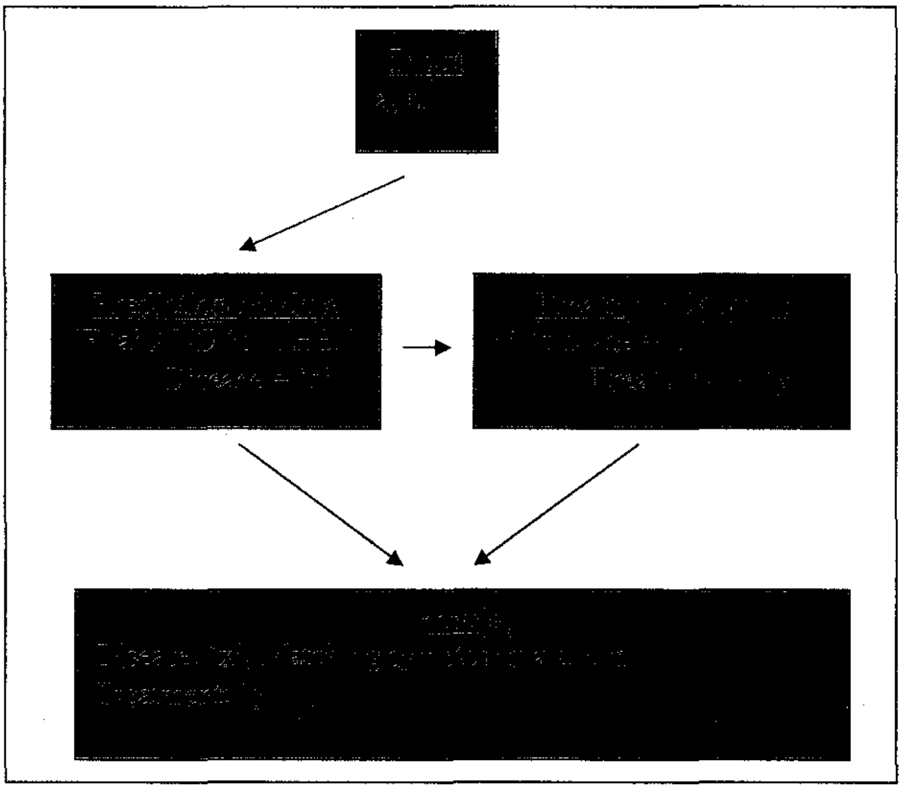


Figure 3.9: Rule-Based Framework

3.6.5 Taxonomy

The medical data in the system rule-based will be classified into three simple categories: symptoms, disease and treatment. They are no sub-categories for all of the three main categories. The rules which are configured in the system will integrate data from all of the three categories. To present the result of the diagnosis, the system will adopt *Simple List* method. Figure 3.10 shows the taxonomy and Figure 3.11 shows the Simple List presentation.

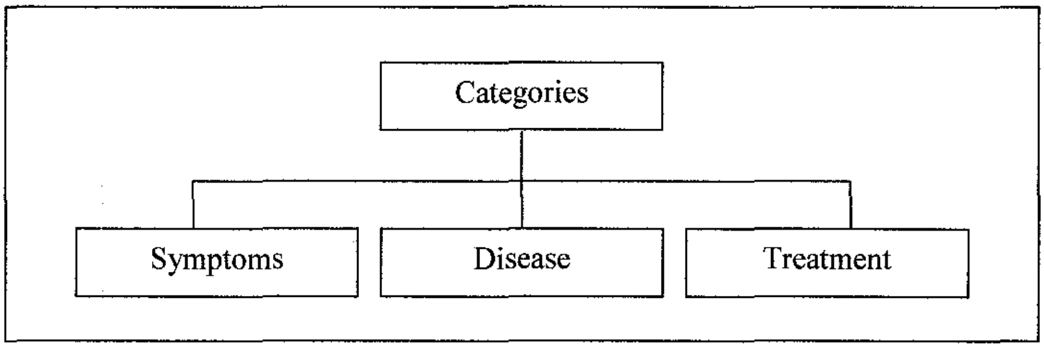


Figure 3.10: Taxonomy

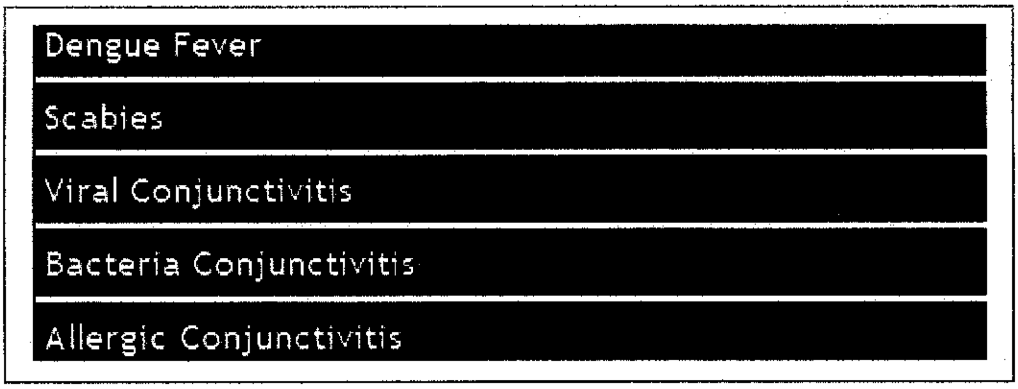


Figure 3.11: Simple List Presentation

CHAPTER 4
RESULTS AND DISCUSSION

4.1 INTERVIEW

4.1.1 Interview Results

An interview session is conducted with Dr. Sofia Zaihan Ismadi from Klinik Zakaria dan Rakan. The interviewee said that since she still practices the old way of detecting disease, where the interviewee will observe the patient symptom, and check for signs. According to the interviewee, symptom is the condition that the patient show or they tell the doctor. Signs on the other hand, are conditions that must be found out. For example, abdominal tenderness, which can only be detected if the doctor touch the patient abdomen and they complain having pain.

