

*Addis Ababa  
University*

*(Since 1950)*



**ADDIS ABABA UNIVERSITY  
SCHOOL OF GRADUATE STUDIES  
SCHOOL OF INFORMATION SCIENCE**

**A CASE BASED REASONING KNOWLEDGE BASED  
SYSTEM FOR HYPERTENSION MANAGEMENT**

**HENOK BEKELE**

**JULY 2011**

ADDIS ABABA UNIVERSITY  
SCHOOL OF GRADUATE STUDIES  
SCHOOL OF INFORMATION SCIENCE

A CASE BASED REASONING KNOWLEDGE BASED  
SYSTEM FOR HYPERTENSION MANAGEMENT

A Thesis Submitted to the School of Graduate Studies of Addis  
Ababa University in Partial Fulfillment of the Requirements for the  
Degree of Master of Science in Information Science

By

HENOK BEKELE

JULY 2011

ADDIS ABABA UNIVERSITY  
SCHOOL OF GRADUATE STUDIES  
SCHOOL OF INFORMATION SCIENCE

A CASE BASED REASONING KNOWLEDGE BASED  
SYSTEM FOR HYPERTENSION MANAGEMENT

By

HENOK BEKELE

Name and signature of Members of the Examining Board

<u>Name</u>	<u>Title</u>	<u>Signature</u>	<u>Date</u>
_____	Chairperson	_____	_____
_____	Advisor	_____	_____
_____	Examiner,	_____	_____

## **Acknowledgement**

First and foremost I would like to thank God and Holy Mother who made all things possible, and granted me success in my thesis work and entire journey.

I am heartily thankful to my advisor, Dr Gashaw Kebede, who patiently advised me with scholar guidance and constructive comments on my work throughout this endeavor, without him I could not reach to this stage. I would also like to acknowledge and extend my heartfelt gratitude to all staff of Brook Medical Service Plc and Bole 17 Health Center, particularly to Dr. Mekonnen Tibebu, Dr. Abebe Desta, Ato Ephrem Bililigne and Sister Amelework, for their unconditional support during the knowledge acquisition.

My deep appreciation and thanks extended to my family for their encouragement and patience with all my pursuit. Finally, I would like to show my gratitude to all my friends particularly to Yosef G/Selassie and Geletaw Sahle for their support.

# Table of contents

Acknowledgement .....	i
List of Tables .....	vi
List of Figures .....	vii
List of Acronyms .....	viii
Abstract .....	ix
CHAPTER ONE .....	1
1. INTRODUCTION .....	1
1.1. Background .....	1
1.2. Statement of the Problem .....	3
1.3. Objective of the Study .....	5
1.3.1. General Objective .....	5
1.3.2. Specific Objective .....	5
1.4. Methodology of the Study .....	6
1.4.1. Literature Review .....	6
1.4.2. Data Collection .....	6
1.4.3. Data Preparation .....	7
1.4.4. Design and Implementation Tool .....	8
1.4.5. Evaluation .....	8
1.5. Scope and Limitation of the Study .....	9
1.6. Significance of the Study .....	10
1.7. Organization of the Thesis .....	10
CHAPTER TWO .....	11
2. LITERATURE REVIEW .....	11
2.1. Artificial Intelligence .....	11
2.2. Knowledge Based Systems .....	12
2.2.1. Architecture of Knowledge Based System .....	12
2.2.2. Knowledge Based System Advantages and Limitations .....	13
2.2.3. Knowledge Based System Development .....	14
2.2.4. Knowledge Elicitation .....	15

2.3.	Case Based Reasoning .....	17
2.3.1.	Cases .....	18
2.3.2.	Styles of Case Based Reasoning .....	19
2.3.3.	Case Based Reasoning Cycle.....	20
2.3.3.1.	Retrieve.....	21
2.3.3.2.	Reuse .....	24
2.3.3.3.	Revise .....	25
2.3.3.4.	Retain.....	26
2.4.	Rule based reasoning.....	26
2.4.1.	Rules .....	27
2.4.2.	Rule Based Reasoning Technique .....	27
2.4.2.1.	Data-Driven Control .....	27
2.4.2.2.	Goal-Driven Control.....	28
2.5.	Case based Vs Rule based reasoning .....	28
2.5.1.	In Knowledge Representation.....	28
2.5.2.	In Handling Missing and Uncertain Knowledge .....	29
2.5.3.	In Problem Solving .....	30
2.6.	Case Based Reasoning Tools .....	30
2.7.	Approaches to Evaluate the Performance of CBR system.....	31
2.8.	Knowledge Based System in Medical Domain.....	32
2.8.1.	Rule Based Reasoning in Medicine .....	33
2.8.2.	Case Based Reasoning in Medicine .....	35
CHAPTER THREE .....		37
3.	KNOWLEDGE ACQUISITION AND MODELLING.....	37
3.1.	Knowledge Acquisition to Identify Case feature .....	37
3.1.1.	Knowledge Acquisition from Relevant Document.....	38
3.1.2.	Knowledge Acquisition from Domain Expert .....	38
3.1.3.	Knowledge Acquisition from Hypertension Cases.....	39
3.2.	Knowledge Modeling.....	39
3.2.1.	Case Concept for Hypertension Management .....	40
3.2.1.1.	Hypertension Complaint's Profile .....	42
3.2.1.2.	Sex of Hypertension Compliant .....	43
3.2.1.3.	Sign and Symptom of Hypertension.....	43
3.2.1.4.	Blood Pressure of Hypertension Compliant .....	45

3.2.1.5.	Risk Factors for Hypertension.....	47
3.2.1.6.	Clinician Decision for Hypertension Management .....	48
3.2.2.	Goal of Antihypertensive Treatment .....	49
3.3.	Hypertension Case Structure.....	50
CHAPTER FOUR.....		54
4.	IMPLEMENTATION.....	54
4.1.	Designing of the Prototype.....	54
4.2.	A CBR Framework for Hypertension Management by Using Nearest Neighbor Retrieval Algorithm .....	55
4.3.	Designing the Retrieval Process.....	57
4.3.1.	Similarity Measurement.....	57
4.3.1.1.	Local Similarity Measurement .....	57
4.3.1.2.	Global Similarity Measurement.....	60
4.3.2.	Ranking Relevant Cases .....	60
4.4.	Designing the Reuse Process .....	61
4.5.	Designing the learning mechanism for prototype .....	61
CHAPTER FIVE .....		63
5.	TESTING AND EVALUATION .....	63
5.1.	Testing and Evaluation of the Prototype .....	63
5.2.	Case Similarity Testing .....	63
5.3.	Evaluation of the Retrieval and Reuse process by Using Statistical Analysis... ..	65
5.3.1.	Evaluation of the Retrieval Process .....	65
5.3.2.	Evaluation of the Reuse Process .....	68
5.3.3.	Comparison of the Prototype Performance with Previous Case Based Reasoning System.....	68
5.4.	Testing the learning mechanism.....	70
5.5.	User Evaluation .....	72
5.6.	Discussion .....	75
5.7.	Sample of the Experiment .....	77
CHAPTER SIX.....		79
6.	CONCLUSION AND RECOMMENDATIONS .....	79
6.1.	Conclusion.....	79
6.2.	Recommendations .....	81

References.....	82
Appendixes .....	90
Appendix I: Interview Questions .....	90
Appendix II: Domain Expert Evaluation Form .....	91
Appendix III: Query Form for the Prototype .....	92
Appendix IV: Sample of Retrieved Cases by the Prototype for Query .....	93
Appendix V: Sample of Proposed Solution for Query .....	94
Appendix VI: Sample of Python Source Code .....	95

## List of Tables

Table 2.1 Comparison of CBR and RBR in terms of Knowledge Representation .....	29
Table 3.1 SBP Classification in Adults.....	45
Table 3.2 DBP Classification in Adults.....	46
Table 3.3 Case Structure for Hypertension Management.....	51
Table 4.1 Attributes Description of the Case Structure .....	58
Table 5.1 Sample of Query for Case Similarity Testing and their Corresponding Cases .....	64
Table 5.2 Query Similarity with their Corresponding Cases from the Case Base.....	64
Table 5.3 Relevant Cases Assigned by Domain Expert for the Sample Test Case .....	66
Table 5.4 Recall and Precision Result for the Sample Test Case .....	67
Table 5.5 Average Recall and Precision Value of the Retrieval process.....	67
Table 5.6 Accuracy Value of the Reuse Process .....	68
Table 5.7 Comparison of the Prototype with the Previous CBR system.....	69
Table 5.8 Performance Evaluation by Group one Domain Experts.....	73
Table 5.9 Performance Evaluation by Group two Domain Experts .....	74