It is found out that doctors usually use 'rule-out' method to detect disease. An example would be, if a patient complains that he/she is having paint chest. They are a lot of disease that can be the cause, say asthma and heart attack. So by asking more question and further checking, they can detect more symptoms or signs that can increase the possibility of a disease and ruling out the others.

In terms of having problem with the current practices of detecting disease, the interviewee said that most doctors are used to it because they are trained in such a way. But sometimes they are situations where doctors are not sure what are the patient problems - because they forget some of the medical information. So they have to treat the patient first and ask the patient to come back for a second check- up to make sure that the doctor's assessment is correct. So, having such system can help them to make better decisions as well as refreshing their memory.

The doctor has used computer system to store patient's data. The system, according to the interviewee, makes doctors work easy, as they can just search the patient data from their computer before doing a check-up, as oppose to using paper and files. The doctor also can just type the prescription into the computer and the nurse/clinic assistant will receive the details through their computer. No need for the doctor to write down the prescription then passes it to the nurse.

Regarding the disease, the interviewee suggested the guidelines by Ministry of Health Malaysia (MoH). The guidelines will provide information and knowledge of the disease, including the symptoms and the treatment. The interviewee also advice for specialist advice on the focused disease.

4.1.2 System requirements

Disease Prediction functions. From the interview, it is found out that the system should be able to receive multiple inputs from the user. Only one type of input is to be catered in the system, pre-defined input, where the system provides a list of input for the user to chose, usually implemented using dropdown menu. The result of the prediction should show the disease predicted with its severity and the treatment suggested.

Patient Record Functions. User wants the system to be able to save the user records in a database. The information needs in the patient's database are the patient's name, address, medical history, date of birth, blood type, weight and height. The system should also be able to generate the patient's identification number automatically. The patient identification number is a unique field or data that is used by the clinics or hospitals to distinguished patients, much like an identification card number.

4.2 3rd VERSION OF THE PROPOSED SYTEM

For the 3rd version of the system, there are four main parts – configuration, create, view and predict. The configuration part is where the user can configure the system settings. The create part is where the user create new patient record to be saved in the system database. The view part is where the user can search and view patient's detail. The main part is the predict part where user can insert the patient's detail and symptom to do diagnosis.

4.2.1 Create

The create function is used by the user to create a new patient record. The patient details are divided to two sections, the personal details and the medical details. For the medical details, the user can add the patient medical history by checking as many checkbox from the checkbox list as they desire. User can add more than one medical history for a patient. Clicking the button Clear Form will clear all the details in the form. By clicking save, the system will insert the patient details into the database. (Figure 4.1)

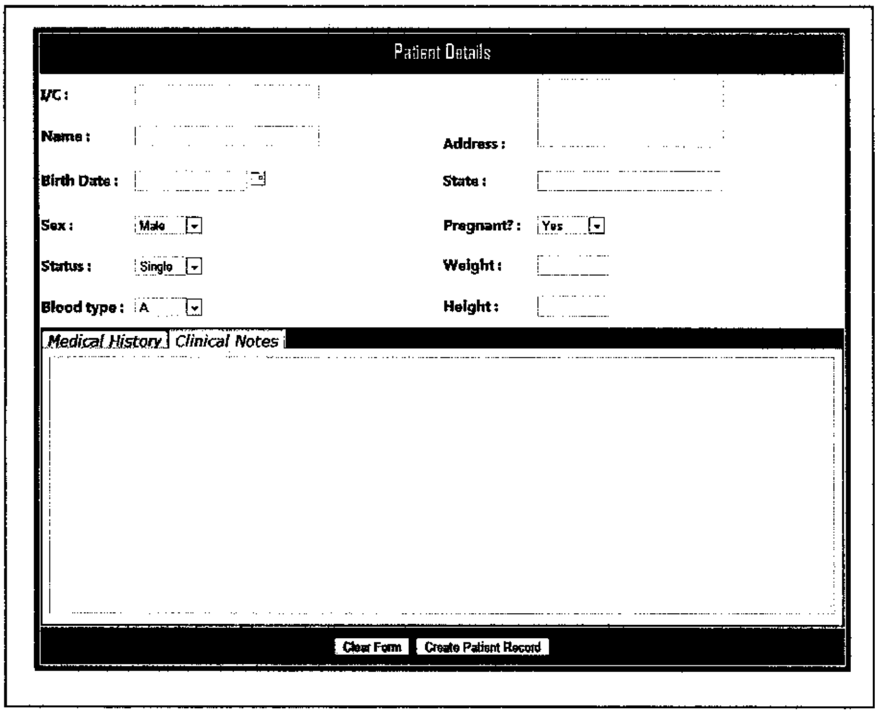


Figure 4.1: Create Patient Interface

4.2.2 View

The view part is divided into two sections. The first section is the search part, where user can search for patient according to 5 different criteria – id, name, gender, status and blood type. The system will list the entire patients that match the selected criteria in a table. (Figure 4.2)