## List of Figures

Figure 2.1 Architecture of a Knowledge-Based System .....	12
Figure 2.2 Development of a Knowledge-Based System .....	14
Figure 2.3 The Case Based Reasoning Cycle .....	21
Figure 2.4 A Nearest Neighbor Algorithm .....	23
Figure 3.1 General Concept for Managing Hypertension Case.....	41
Figure 3.2 Concept for Hypertension Compliant Profile.....	42
Figure 3.3 Concept for Sex of the Compliant.....	43
Figure 3.4 Sign and Symptom of Hypertension Patients.....	44
Figure 3.5 Concepts of Blood Pressure of Compliant .....	46
Figure 3.6 Risk Factor of Hypertension.....	47
Figure 3.7 Concept for Clinician Decision for Hypertension Case .....	48
Figure 3.8 Goal of Antihypertensive treatment .....	49
Figure 4.1 Architecture of Case Based Reasoning for Hypertension Management .....	56
Figure 5.1 Revise Process for the newly Solved Case.....	71
Figure 5.2 Retaining Process of the newly Solved Case .....	71

## **List of Acronyms**

**ADIS**- Acquired Immunodeficiency Syndrome

**AI**- Artificial Intelligence

**CBR**-Case Based Reasoning

**DBP**-Diastolic Blood Pressure

**HIV**- Human Immunodeficiency Virus

**JCOLIBRI**- Java Class Ontology Libraries Integration for Building Reasoning

Infrastructure

**KBS**-Knowledge Based System

**RBR**-Rule Based Reasoning

**SBP**-Systolic Blood Pressure

## **Abstract**

Hypertension is a growing public health problem. In this paper, the potential of case based reasoning approach for hypertension management have been investigated. In order to conduct the research, the required knowledge for the study have been collected from hypertension compliant card histories, domain experts and other relevant documents through semi-structured interview and document analysis methods of knowledge elicitation. Then the knowledge is modeled in hierarchical tree manner and case structure of the case base is constructed. Forty five hypertension cases are collected from Brook Medical Service Plc and Bole 17 Health Center to construct the case base.

The case based reasoning prototype for hypertension management is implemented by using python 2.6. Nearest Neighbor retrieval algorithm, voting method, domain expert feedback and incremental learning are used for the retrieval, reuse, revise and retain tasks of the prototype respectively. The collected hypertension cases are represented in the form of Feature-vector case representation approaches.

The prototype is evaluated by using both statistical analysis and user evaluation. The statistical analysis uses leave-one-out cross validation testing proportion for both the retrieval and reuse processes. The retrieval performance of the prototype shows average value of 86.1% recall and 60% precision, while the performance of the reuse process shows an average value of 88.89% accuracy. The over all performance of the prototype as it is evaluated by domain experts is 3.99 out of 5. Given all these results, the performance of the prototype is promising.

All in all the study achieves its objective by developing the prototype with promising performance and user acceptance, and demonstrating case based reasoning approach in designing knowledge based system for hypertension management.

# CHAPTER ONE

## INTRODUCTION

### 1.1. Background

A state of elevated systemic blood pressure, which is commonly asymptomatic, is called Hypertension. Hypertension puts a strain on the heart by increasing its need for oxygen, it makes the heart to work harder than normal. Over time this also causes the walls of the arteries harden. People with high blood pressure often do not feel sick. Since hypertension may cause no symptoms at all for a long time, sometimes it is referred to as “the silent killer”. Patients’ organs and tissues can be damaged by the diseases without showing any external symptoms. It is more common in men than women, and more common in older people than younger people (STGLGH, 2010; SOH, 2011).

There are two main types of hypertension known as primary or essential hypertension and a secondary hypertension. Primary or essential hypertension is the one that is difficult to identify its causes; however the causes are easily identifiable in secondary hypertension (SOH, 2011).

Hypertension plays a major role in the development of other diseases such as ischemic heart disease, cardiac and renal failure. As a result of this, treating hypertension has been associated with about a 40% reduction in the risk of stroke. Even though the treatment of hypertension has shown to prevent different kind of cardiovascular diseases and to extend and enhance life, hypertension remains inadequately managed everywhere (Whitworth, 2003).

Hypertension is an important public-health challenge in the world due to increasing longevity and prevalence of contributing factors such as overweight and obesity, cigarette smoking, physical inactivity, unhealthy diet, stress, dietary salt intake, and alcohol use (Whitworth, 2003; Tesfaye et al, 2009). While the current prevalence of hypertension seems constant in developed countries, in many developing countries it has been increased dramatically, particularly in urban societies (Whitworth, 2003; Tesfaye et al,

2009). However a research indicate that in 2003, out of 167 countries, 61% of the countries don't have national hypertension guidelines and 45% of the countries' health professionals were not trained to manage hypertension (Whitworth, 2003).

Integrating computer technology and artificial intelligence into health services is one of the approaches for addressing shortage of qualified health professional, experts, advisers and trainers on the area. Application of knowledge based systems is one of the mechanisms that improve health service qualities (Sajja and Akerkar, 2010).

The concept of knowledge based systems is derived from the field of artificial intelligence (AI). AI intends understanding of human intelligence and building of computer programs that are capable of simulating or acting one or more of intelligent behaviors (Tan, 2008; Sajja and Akerkar , 2010).

Knowledge based system are system which try to solve problem in a human expert like fashion by using knowledge of application and problem solving technique. Knowledge based systems are becoming a vital application as aids to clinical decision making across the medical spectrum (Carson and Cramp, 2002). Every knowledge based system has two building blocks which are known as knowledge Base and inference engine (Sajja and Akerkar, 2010).

There are different ways of representing knowledge in the knowledge base; the two of such techniques are cases and rules (Salem, 2007; Schmidt et al, 2001). Rules represent general knowledge of the domain, whereas cases represent specific knowledge. Rule based systems solve problems from scratch, while case based systems use pre-stored situations to deal with similar new instances.

Inference engine, on the other hand, represents the reasoning technique that manipulates, uses and controls the knowledge to solve the problems. It infers the knowledge available in the knowledge base. Case based reasoning and rule based reasoning are two examples of reasoning techniques (Sajja and Akerkar, 2010).

Case based reasoning methodology provides a foundation for a new technology of building intelligent computer aided diagnoses systems. It addresses the challenges that are found in the rule-based technology as follows (Salem, 2007).

- It provide adaptation solution by reasoning from analogy of past cases,
- It supports scalability by allowing adding knowledge

## **1.2. Statement of the Problem**

Hypertension is a growing public health problem which greatly contributes to cardiovascular diseases. A research show that in 2000 over 25% of the world's adult population had hypertension and the proportion is expected to increase to 29% by 2025 (Getahun et al, 2010). In 2000 the estimated total number of people with hypertension was 972 million, among which 333 million people found in economically developed countries, and the rest 639 million patients found in economically developing countries. The number of people with hypertension in economically developed countries is predicted to increase by 24% from 333 million to 413 million, while a rise of 80% is predicted for economically developing countries from 639 million to 1.15 billion in 2025 (Lekoubou et al, 2010).

In Ethiopia, by the year 2000/01 hypertension accounted for 1.4 percent of all deaths reported to Federal Ministry of Health of Ethiopia. It was the 7th leading cause of death in the country (Getahun et al, 2010). The use or misuse of addictive substances, such as cigarettes, alcohol, and khat (*Catha edulis* Forsk) which are the primary factors for hypertension are increasing in the country. Due to this and other factors hypertension is widely spread in the country and may represent a silent epidemic in the population (Tesfaye et al, 2009).

While the burden of chronic diseases in sub-Saharan Africa is very high, the current density and distribution of health workforce suggest that Sub-Saharan Africa cannot respond to the growing demand for chronic disease care (Lekoubou et al, 2010). In Ethiopia the number of physicians and nurses for 100,000 peoples are 3 and 6 respectively which clearly shows the shortage of qualified human power (Lekoubou et al,

2010). A research shows from 167 countries, 45% health professionals were not trained to manage hypertension (Whitworth, 2003).

Even though the prevalence of hypertension is high, and its treatment also associated in reduction of different kind of cardiovascular diseases, hypertension remains inadequately managed everywhere (Whitworth, 2003).

In Ethiopia, there are different studies that are conducted to investigate the applicability of knowledge based system in supporting medical and other service but some of them uses a rule based reasoning techniques. For instance in health area, Anteneh (2004) attempt to design a prototype knowledge based system for antiretroviral therapy by using rule based reasoning technique; Rediet (2006) developed a prototype of knowledge based system for HIV pretest counseling by using rule based reasoning. Even though Rule Based Reasoning technique has some advantage in developing knowledge based system, it has drawback when it is applied to medical domain. The one of the main limitations is that knowledge acquisition process in medical domain is challenging and it is difficult to represent in all the knowledge in the form of rule.

Alemu (2010) tried to design a knowledge based system for AIDS resource center by using a case based reasoning approach. But his work focus on the retrieval task of the case based reasoning model. In addition to these, Yemsirach (2010) tried to investigate the case based reasoning approach in designing legal knowledge based system. However, as to the knowledge of the researcher there is no attempt that tried to investigate the case based reasoning approach in designing knowledge based system for hypertension management.

This study is conducted with the aim of filling the gap which is stated in the above section by investigating the case based reasoning approach in designing knowledge based system that can diagnosis and treat hypertension cases. As knowledge based systems are useful when there is a shortage of experts, and when intelligent assistance or training are required for the decision making (Sajja and Akerkar, 2010).

### **1.3. Objective of the Study**

The study has the following general and specific objectives

#### **1.3.1. General Objective**

The main objective of this study is to investigate the applicability of case based reasoning approach in development of knowledge based system for hypertension management, i.e. to investigate the applicability of the approach in providing relevant cases and proposing solution to the new hypertension cases form already solved case.

#### **1.3.2. Specific Objective**

To achieve the above general objective of the study, the researcher set the following specific objectives:

- To review literature in order to have an understanding on concepts, principles and technologies of knowledge based system
- To make interview and discussion with domain experts, and to analyze documents so as to have a required knowledge for building prototype knowledge based system for hypertension management
- To model the required knowledge for hypertension management
- To develop a case structure consisting of necessary attributes that have a direct impact on hypertension management
- To build the case base for the hypertension management prototype from hypertension compliant cases
- To develop a prototype knowledge based system for hypertension management by using case base reasoning technique
- To test and evaluate the prototype system
- To make conclusions and recommendation for further research

## **1.4. Methodology of the Study**

The following method and technique are employed for the study, in order to achieve the general and specific objective of the research.

### **1.4.1. Literature Review**

Extensive literature reviews from different books, journals articles, thesis and the Internet are conducted, so as to have a solid and concrete understanding on principles, techniques and tools of knowledge based system with a special emphasis on case based reasoning. Furthermore, researches that are conducted on knowledge based system in medical domain and other related work are reviewed.

### **1.4.2. Data Collection**

Both primary and secondary data collection methods are employed to collect the required domain knowledge for the study. As primary sources, health professionals from Brook Medical Service Plc and Bole 17 Health Center are interviewed. In addition, relevant literature from all possible sources, including journal articles, guideline for hypertension case management, hypertension related books, thesis, the Internet and related websites are reviewed as secondary sources.

Six domain experts are selected by using purposive sampling techniques. This technique is selected by the researcher as the research need in-depth investigation on hypertension management. Purposive sampling techniques enable to select sample which can provide the needed information (Tan, 2008). Domain experts are selected based on their educational qualifications related to the domain area, year of experience and willingness. From the six experts, two of them are medical doctors who specialized in internist and having more than twenty years of experience. Three of them are health officers with more than 3 years experience. The remaining one is a nurse with BSc degree having more than 10 years of experience.

A semi-structured interview is conducted with the selected health professionals in order to acquire the necessary knowledge for the study. The main reason that the researcher

used a semi-structured interview compare to other type of interview is that semi-structured interview guide interviewer by providing both types of closed-ended and open-ended questions. It allows the interviewer to change the order of the questions and add new questions based on the context of the participant response so as to get depth knowledge.

The interview focuses on the concepts, which the health professionals focus on, during hypertension management. The main semi-structured questions that are used in the interview is attached on Appendix I. Making interviews with domain experts in work time are likely to be interrupted, as the domain experts are usually busy on work time. As a result of this, experts are interviewed after work hours and on weekends.

After the necessary attributes for hypertension management are identified, the concepts are modeled by using hierarchical tree. Then the case structure that comprise variables, which have direct impact on hypertension management, is constructed. Finally the hypertension cases, which are used to build the case base, are collected from Brook Medical Service Plc and Bole 17 Health Center. The cases are collected by the researcher with the help of domain experts from hypertension compliant history cards.

### **1.4.3. Data Preparation**

In this step the collected hypertension cases are arranged into a form that is suitable for the selected retrieval algorithm and case based reasoning tools. A hypertension cases that have noisy data are removed from the collected hypertension cases. The reason for removing these cases is to avoid uncertain result on the study. After cases that have noisy data are removed from the collected hypertension cases, the collected cases are represented in a Feature-vector approach to develop the case base. This approach uses attribute-value pair format. The main reason for using this kind of case representation is that it represents the case in simple and clear manner. Feature-vector approach allows using Nearest Neighbor algorithm (Bergmann et al, 2005). A total of 45 cases have been used to build the case base.

#### **1.4.4. Design and Implementation Tool**

In designing the case based reasoning system prototype for hypertension management, the Nearest Neighbor retrieval algorithm and voting method are used for retrieval and reuse task respectively. The main reason for using Nearest Neighbor algorithm is that it retrieves cases which match partially with the new case (Salem et al, 2005; Martin, 1995). According to Mishra and Sahu (2011), Nearest Neighbor algorithm has the advantage of simplicity in retrieving relevant cases. It is also suitable when there are attributes that have numeric value (fang and Songdong, 2007).

A voting method is used for the reuse process. This method is tested by Salem et al (2005) on their research for supporting diagnosis of heart diseases and it proposes solutions correctly. The revise process is designed by using a way of getting domain experts feedback for the proposed solution based on its consequence for the solved case. Incremental learning is used to design the retain process of the case based reasoning prototype.

The tool that is used to implement the case based reasoning prototype for hypertension management is python 2.6.4. Python is open source software (Ascher, 2002). The reason that python is used in the study is that the researcher is familiar with python programming and python is easily accessible to the researcher as compared to other tools used to design case based reasoning system.

#### **1.4.5. Evaluation**

The evaluation of the case based reasoning system in this research is conducted through statistical analysis, learning testing and user evaluation. The statistical analysis is done for both retrieval and reuse process of the case based reasoning prototype.

The statistical analysis uses the 45 cases that make up the case base as training and testing data. The statistical evaluation uses leave-one-out cross validation testing proportion, where each case in turn is left out, and the learning method is trained on all the remaining cases, i.e. the evaluation is done for all cases by making one of the cases as a testing data and the rest of the cases as a training data (case base). The main reason that

the researcher uses leave-one-out cross validation is that it is common evaluation strategy in case based reasoning (Jagannathan, et al, 2010 ) and it provide almost unbiased estimate of generalization performance (Cawley and Talbot, 2008).

The researcher conduct 45 experiments for both retrieval and reuse task of the case based reasoning system. The performance of the retrieval task of the case based reasoning system is measured by using recall and precision, and the performance of the reuse task is measured by using accuracy.

The learning testing is conducted to test the learning mechanism of the prototype from solved hypertension cases for future use. Finally, user evaluation has been conducted by group of domain experts. The domain experts prepare query (new cases) and they feed it to the system. After they look up the result of the system for the query, they evaluate the performance of the system.

## **1.5. Scope and Limitation of the Study**

Even though there are different approaches to design knowledge based system, the study focus only on case based reasoning approach. The study is limited to develop a prototype knowledge based system for the purpose of hypertension management. The prototype is limited to diagnosis and recommends appropriate kind of hypertension treatment (only life style modification, first line drug, urgency drug or an emergency treatment) for hypertension cases. But it doesn't deal with the management of drug response to hypertension patient who starts antihypertensive drug treatment.

As the cases for the case base of the prototype are collected from hypertension complaints' record files, the cases structure for the prototype is made up of attributes that are recorded in the complaints' record files. Due to these attributes like weight, race and height which are considered in hypertension management are not included in the case structure for the prototype.