Upon clicking the id of the patient, the user will be linked the particular patient's profile, which is the second section of the view part. In the patient's profile, the user will have the option to edit the patient's details. The patient's profile page will also show the history of the treatments receive by the patient, in the Treatment Receive table. Clicking the link in the table will linked the user to the treatment particulars. The user has the option to do diagnosis to the patient by clicking the link 'Go' in the Diagnose field. This link will take the user to the diagnose page. (Figure 4.3)

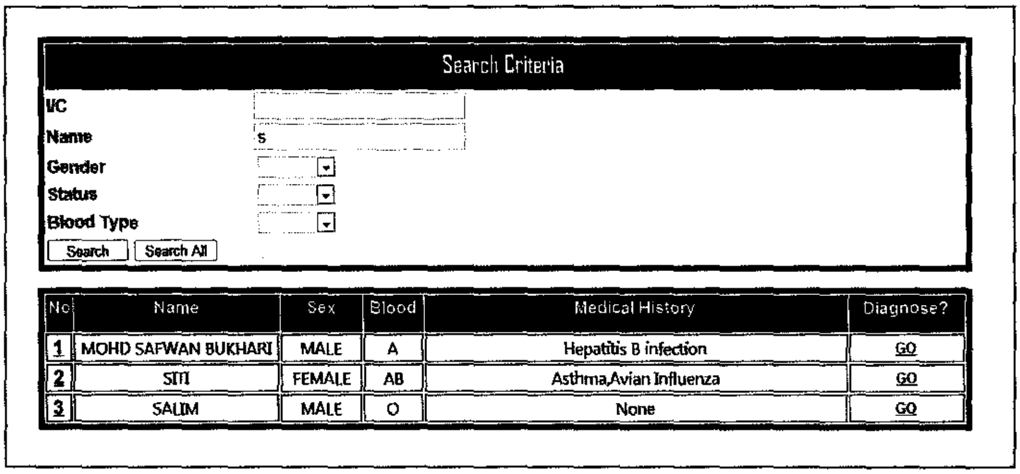


Figure 4.2: Search Patient Interface

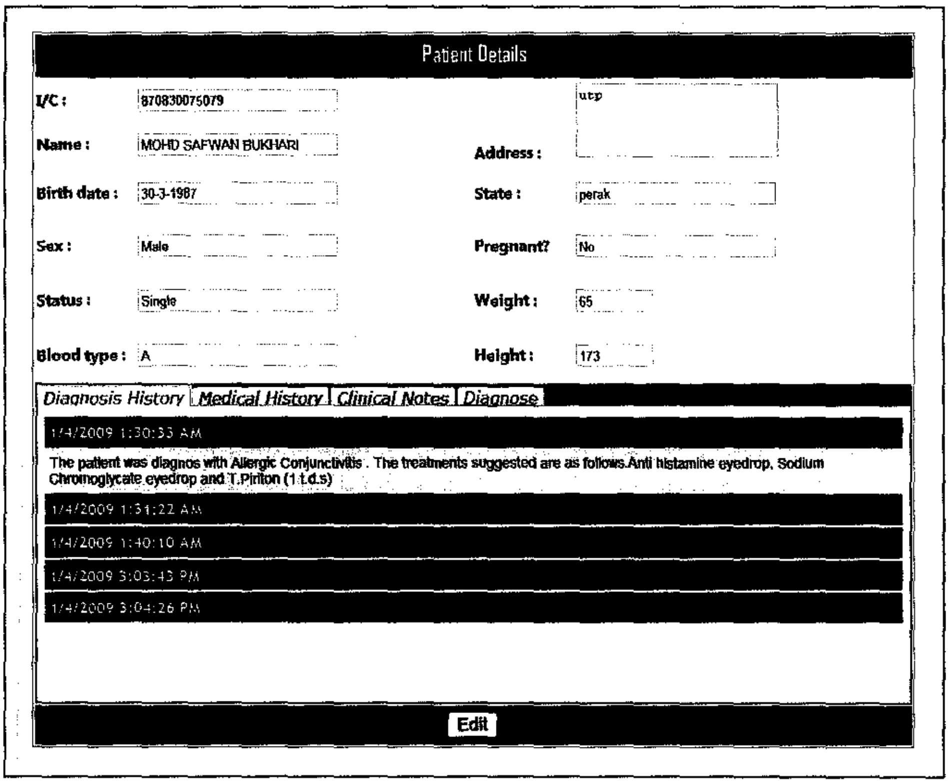


Figure 4.3: Patient Profile Interface

4.2.3 Configuration

The main purpose of it is to allow user to set the settings of the system. This allows the system to be flexible and dynamic. The first part is the option for the medical history checkbox list option. The medical history checkbox list is use to insert the patient's medical history in the *create* and *view* part. User can either add or delete existing options and it will be reflected in the medical history checkbox list. (Figure4.4)

The second part is the symptom checkbox list. The symptom checkbox list is use to insert the patients symptoms which are to be diagnose in the diagnosis page. Same as the first part, user can either add or delete existing options and it will be reflected. (Figure4.5)

The third part is the rules. The rules are to be used by the systems intelligence to do the diagnosis. Rules can be constructed by selecting a number of symptoms that would lead to a certain outcome. (Figure4.6)

The final part is the treatment. Treatment is to be used by the system to display suggested treatment for a particular disease. (Figure4.7)

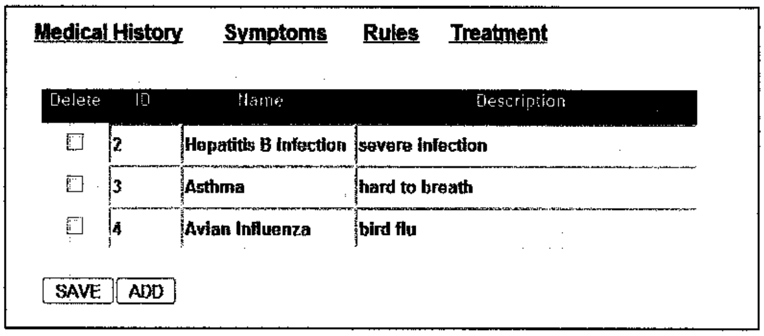


Figure 4.4: Medical History Interface

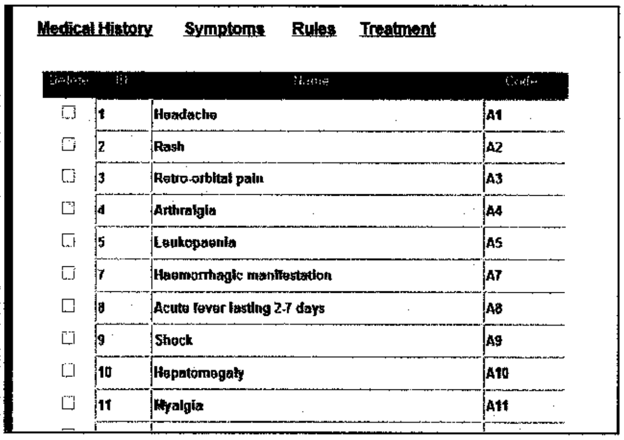
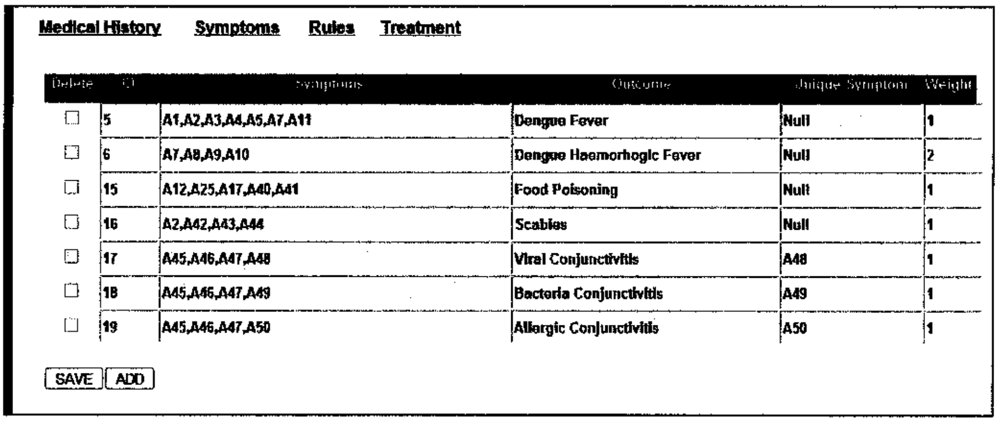


Figure 4.5: Symptom Interface

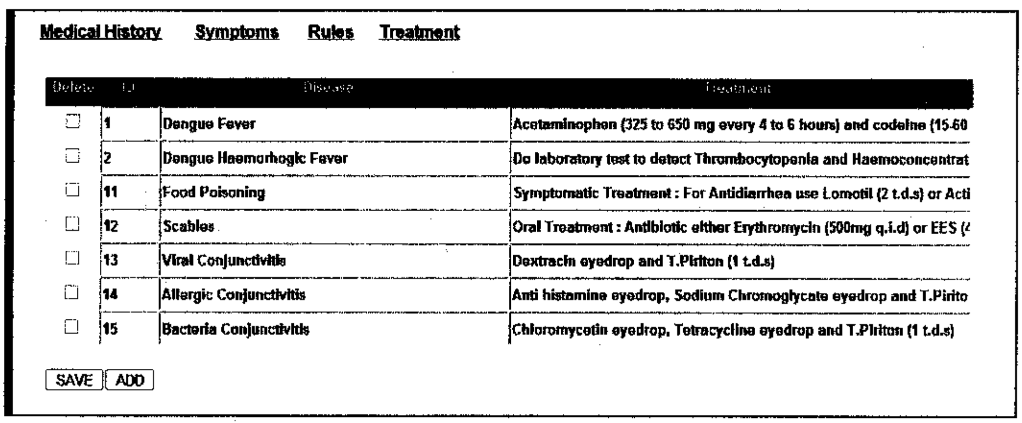


Medical History Symptoms Rules Treatment

Delete	ID	Symptoms	Outcome	Unique Symptom	Weight
☐	5	A1,A2,A3,A4,A5,A7,A11	Dengue Fever	Null	1
☐	6	A7,A8,A9,A10	Dengue Haemorhogic Fever	Null	2
☐	15	A12,A25,A17,A40,A41	Food Poisoning	Null	1
☐	16	A2,A42,A43,A44	Scabies	Null	1
☐	17	A45,A46,A47,A48	Viral Conjunctivitis	A48	1
☐	18	A45,A46,A47,A49	Bacteria Conjunctivitis	A49	1
☐	19	A45,A46,A47,A50	Allergic Conjunctivitis	A50	1