## **1.6. Significance of the Study**

Knowledge based system try to solve problem in a human expert like fashion by using knowledge of application(expert) and problem solving technique. Thus the study can benefit the country and the society by supporting activities which can be conducted to improve hypertension management. The study can help health professionals by providing them with a quick reference for hypertension complaints' cases. The output of research can also help medical students when they are being trained on hypertension management. This research can serve as a base for future researchers in the area.

## **1.7. Organization of the Thesis**

The thesis is organized in to six chapters. The first chapter present background information about hypertension, knowledge based system, statement of the problem, objective of the study, scope of the study and the methodology employed in the study. The second chapter deal with the literature review which discuss about artificial intelligence, knowledge based system, knowledge elicitation technique, cases based reasoning, case based reasoning tools and rule based reasoning. This chapter also discusses about the evaluation technique for case based reasoning system. Reviews on related research work in health are included in the chapter.

Chapter three deal with knowledge acquisition and modeling. It discusses about how the required knowledge is collected from hypertension cases, domain experts and other relevant documents. It presents models that show what things are considered in the hypertension management. This chapter also presents the case structure framework that is used for purpose of building the case base.

Chapter four deal with implementation of the case based reasoning system. It presents the framework of the case based reasoning prototype and shows how the system is implemented by using python. Chapter five presents the evaluation result of the system and discussion on the result. The last chapter summarizes the main points presented in this thesis and provides future research lines.

# CHAPTER TWO

## LITERATURE REVIEW

### 2.1. Artificial Intelligence

The field of Artificial intelligence (AI) tries to understand how we think, that is how we can perceive, understand, predict, and manipulate a world. AI also tries to develop intelligent entities that can handle problem in the same ways as humans do (Russell and Norvig, 2003).

The name of AI was coined in 1956 (Detore, 1989). Various definitions for AI are proposed. Russell and Norvig (2003) try to classify the definitions into two dimensions

- i. definitions that are concerned with thought processes and reasoning, these are
  - “The exciting new effort to make computers think ... machines with minds, in the full and literal sense”
  - “The study of the computations that make it possible to perceive, reason, and act”
- ii. definitions that address behavior
  - “the art of creating machines that perform functions that require intelligence when performed by people”
  - “AI... is concerned with intelligent behavior in artifacts”

AI started with a goal to replicate human level intelligence in a machine (Brooks, 1991). The field of Artificial intelligence adopts different ideas, viewpoints and techniques from different disciplines such as Philosophy, Mathematics, Economics, Neuroscience, Psychology, Computer engineering, Control theory and cybernetics, and linguistics (Fogel, 2006; Russell and Norvig, 2003)

AI has different branches. Some of the branches are general purposes that deal with learning and perception, and others are specific task such as playing chess, providing mathematical theorems and diagnosing diseases in medical area (Binjie, 2010; Russell and Norvig, 2003).

## 2.2. Knowledge Based Systems

The concept of knowledge based systems is derived from the field of AI . AI intends understanding of human intelligence and building of computer programs that are capable of simulating or acting one or more of intelligent behaviors. Intelligent behaviors include cognitive skills like thinking, problem solving, learning, understanding, emotions, consciousness, intuition and creativity, language capacity, etc. These days some of the behaviors such as problem solving, learning and understanding are handled by computer programs (Sajja and Akerkar, 2010; Tan,2008)

Computer programs that try to solve problems in a human expert-like fashion by using knowledge about the application domain and problem solving techniques are known as Knowledge based system (KBS) (Sajja and Akerkar, 2010; Speel et al, 2001). Human experts use the knowledge they have about the domain and techniques that lead how to use the knowledge to solve problems. Knowledge based systems handle problems in the same way. They represent the knowledge about the application domain and they use one or more techniques that guides on how to use the knowledge to solve problems (Sajja and Akerkar, 2010).

### 2.2.1. Architecture of Knowledge Based System

As shown in figure 2.1 below, every KBS has two building blocks. These are: (Sajja and Akerkar, 2010; Speel et al, 2001)

- knowledge base and
- inference engine

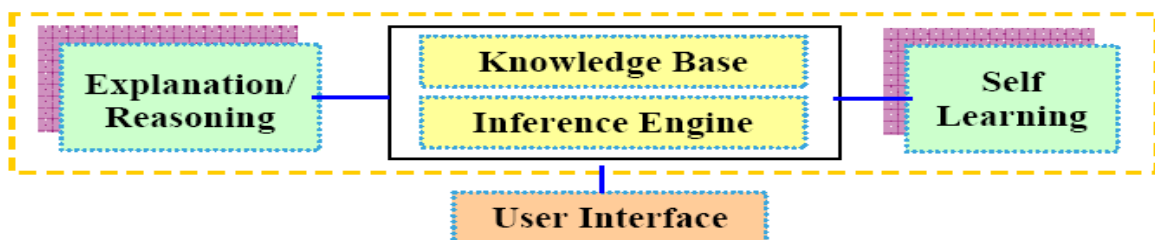


Figure 2.1 Architecture of a Knowledge-Based System (Sajja and Akerkar, 2010;)

Knowledge base contains all necessary knowledge about the domain that is required to handle problems. The knowledge can be acquired from experts, documents, books and/or other sources. It is formalized and organized with a technique called knowledge representation. There are several techniques to represent knowledge in the knowledge base. Representing knowledge in form of rules and cases are examples of the techniques (Salem, 2007; Schmidt et al, 2001; Montani et al, 1991), it will be discussed later.

The second component of a KBS is inference engine. After the system gets the required knowledge, it needs to be instructed how to use the knowledge in solving problems. Inference engine represents the reasoning technique that manipulates, uses and controls the knowledge to solve problems. Case based reasoning and rule based reasoning are two examples of reasoning techniques (Sajja and Akerkar, 2010). This will also be discussed later

### **2.2.2. Knowledge Based System Advantages and Limitations**

Knowledge based systems allow documentation of expert's knowledge as well as utilization of the knowledge for problem solving purpose in cost effective way. It also increases the quality of decision making process. It has strong potential in reducing training cost in different areas that need decision making and it has strong ability in maintaining consistent expert knowledge (Sajja and Akerkar, 2010; Tan, 2008).

KBS are more advantageous than traditional computer information systems in many ways. Specially:

- when there is shortage of expert,
- when decision making for problem solving needs intelligent assistant
- When expertise is needed to be stored for future use.

However Knowledge Base Systems have some major limitations due to the following main reasons: (Sajja and Akerkar, 2010)

- Abstract nature of the knowledge.
- Limitations of cognitive science and other scientific methods.

Due to this and other factors acquisition, representation and manipulation of the large volume of the knowledge is the major problem.

### 2.2.3. Knowledge Based System Development

The development of knowledge base system has many processes and its general view is depicted in figure 2.2.

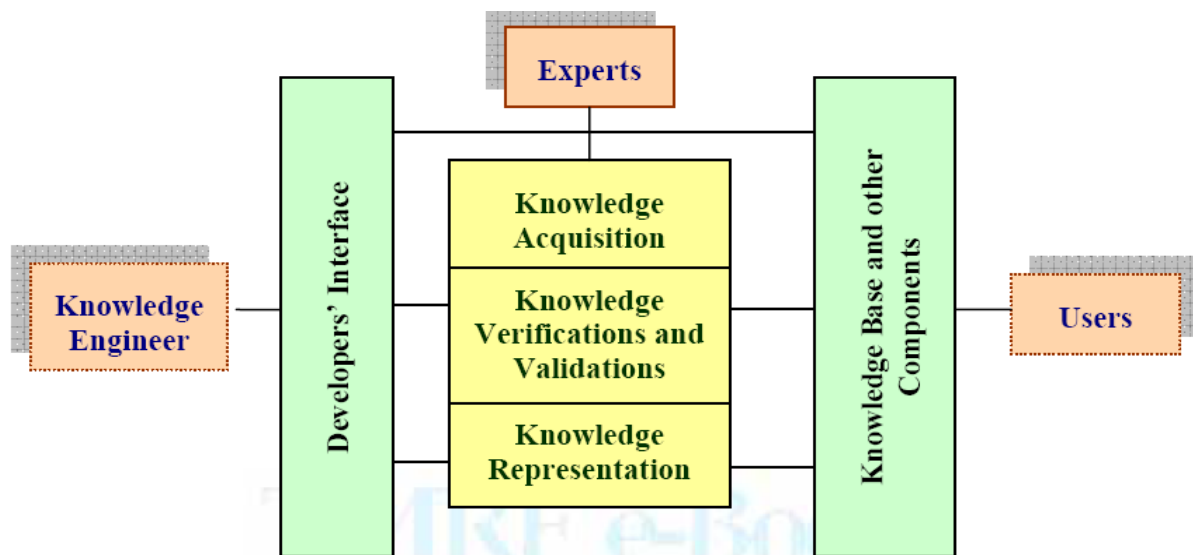


Figure 2.2 Development of a Knowledge-Based System (Sajja and Akerkar, 2010)

Expert is the one who has knowledge in his mind and the knowledge is stored in abstract way. The Knowledge Engineer is responsible person to acquire, transfer and represent the experts' knowledge in form of computer system.