SAVE ADD

Figure 4.6: Rules Interface



Medical History Symptoms Rules Treatment

Delete	ID	Disease	Treatment
☐	1	Dengue Fever	Acetaminophen (325 to 650 mg every 4 to 6 hours) and codeine (15-60
☐	2	Dengue Haemorhogic Fever	Do laboratory test to detect Thrombocytopenia and Haemoconcentrat
☐	11	Food Poisoning	Symptomatic Treatment : For Antidiarrhea use Lomotil (2 t.d.s) or Acti
☐	12	Scabies	Oral Treatment : Antibiotic either Erythromycin (500mg q.i.d) or EES (/
☐	13	Viral Conjunctivitis	Dextracin eyedrop and T.Piriton (1 t.d.s)
☐	14	Allergic Conjunctivitis	Anti histamine eyedrop, Sodium Chromoglycate eyedrop and T.Pirito
☐	15	Bacteria Conjunctivitis	Chloromycetin eyedrop, Tetracycline eyedrop and T.Piriton (1 t.d.s)

SAVE ADD

Figure 4.7: Treatment Interface

4.2.4 Diagnose

This is the most important part of the system. This is where the system will utilize the rules and use it to make prediction. User will input the symptoms of the patients and the system will use the inputs as well as the patient's medical history to do diagnosis. Treatment will also be suggested by the system according to the patient's problem. (Figure4.8)

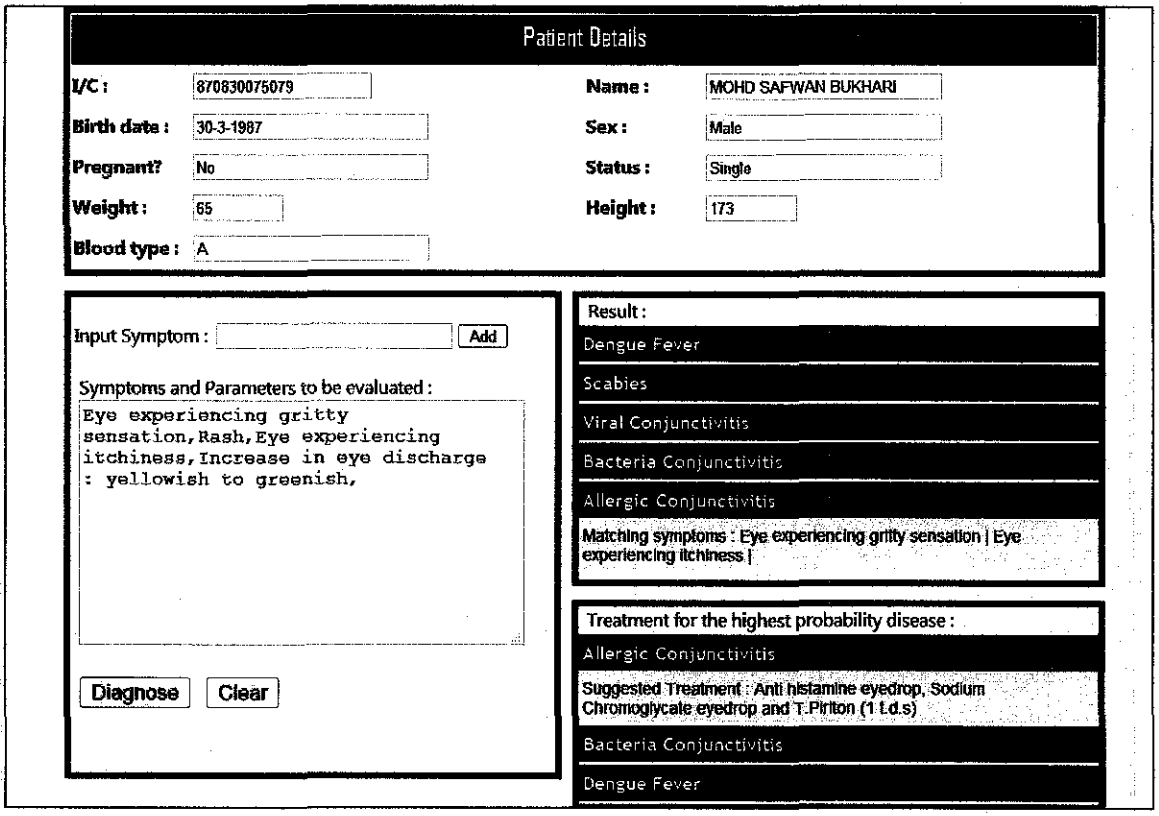


Figure 4.8: Diagnose Interface

4.3 TEST CASES

The test cases were done by consulting a doctor from UTP Clinic. The doctor gave a list of 5 diseases and the symptoms that would lead to that particular disease diagnosis. Then rules are configured according to the information acquired. The system is then tested by doing a series of test cases, where several symptoms are inserted and the diagnoses given by the system is compared to the final diagnosis that the doctor provided.

Test Case 1:

Configured Rules Involved: IF (*rash, severe itchiness at night, infected rash , excoriation mark*) THEN *scabies*
Symptoms Entered : Rash Severe itchiness at night Infected rash Excoriation marks Headache Fever
Final diagnosis provided by doctor : Scabies.
Diagnoses given by the system : Dengue Fever, Food Poisoning, Scabies
Was the expected final diagnosis given : Yes, Scabies.

Test Case 2:

Configured Rules Involved: IF (*vomiting ,abdominal pain or discomfort, fever ,diarrhea, nausea*) THEN *food poisoning*
Symptoms Entered : Headache Vomiting Abdominal pain or discomfort Fever Diarrhea Myalgia

Nausea
Final diagnosis provided by doctor : Food Poisoning.
Diagnoses given by the system : Dengue Fever, Food Poisoning
Was the expected final diagnosis given : Yes, Food Poisoning.

Test Case 3:

Rules Configured Involved: IF (*eye experiencing itchiness, eye experiencing redness, eye experiencing gritty sensation, warm and watery eye*) THEN *viral conjunctivitis*
Symptoms Entered : Eye experiencing gritty sensation Eye experiencing itchiness Eye experiencing redness Warm and watery eye Cough Fever
Final diagnosis provided by doctor : Viral Conjunctivitis.
Diagnoses given by the system : Food Poisoning, Viral Conjunctivitis, Bacterial Conjunctivitis, Allergic Conjunctivitis
Was the expected final diagnosis given : Yes, Viral Conjunctivitis.

Test Case 4:

Rules Configured Involved: IF (*eye experiencing itchiness, eye experiencing redness, eye experiencing gritty sensation, Increase in eye discharge : yellowish to greenish*) THEN *bacterial conjunctivitis*
Symptoms Entered : Eye experiencing gritty sensation Eye experiencing itchiness

Eye experiencing redness Increase in eye discharge : yellowish to greenish Rash Nausea
Final diagnosis provided by doctor : Bacterial Conjunctivitis.
Diagnoses given by the system : Food Poisoning, Viral Conjunctivitis, Bacterial Conjunctivitis, Allergic Conjunctivitis, Dengue Fever, Scabies
Was the expected final diagnosis given : Yes, Bacterial Conjunctivitis.

Test Case 5:

Rules Configured Involved:
IF (*eye experiencing itchiness, eye experiencing redness, eye experiencing gritty sensation, severe itching of the eye*) THEN *allergic conjunctivitis*
Symptoms Entered : Eye experiencing gritty sensation Eye experiencing itchiness Eye experiencing redness Severe itching of the eye Myalgia
Final diagnosis provided by doctor : Allergic Conjunctivitis.
Diagnoses given by the system : Viral Conjunctivitis, Bacterial Conjunctivitis, Allergic Conjunctivitis, Dengue Fever
Was the expected final diagnosis given : Yes, Allergic Conjunctivitis.

4.4 TEST CASES ANALYSIS

From the 5 test cases conducted, the system provides 19 possible diseases and 5 of the listed disease are as the same as the doctor's. For each particular case, the system managed to come out with a disease similar to the final diagnosis given by the doctor. So from the test the system accuracy is 100%. The accuracy of the system depends on the rules configured to it. By right, if the rules configured are accurate and credible, the system can provide the correct final diagnoses.

But the system also provided other disease (differential diagnoses) other than the final diagnose when displaying its result. The clinicians themselves must make decisions which are the most likely diagnose. The system act as a reminder by listing a number of possible diseases. The fewer the disease the system managed to list down, the better it is for the clinicians to make decision. Below is the pie chart that shows the distribution of the final diagnoses and the differential diagnoses in terms of percentage, provide by the system according to the test cases.

Table 4.1: Distribution of final diagnose and differential diagnose

Correct final diagnosis	5
Differential diagnosis	14
Total diagnosis	19

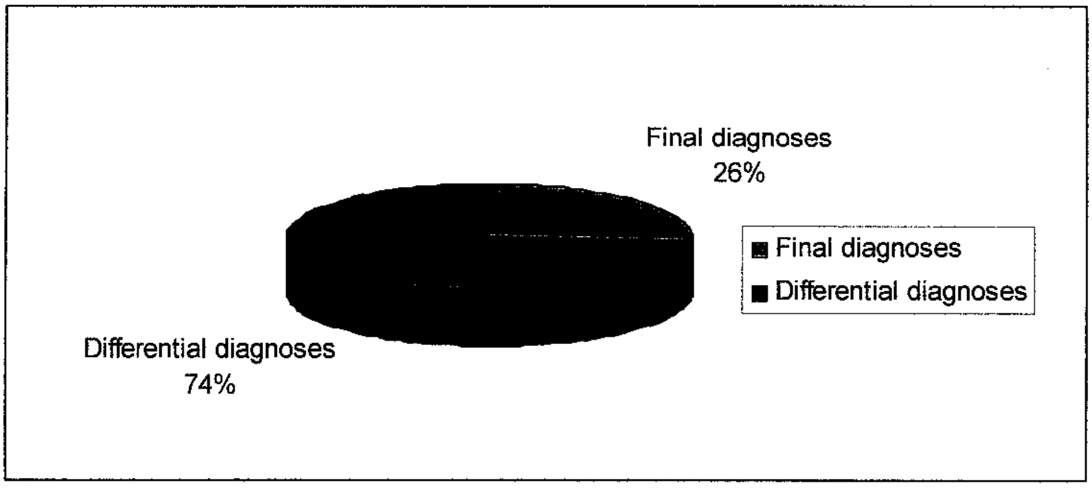


Figure 4.9: Pie Chart Showing The Distribution Of Final Diagnose And Differential Diagnose

CHAPTER 5
CONCLUSION AND RECOMMENDATION

5.1 Conclusion

The initial stage of this project had given the exposure of understanding of how to develop a knowledge-based system and rule-based reasoning. The next part is to capture the medical knowledge from either reading or by interviewing the experts. Using the rules based system allows the system to function immediately, as long as the rules have been stored in the system. The system intelligence does not need to be train. But in order for the system to be accurate, the rules must be developed with the experts of the field, in this case the clinicians.