Knowledge acquisition is the general term used for the process of developing a computational problem-solving model (Clancey, 1984). It includes interviews,

questionnaires, record reviews and observation to acquire factual and explicit knowledge. This process through which knowledge engineer gets knowledge from the domain expert is called knowledge elicitation. It will be discussed in next section.

The knowledge which is captured through knowledge acquisition process should be documented in a knowledge representation method. The knowledge engineer has the responsibility to select appropriate knowledge presentation scheme and inference engine that is natural, efficient, transparent, and developer friendly. Then the captured knowledge will be stored in the knowledge bases (Sajja and Akerkar, 2010; Tan, 2008).

#### **2.2.4. Knowledge Elicitation**

Knowledge elicitation is the process through which knowledge engineer get knowledge form the domain expert. Knowledge elicitation allows to obtain knowledge from Explicit knowledge such as books, manuals... and from Tacit knowledge by using different kind of technique (Shadrick et al, 2005; Burge, 1998). There are many different knowledge elicitation techniques; the most common are (Burge, 1998):

- Interview
- Observation
- Document Analysis

##### **Interview**

Interview technique involves asking the domain experts on how they perform their tasks based on their knowledge. Based on its structure, interview is classified as unstructured, semi-structured and structured interview (Burge, 1998).

A structured interview is an interview which has a set of predefined questions that will be answered by the expert. The order of the questions is also predefined. This is type of interview allows to minimize the effects of interviewer on the research results (Clark et al, 2008).

A semi-structured interview is an interview which has a guide that usually includes both types of closed-ended and open-ended questions. It is more flexible than structured one. In this kind of interview the interviewer has a chance to change the order of questions to be asked and to add questions based on the context of the participants' responses.

Unstructured interview is a kind of interview that doesn't predetermine both the questions and the answer categories. But it is not random and non-directive. It basically rely on the interaction between the researcher and the expert (Clark et al, 2008).

The questions that are used in the interview and the ability of the experts to express (verbalize) their knowledge are the factors for the success of an interview (Shadrick et al, 2005).

### **Observation**

Observation is one of the most popular and most used techniques of knowledge elicitation. In this technique, the knowledge engineer observes the expert at work, trying to understand and duplicate the expert's problem-solving methods. This technique prevents the knowledge engineer from interfering in the process; however it does not provide any insight into why decisions are made (Clark et al, 2008; Burge, 1998).

### **Document Analysis**

This technique collects information from existing documents. These documents include promotional literature, brochures, manuals, employee handbooks, reports, glossaries, course texts, and existing training materials (Clark et al, 2008; Burge, 1998).

Knowledge Elicitation methods can be classified in many ways. Direct and Indirect is one of the most common ways of categorizing knowledge elicitation. This way of classification is by how directly knowledge engineers obtain information from the domain expert. Direct methods involve directly questioning a domain expert on how they do their job. In order for direct methods to be successful, the domain expert has to reasonably articulate and willing to share information. Interviewing is an example of this

type of knowledge elicitation. However in Indirect methods the needed information is not requested directly. Instead, the results of the knowledge elicitation session must be analyzed in order to extract the needed information. Indirect methods are thought to be more suitable when knowledge is not easily expressed by the expert. Role Playing is an example of indirect methods (Shadrick et al, 2005; Burge, 1998).

### **2.3. Case Based Reasoning**

Human beings have the ability to handle situations by remembering (recalling) the past experiences that we have on similar situations. As Kolodner (1992) indicate we are likely to observe peoples who try to handle problem by relating it with other experienced situations. We normally learn from our successful and wrong activities to handle future similar situations in the right way and not to repeat our mistakes. Remembering and reusing previously solved problems, and learning from experiences for future use, is natural and useful (Aamodt and Plaza, 1994; Kolodner, 1992).

KBS that use case based reasoning is known as the case based system (Kyung and Dongkon, 1999). Case based systems are designed to work in the same way with the basic idea of similar problems have similar solutions. It is a KBS that solves problems by remembering similar past situation and reusing its solution and lesson learned from it (Aamodt and plaza, 1994; Kolodner, 1992).

Case based system represents situations or domain knowledge in the form of cases and it uses case based reasoning technique to solve new problems or to handle new situations (Kyung and Dongkon, 1999).

Case can be defined as

“...a contextualised piece of knowledge representing an experience that teaches a lesson fundamental to achieving the goals of a reasoner.” (Kolodner, 1992)

Case based reasoning is also defined as

“..... adapting old solutions to meet new demands; using old cases to explain new situations; using old cases to critique new solutions; or reasoning from precedents to interpret a new situation (much like lawyers do) or create an equitable solution to a new problem (much like labor mediators do)”(Kolodner, 1992)

### 2.3.1. Cases

A case refers to specific experience or knowledge tied to specific situation that is worth remembering for future use. So cases in the knowledge base represent collection of specific experienced captured and learned situations of the application domain (Aamodt and plaza, 1994; Kolodner, 1992). The structure and content of cases highly affect a case based reasoner performance (Aamodt and plaza, 1994).Each case has to have the following three parts: (Bergmann et al, 2005)

- **Situation/Problem description:** describes specific circumstances, states of a situation, and state of the environment when this particular case is recorded.
- **Solution:** provides how the problem described in the problem description was solved or treated in a particular instance.
- **Outcome:** describes the final result or consequence and feedback gained from following the proposed solution

As indicated by Bergmann et al (2005) case representation in case based reasoning uses similar knowledge representation formalisms from AI to represent the experience contained in the cases for reasoning purposes. Different case representations have been proposed. The three classical types of case representations are: (Bergmann et al, 2005),

- **Feature vector (or propositional) cases:** Feature-vector approaches represent a case as pairs of attribute-value format. It is similar to the propositional representations used in Machine Learning. It supports Nearest Neighbor matching and instance-based learning.

- **Structured (or relational) cases:** The structured approach represents cases around frame-based formalism like relational representations in Machine Learning.
- **Textual (or semi-structured) cases:** Textual case representations decompose the text that constitutes a case into information entities (IEs). An IE is a word or a phrase contained in the text that is relevant to determine the reusability of the episode captured in the case.

### 2.3.2. Styles of Case Based Reasoning

There are two styles of case based reasoning: (Kolodner, 1992; Salem, 2007)

- problem solving and
- interpretive

#### **Problem solving case based reasoning**

Problem solving case based reasoning reuses old similar cases to understand the problem, suggest a solution, and/or to keep it from failure. Almost right solutions for new problems can be found from old solutions and old solutions can provide warnings of potential mistakes or failures.

Problem solving style intensively applies adaptation processes to generate solutions and interpretive processes to judge derived solutions (Kolodner, 1992). Problem solving case based reasoning is useful for a wide variety of problem solving tasks, including planning, diagnosis, and design (Kolodner, 1992; Salem, 2007). In each of these, cases are useful in suggesting solutions and in warning of possible problems that might arise.

#### **Interpretive style**

Interpretive case based reasoning evaluate new situations (new cases) in the perspective of old situations (old cases). It uses old cases to provide justifications for solutions, allowing evaluation of solutions when no clear-cut methods are available. This kind of cased based reasoning is used in courts, as the lawyers use old cases to justify an

argument in for the new case. It is also useful for situation classification, evaluation of a solution, argumentation, justification of a solution, interpretation, or plan, and projection of effects of a decision or plan. (Kolodner, 1992; Salem, 2007). Interpretive style can also be found in problem solving (Kolodner, 1992).

Both styles of case based reasoning depend heavily on a case retrieval mechanism that can recall useful cases at appropriate times, and in both, storage of new situations back into memory allows learning from experience (Kolodner, 1992).

### **2.3.3. Case Based Reasoning Cycle**

Case based reasoning, as its name indicates, uses cases to reason about a given problem. Generally case based reasoning process is divided in to four; (Aamodt and plaza, 1994; Lopez et al, 2006;. Kolodner, 1992),

1. RETRIEVE the most similar case or cases;
2. REUSE the case or cases to attempt to solve the problem;
3. REVISE the proposed solution if necessary, and
4. RETAIN the new solution as a part of a new case

When a problem occurs, the method searches its knowledge base and retrieves the most similar case or cases. The information and knowledge in the retrieved case is reused to propose a solution for the problem. The proposed solution is then evaluated to check that the problem is solved successfully or failed. A case based system updates its knowledge from its experiences, so based on the evaluation result the method retains the new experience learned from this problem solving process (Aamodt and plaza, 1994; Lopez et al, 2006; Kolodner, 1992).

Each of the processes is discussed in the following sections. The Figure 2.3 below shows sequence of the processes.



Various researchers describe it in different ways. Aamodt and plaza (1994), subdivided **Case Retrieval** into three subtasks;

- i. Identify features: involves indexing the problem with the most descriptive features in order to match it with indexed saved cases. In another words, it identifies its descriptive properties and take out the properties which doesn't describe the problem strongly.
- ii. Initially Match: finding previous cases that match with the problem at hand and it retrieves a set of plausible candidates. That means it involves searching and similarity assessment to produce a set of similar cases.
- iii. Select: selecting the best-matched case from the set of similar cases. It is based on the similarity assessment result that the best matched case or set of cases is selected as output of the retrieval process.

On the other hand, Kolodner (1992) subdivides the retrieving process into;

- i. Recall previous cases; the aim of this step is to retrieve "good" cases that have the potential to make relevant predictions about the new case
- ii. Select the best subset; the goal of this step is to select the best cases from the result of the first step. Sometimes it is appropriate to choose one best case, sometimes a small set is needed.

The quality of the retrieval process depends on its descriptive feature identifying algorithm, searching algorithm and similarity assessment method. The two of most well known algorithm for case retrieval are: (Singh et al, 2007; Watson & Marir, 1994)

- Nearest Neighbor
- Induction

These methods can be used alone or combined into hybrid retrieval strategies.

## Nearest neighbor algorithm

The Nearest Neighbor algorithm measure the similarity of stored cases with a new input case, based on matching a weighted sum of features. (Kyung and Dongkon, 1999; Watson & Marir, 1994; Singh et al, 2007). When a new case doesn't exactly match with old cases then this algorithm will return nearest match from case based reasoning library. It is suitable when there are attributes that has numeric (continuous) value (fang and Songdong, 2007). But the retrieval time by this algorithm increases linearly as the case in the case base increases.

The algorithm for Nearest Neighbor is as follows:( Salem et al, 2005)

*For each feature in the input case :*

*Find the corresponding feature in the stored case*

*Compare the two values to each other and compute the degree of match*

*Multiply by a coefficient representing the importance of the feature to the match*

*Add the results to derive an average match score*

*This number represents the degree of match of the old case to the input.*

Nearest Neighbor algorithm can be represented in the following equation (Watson & Marir, 1994).

$$NN(I, R) = \frac{\sum_{i=1}^n w_i \times sim(f_i^I, f_i^R)}{\sum_{i=1}^n w_i}$$

*Figure 2.4 Nearest Neighbor Algorithm*

Where  $w$  is the importance weighting of a attribute,

$sim$  is the local similarity function, and

$f_i^I$  and  $f_i^R$  are the values for attribute  $i$  in the input case (I) and a case in the case base (R) respectively.

$n$  is number of attributes in the case

## **Induction**

Induction algorithm tries to extract rules or construct decision trees from previously solved cases. In case based reasoning systems, it analyzes the case base in order to construct a decision tree that classifies the cases. The most popular induction algorithm in case based reasoning is called ID3. It uses a heuristic called information gain to find the most promising attribute on which to divide the case base (Salem et al, 2005; Singh et al, 2007)

Induction algorithm is helpful when a single case feature, which is dependent upon others, is required as a solution. This algorithm identify which features do the best job in discriminating cases, and generate a decision tree type structure to organize the cases in memory (Watson & Marir, 1994).

### **2.3.3.2. Reuse**

The knowledge in the retrieved case is an input for reuse process to propose a solution for the new problem. This process found next to retrieval and followed by revise and retains (Lopez et al, 2006).

As indicated by Aamodt and plaza (1994), reusing of the retrieved case solution in the context of the new case is based on the differences among the past and the current case, and what part of a retrieved case can be transferred to the new case.

Proposing a solution can be performed into two ways: reusing the solution as it is or by adapting it. When the selected case and the new case do not have significant difference, the solution in the selected case will be proposed as it is for the new problem. Whereas, if there is a significant difference between them, the solution in the selected case is adapted based on the unique feature of the new case, this process is known as adaptation (Lopez et al, 2006).

Adaptation methods differ in complexity with respect to two dimensions: what is changed in the retrieved solution, and how the change is achieved. As mentioned by Lopez et al(2006) , there are three types of adaptation method.

- *Substitution:* adaptation that simply replaces some part(s) of the retrieved solution,
- *Transformation:* adaptation that alters the structure of the solution
- *Special:* adaptation that applies specialized heuristic knowledge to repair the retrieved solution, or replays the method used to derive the retrieved solution for the new problem.

However the above techniques of adaptations are applied for cases that have more than one attribute in the solution part of the case. If only one attribute make up the solution part of the case then a voting adaptation technique can be applied at the reuse process to propose the solution for the new case. Voting technique works by looking the solution of the closest retrieved cases. If there is more than one solution from the closest retrieved cases then the one that occur frequently will be proposed as a solution for the problem (Salem et al, 2005).

### **2.3.3.3. Revise**

In case based systems proposing a solution is not the only goal, it also aims to learn from the consequence of applying the proposed solution. This process evaluates how good the proposed solution is for the given problem. The evaluation is performed by using simulator, by getting feedback from a human expert of the application domain or by applying it in the real world and see the result. This process may take hours, days or months until the result is being realized. The system learns from the result whatever it is: success or failure. If it is failure, the fault needs to be repaired and explanation of why the failure occurs should be given to prevent future similar problems from such kind of failures (Aamodt and plaza, 1994).

#### **2.3.3.4. Retain**

Case based systems upgrade their domain knowledge by learning from new experiences obtained while problems are solved. After the proposed solution for the given problem is evaluated in the revise process, the retain process identifies useful and worth remembering new experiences and decides how to merge with existing knowledge. This type of learning is known as incremental learning because it always adds knowledge that is new and useful in addition to the existing knowledge (Aamodt and plaza, 1994).

The new experience may be success or failure. If it is success, the retain process keeps how the problem is solved by modifying existing cases or by creating a new case if it has significant difference with the existing ones (Watson & Marir, 1994). The advantage of keeping failure processes is to prevent future similar problems from similar failure. The failure can be task failure where the solution is unsuccessful or expectation failure where the observed solution is different from the expected solution.

### **2.4. Rule based reasoning**

Rule based reasoning is one of AI reasoning techniques. It reasons from domain knowledge represented in set of rules. Rules are one of the most popular knowledge representations. The basic form of a rule is if <conditions> then <conclusion> where <conditions> represent premises and <conclusion> represents associated action for the premises. When the conditions of a rule are satisfied, the conclusion becomes active; often the rule is said to be applied (Prentzas and Hatzilygeroudis, 2007).

A rule based system has one more component, which is known as working memory, in addition to knowledge base and inference engine. The inference engine receives a problem from the working memory and provides the reasoning result to the working memory. The working memory contains the description of the problem and updates its content based on the reasoning results received from the inference engine. The rules in the knowledge base and the reasoning method used by the inference engine are discussed below.