5.2 Recommendation

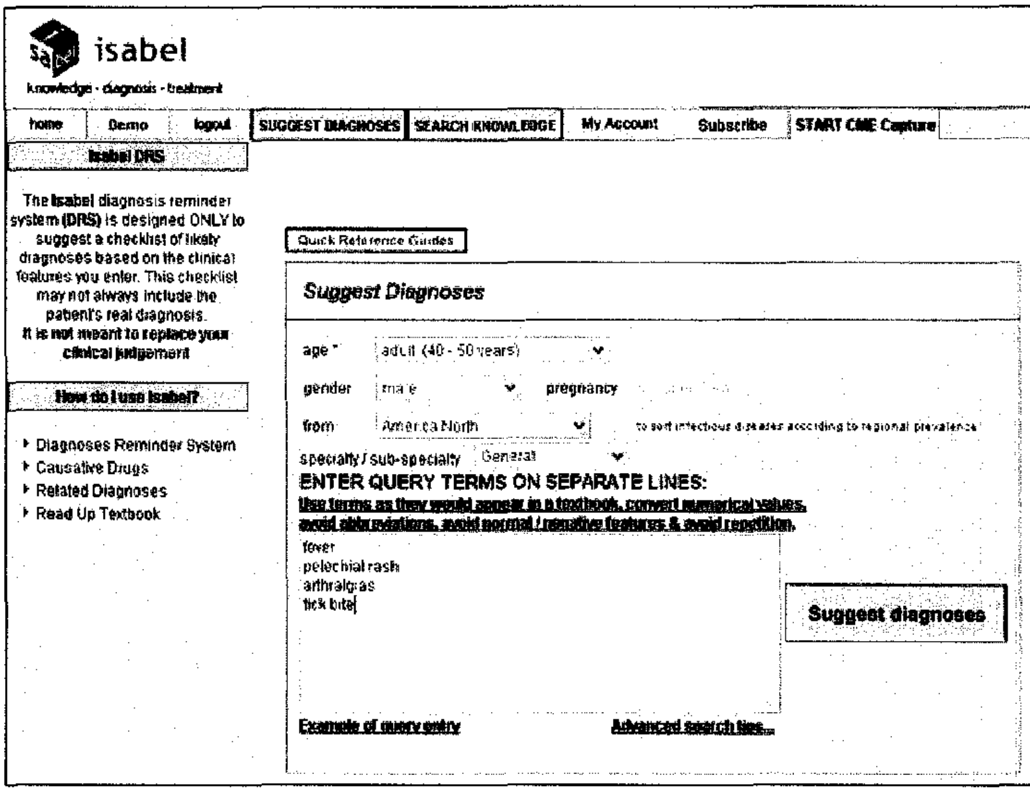
For the system itself, I would recommend two enhancements. First is to use multi modal reasoning method for the system's artificial intelligence. By using the multi modal reasoning method, it will combines both of the advantages of rule-based reasoning and case-based reasoning, thus making the system more accurate and credible. Secondly is to use natural text as the input method. The user will key in inputs in the textbox, each line will be assume as a symptom. The system will then find the most similar word in the system's library compared to the input that user provided for each line, and that word will be selected as a symptom to be use for predictions.

Besides using it for medical diagnosis, one possible proliferation of rule-based reasoning to expand is in the finance field. One example is on security trading and portfolio analysis. Rule-based can be used to evaluate rate return and risk exposure, cutting short the time-consuming portfolio composing process.