### **2.4.1. Rules**

Normally rules represent what to do or not to do while certain situations are met. Similarly, the application domain knowledge is represented with set of rules that represent the facts that would be true when some conditions are given true. A typical rule has a format of If <conditions> then <conclusion>; where conditions represent premises or facts, and conclusion represents associated actions for the premises. The condition might be a premise or set of premises that are connected with logical operators: AND & OR. The conclusion can be an action to be taken or facts that are inferred from the given premises (Prentzas and Hatzilygeroudis, 2007)

Frequently used means of acquiring rules is, interviewing of the domain experts. Rules represent general knowledge of the application domain. They preserve modularity and ease of explanation because they are used in a direct fashion as acquired from experts. Its shortcoming is its difficulty in acquiring complete and perfect knowledge in a complex domain due to the experts may be incapable of expressing their knowledge or unavailability of some experts. In addition, sometimes representing the domain with only general knowledge may not be enough (Prentzas and Hatzilygeroudis, 2003; Prentzas and Hatzilygeroudis, 2007).

### **2.4.2. Rule Based Reasoning Technique**

Rule based reasoning technique represents how a system solves a problem by using knowledge of the application domain that is represented in form of rules. There are two rule based reasoning methods (Buchanan and Duda, 1982).

1. Data-driven also called forward chaining
2. Goal-driven also called backward chaining

#### **2.4.2.1. Data-Driven Control**

In this process, it receives a problem description as a set of conditions and tries to derive conclusions as a solution. With data-driven control, rules are applied whenever their left-hand-side conditions are satisfied. To use this strategy, one must begin by entering

information about the current problem as facts in the database (Buchanan and Duda,, 1982). Here we assume that a rule is applicable whenever there are facts in the database that satisfy the conditions in its left-hand side. If there are no applicable rules, there is nothing to be done, except perhaps to return to the user and ask him or her to supply some additional information.

#### **2.4.2.2. Goal-Driven Control**

This strategy focuses its efforts by only considering rules that are applicable to some particular goal. It is similar with forward chaining in most process, the big difference is it receives the problem description as set of conclusions, instead of conditions, and tries to find the premises or causes of the conclusions (Buchanan and Duda, 1982).

Goal-driven control is variously known as top-down, backward chaining, or consequent reasoning. A primary virtue of this strategy is that it does not seek information and does not apply rules that are unrelated to its overall goal (Buchanan and Duda, 1982).

### **2.5. Case based Vs Rule based reasoning**

The two most known approaches for problem solving in intelligent systems are case based and Rule based reasoning.

#### **2.5.1. In Knowledge Representation**

One of the main difference between case based and rule based reasoning system is on the method in which knowledge is stored and used (Knowledge Representation) (Holt and Benwell, 1999)

Rules and cases are alternative ways of representing knowledge of an application domain. Rules are used in rule based reasoning while cases are used in case based reasoning. Rules are suitable to represent general and normative knowledge, whereas cases are suitable for detailed specific situations (Holt and Benwell, 1999).

Cases are capable of representing specific historical knowledge. Cases are natural and easy to obtain. They can be collected from historical records, repair logs, or other

sources; eliminates challenging of knowledge acquisition from experts (Holt and Benwell, 1999).

Rules in a rule-based system have abilities to represent experiential knowledge acquired from experts in a direct fashion. The problem here is that it is difficult to acquire complete and perfect knowledge in a complex domain. Commonly way of acquiring these rule are through interviewing experts, but the expert may be incapable of expressing his/her knowledge and it also takes time (Prentzas and Hatzilygeroudis, 2003; Prentzas and Hatzilygeroudis, 2007). The summary of comparison in terms of knowledge representation is found in the table 2.1 below.

<b>Case based reasoning</b>	<b>Rule based reasoning</b>
Cases are chunks of domain knowledge.	Rules are independent, consistent pieces of domain knowledge.
Knowledge is stored in the form of case.	Knowledge is extracted from experts and encoded in rules.
Cases are constants	Rules are patterns

*Table 2.1: Comparison of CBR and RBR in Terms of knowledge Representation (Holt and Benwell, 1999)*

### **2.5.2. In Handling Missing and Uncertain Knowledge**

Case based reasoning uses partial matching to reason about a problem. If some of the given problem descriptions match with a case, the case is applicable to propose a solution. It also tries to handle novel problems by referring several cases for the different descriptive features of the problem or by adapting a case which has some kind of similarity (Prentzas and Hatzilygeroudis, 2003; Prentzas and Hatzilygeroudis, 2007)..

Rule based reasoning uses perfect matching to apply a rule for a given problem. It doesn't handle missing information and unexpected data values. For example, if a problem misses some of a rule's features, the rule is considered as not applicable and the reasoner doesn't draw a conclusion from the given conditions. It also doesn't solve problems that

don't perfectly match with the available rules in the knowledge-base (Prentzas and Hatzilygeroudis, 2003; Prentzas and Hatzilygeroudis, 2007).

### **2.5.3. In Problem Solving**

A case based reasoning method cut down most of the problem solving processes by applying a case that has already solved while a rule-based reasoning method solves problems from scratch whether the same problem has been solved successfully in past or not (Prentzas and Hatzilygeroudis, 2007).

Some of the advantages of CBR over rule-based systems in problem solving are: (Holt and Benwell, 1999).

- It works better for problems where problems have many exceptions to rules,
- It works better when problems are not fully understood,
- It learn from experience, that is keep up with knowledge that workers learn in their daily experiences, indicating an ability to store temporal information,
- It represents the expert's knowledge more accurately in the form of particular experiences (cases) rather than in the form of rules,
- It offers cost-effective solutions to knowledge acquisition bottleneck problems.

## **2.6. Case Based Reasoning Tools**

There are different types of tools that can be used for developing a case base reasoning system. Most of the case based reasoning tools are commercials. The following case based reasoning tools are indicated on the paper of Ashraf and Iqbal (2008):

**ReCall:** This case based reasoning tool is written in C++ language. It provides both Nearest Neighbor and inductive retrieval algorithm. It can run on Windows and UNIX workstations under Motif, Sun, HP series 700 and DEC Alpha, designed in open architecture that allows the user to add case based reasoning functionality in the applications.

**ReMind:** It was basically developed for Macintosh by Cognitive Systems Inc. but now it works on Windows and UNIX. ReMind offer different kind of retrieval algorithm such as, Nearest Neighbor and inductive.

**CBR-Express:** This case based reasoning tool primarily designed for help desk applications. It provides a comfortable user interface and fast retrieval speed. It has simple case structure and Nearest Neighbor retrieval algorithm cases.

**Kate:** This tool is developed by AcknoSoft that can run on MS Windows, Mac, or SUN. . Kate is made up of Kate-Induction, Kate-CBR, Kate-Editor and Kate-Runtime. This tool support both kind of Nearest Neighbor and inductive retrieval algorithm.

**JCOLIBRI:** is a technological evolution of COLIBRI and it is object-oriented framework in Java which is designed for building Case Based Reasoning systems. (Recio-Garcia and Daz-Agudo, 2004; Lotfy- Abdrabou and Salem, 2010). The design of the JCOLIBRI framework comprises a hierarchy of Java classes and a number of XML files. It support Nearest Neighbor retrieval algorithm.

All the tools indicated on the above are commercial except JCOLIBRI and they are difficult to be accessible by the researcher. Even though JCOLIBRI is open source software it doesn't allow voting automatically for the reuse task, which was used in reuse process of the study. As a result of this python 2.6 is used to develop the case base reasoning system for the study.

## **2.7. Approaches to Evaluate the Performance of CBR system**

Evaluation of knowledge based system includes both system performance (statistical analysis) and user acceptance (Buchanan and Forsythe; 1991). The statistical analysis for case based reasoning can be conducted for both retrieval and reuse process. The first task of case based reasoning is to retrieve cases that are relevant to the new case (Aamodt, and Plaza, 1994). As retrieval task of the case based reasoning system aims to retrieve relevant cases from the case base, precision and recall are useful measures of retrieval performance in case based reasoning (McSherry, 2001). Recall is defined as the ratio of

the number of relevant cases returned to the total number of relevant cases for the new case in the case base (Junker et al, 1999; Losee, 2000; McSherry, 2001). Where as precision is the ratio of the number of relevant cases returned to the total numbers of cases for a given new case (Junker et al, 1999; Losee, 2000; McSherry, 2001).

$$\text{Recall} = \frac{\text{Number of relevant cases retrieved}}{\text{Total number of relevant cases}}$$

$$\text{Precision} = \frac{\text{Number of relevant cases retrieved}}{\text{Total number of cases retrieved}}$$

Accuracy is another important measurement to evaluate the case based reasoning performance (McSherry, 2001). Accuracy measurement can be used to evaluate the reuse task, as the aim of the reuse process in case based reasoning is to solve problems correctly. This measurement had been used by Salem et al (2005) on their research. Accuracy is defined as the ratio of the number of correctly recommended cases to the total number of tested cases (Junker et al, 1999).

$$\text{Accuracy} = \frac{\text{Total number of correctly recommended cases}}{\text{Total number of tested cases}}$$

Only system performance evaluation based on statistical analysis does not assure the applicability of the system in the real life. Even though system that achieves good system performance statistically, it may not be comfortable to the user in solving the particular problem (Buchanan and Forsythe, 1991). As a result of this user acceptance is conducted to assess the applicability of the system for the real life.