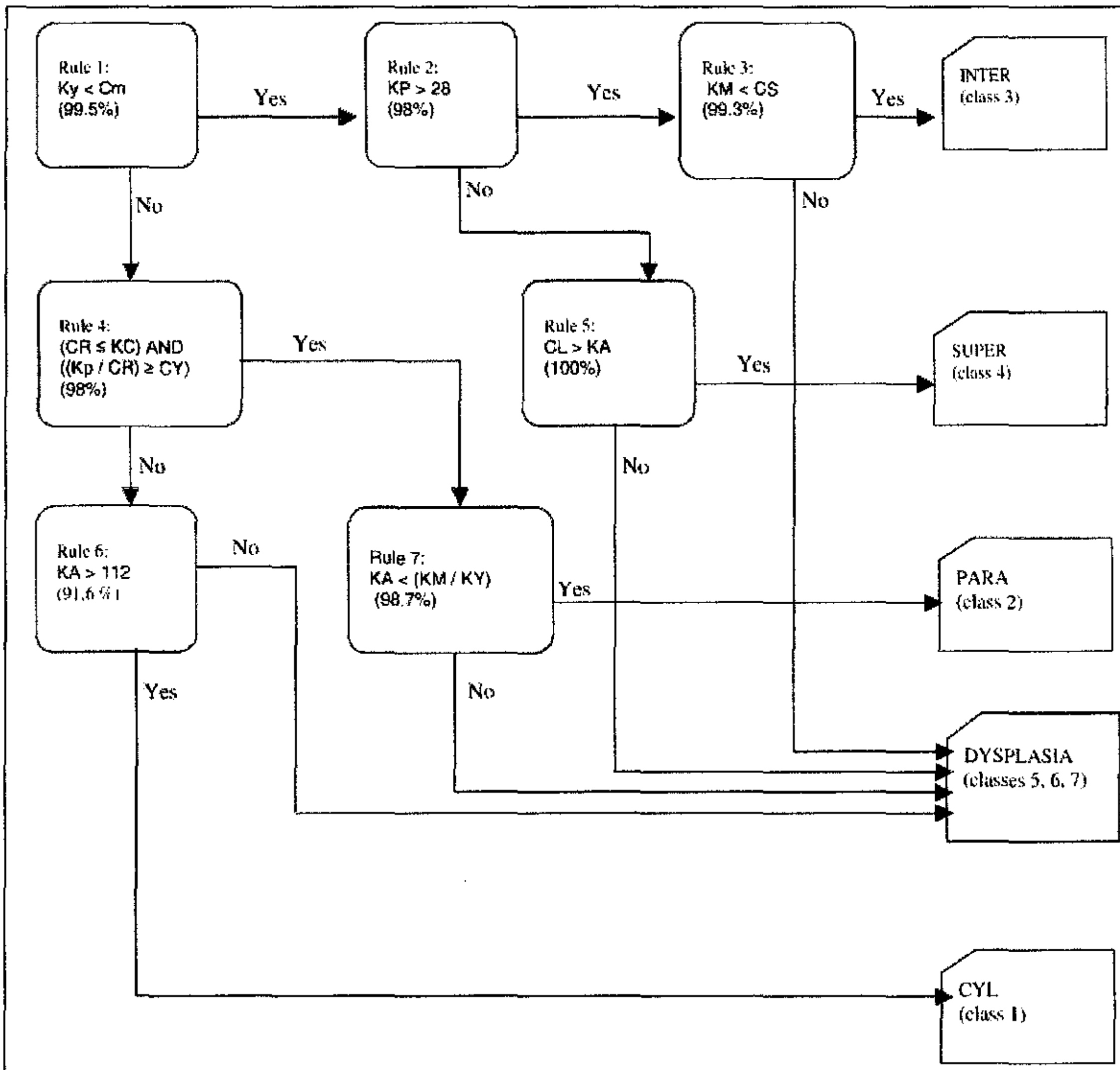
APPENDIX A: Interview Question

1. Do you use any computer system to assist you in diagnosis or disease detection? Yes? Can you explain about the system? No? Why?
2. What are your current way of detecting asthma and doing diagnosis on patient?
3. What do you with the data that that you receive from patient, either by asking, observing or checking?
4. Do you have any computer system to store the patient medical history? Yes? Can you explain about the system? No? Why?
5. Do you use your patient medical history in doing diagnosis?
6. How many input that you need to considered before doing a diagnosis?
7. What problem or inefficiency do you face in the current method of diagnosis?
8. If there is a system to assist you in diagnosis, what are the requirements that you need from the system?
    a. What input should the system take?
    b. What type or class of data should the system considered while doing analysis?
    c. What output do you expect the system to gives you?

APENPDIX B: ISABEL date entry screen



APPENDIX C: Rule flowchart for classification between different types of cervical cells



Legend:

Kerne_A nucleus area	Cyto_A cytoplasm area	K/C nucleus area / cytoplasm area	Kerne_Ycol nucleus intensity	Cyto_Ycol cytoplasm intensity	KerneShort nucleus shortest diameter	KerneLong nucleus longest diameter	KerneElong nucleus elongation	KerneRund nucleus roundness	CytoShort cytoplasm shortest diameter	CytoLong cytoplasm longest diameter	CytoElong cytoplasm elongation
KA	CA	KC	KY	CY	KS	KL	KE	KR	CS	CL	CE
CytoRund cytoplasm roundness	KernePer Nucleus perimeter	CytoPeri cytoplasm perimeter	KerneX nucleus x coordinate	KerneY nucleus y coordinate	CytoX cytoplasm x coordinate	CytoY cytoplasm y coordinate	KernePos nucleus position in cytoplasm	KerneMax nucleus maxima	KerneMin nucleus minima	CytoMax cytoplasm maxima	CytoMin cytoplasm minima
CR	KP	CP	Kx	Ky	Cx	Cy	Kp	KM	Km	CM	Cm

APPENDIX D: Gantt Chart

Activities / Week	1	2	3	4	5	6	7	8	9	10	11	12	13	14
PLANNING														
Selection of Project Topic	■	■	■											
Preliminary Research Work		■	■	■										
Submission of Preliminary Report														
Seminar 1(optional)					■	■	■							
Paper Research on Artificial Intelligence and Knowledge Based System					■	■	■							
Submission of Progress Report														
Seminar 2 (compulsory)									■	■	■	■		
ANALYSIS														
Interview with clinicians, develop system architecture and knowledge based structure								■	■	■	■	■		
Submission of Interim Report Final Draft														
Oral Presentation														

Activities / Week	15	16	17	18	19	20	21	22	23	24	25	26	27	28
DESIGN														
Developing prototype – ver. 1	■	■	■											
Submission of Progress Report 1														
Developing prototype – ver. 2				■	■	■	■							
Submission of Progress Report 2														
Seminar (compulsory)									■	■	■			
Developing prototype – ver. 3								■	■	■	■	■		
IMPLEMENTATION														
Poster Exhibition														
Submission of Dissertation (soft bound)														
Oral Presentation														
Submission of Project Dissertation (Hard Bound)														