## **2.8. Knowledge Based System in Medical Domain**

Medical domain is becoming popular research area for investigating the application of Information Technology. The advent of AI in the 1970's has brought expert systems in different domains; the most popular application has been in the area of health and medicine (Masizana-Katongo et al, 2009). Different researches have been conducted in health and medicine area, the most well known application are described as follows:

### **2.8.1. Rule Based Reasoning in Medicine**

#### **MYCIN**

MYCIN is a computer program designed to aid physicians in the diagnosis and treatment of meningitis and bacteremia infections. It is one of the most popular medical expert system used to assist to diagnose and treat blood diseases. It was developed at Stanford University in the mid-1970s. It was the first in demonstrating how a system can be used to successfully perform medical diagnosis (Masizana-Katongo et al, 2009).

#### **PACE**

PACE stands for Patient Care Expert System. It was designed in 1977 with the intention of making intelligent selection from vast and ever changing information related to health so as to facilitate patient care.

PACE is made up of the most extensive regularly maintained computer based knowledge base of nursing and software that permits easy care plan development. The knowledge base is used by the system software to make easy care plan development (Evans, 1990)

At first the system was designed for educational system for nursing profession. After years PACE has evolved and passed through many development generations. Now PACE has become an advanced clinical management system that can assist the entire health care field to diagnose and care for patients with pulmonary diseases, (Masizana-Katongo et al, 2009).

#### **MITIS**

It is another expert system that was developed in 2004 at the National Technical University of Athens. It was designed to help in managing and processing datas related to obstetrical, gynaecological, and radiological. MITIS is a WWW-based medical information system based on three-tier client–server architecture and designed to provide mainly gynecologists with unified patient management capabilities, either internal in the

hospital or external at a private office (Matsopoulou et al, 2004; Masizana-Katongo et al, 2009).

Anteneh (2004) has conducted a study to investigate the rule based reasoning approach in designing and developing knowledge based system prototype for antiretroviral therapy. He explores the applicability of knowledge based system in assisting the choice of drug for individual patients in the area.

Redit (2006) conducted a study to investigate KBS for HIV pre-testing counseling. The main objective of her work is to look into the feasibility of employing the expert system in the area of pre-test counseling by using the knowledge based system technology. She used a rule based reasoning technique and Pro Gold expert system shell in developing the prototype. The prototype is able to show the applicability of the technology to the area at a satisfactory level.

All the knowledge based systems that are discussed in the above section are made up of rule based reasoning technique. Even though Rule Based Reasoning has some advantage in developing KBS it has drawback when it is applied to medical domain (Luan, 2005).

- The knowledge acquisition process in medical domain is challenging and it is difficult to represent in the form of rule.
- Rule based reasoning doesn't solve problems using past experience. It uses rule that means it requires all facts to be known and being represented in the form of rule. However it is difficult to know all the knowledge in medical domain and the knowledge in medical domain are highly context dependent.
- Rule based system has poor scalability. It is difficult to add or update the existing knowledge. To update the existing knowledge of rule based system, mostly it requires redesign of a part of the system which can cause serious problem in many medical applications.

## 2.8.2. Case Based Reasoning in Medicine

Case based reasoning approach provides a base for a new technology of building intelligent computer aided diagnoses systems that directly solve the limitation found in Rule based reasoning techniques, e.g. the problems of knowledge acquisition, remembering, robust and maintenance (Salem, 2007).

Case based Reasoning approach has been investigated in improving human decision-making and has become popular in developing knowledge-based systems in health and medicine (Salem et al, 2005). CBR is appropriate in medicine for some important reasons; cognitive adequateness, explicit experience, duality of objective and subjective knowledge, automatic acquisition of subjective knowledge, and system integration. Different researches have been conducted in medical domain that employed case based reasoning. Some CBR-systems are:

CASEY designed to give heart failure diagnosis. It employs three steps for its reasoning functionality: the first one is that it searches for similar cases, the second step is the determination process concerning differences and their evidences between a current and a similar case, and in the third step it transfers the diagnosis of the similar to the current case or if the differences between both cases are too important it uses general adaptation operators for modifying the diagnosis (Salem et al, 2005; Schmidt et al, 2001).

Alemu (2010) investigated the potential of case based reasoning in solving complex side effects of HIV/AIDS cases for person living with HIV/AIDS who have begun antiretroviral therapy. He used JCOLIBRI version 1.1 in designing the prototype. The system registers 72% and 63% of recall and precision respectively (Alemu, 2010).

Other case based reasoning systems have been developed to investigate the technique in different medical area. Some of the systems are as follows: (Salem et al, 2005)

- GS.52 developed which give diagnostic support system for dysmorphic syndromes,
- NIMON system that monitors a renal function

- ICONS that presents a suitable calculated antibiotics therapy advise for intensive care patients

Decisions about the management of hypertensive patients should not only take blood pressure levels into account, but also the presence of other cardiovascular risk factors, target organ damage, and associated clinical conditions. However as a research shows from 167 countries, 45% countries' health professionals were not trained to manage hypertension (Whitworth, 2003). And in Addis Ababa hypertension is widely spread (Tesfaye et al, 2009). Therefore the main aim of the study is to investigate the applicability of cases based reasoning technique in designing KBS for hypertension management.

## CHAPTER THREE

### KNOWLEDGE ACQUISITION AND MODELLING

#### 3.1. Knowledge Acquisition to Identify Case feature

One of the most important processes in knowledge based system development is knowledge acquisition. How knowledge is obtained and where it is obtained determines the usefulness of the system (Fredlund et al, 1996). Knowledge acquisition is a vital stage in the development of knowledge based system. It is an important obstacle and time consuming part when constructing expert systems. Knowledge acquisition is referred to as a process of eliciting, structuring and representing knowledge from some knowledge source, usually human experts, in order to build knowledge based system (Sagheb-Tehrani, 2009). As indicated by Jones (1989) knowledge elicitation and structuring are the two most important activities of knowledge acquisition processes that are carried out by knowledge engineer in order to build knowledge based system.

A general sequence of events carried out by knowledge engineer for knowledge acquisition in designing knowledge based system is as follows: (Miller, 2009)

- i. Eliciting data and information from the domain expert
- ii. Interpreting the data and information and making conclusion about the expert's underlying knowledge and reasoning processes
- iii. constructing model which describes the expert's knowledge
- iv. Repeat steps i-iii as the knowledge based system evolves into a functional system

The sequence of knowledge acquisition process in case based reasoning system is basically the same with the above sequence for Knowledge acquisition. Knowledge acquisition in case based reasoning carries out problem analysis, which involves transforming the information taken from the domain expert into the problem and solution fields in the case based data structure (Miller, 2009)

In this chapter, the researcher identifies the variables and concepts that are used in hypertension management and model it by using hierarchical tree structure. Furthermore, the hypertension case structure, which is used as a framework for the case base in the case based reasoning system, is constructed. since the research investigates the applicability of case based reasoning in health specifically for hypertension management, the knowledge for this research is collected from the hypertension cases, domain expert and relevant documents through knowledge elicitation techniques of semi-structured interview and document analysis.

### **3.1.1. Knowledge Acquisition from Relevant Document**

In order to elicit knowledge from the domain experts and from hypertension cases in a proper manner, the researcher reviews different relevant documents related to hypertension management. These documents are

- articles that are published in different journals,
- vouchers and
- Standard treatment guideline (STG) for Ethiopians health professional.

In addition, the researcher identifies variables that are considered in hypertension management. Most of the identified variables are the same with the variables that are identified through semi-structured interview and hypertension case analysis. However some other documents indicate that race should be considered in hypertension management, and it is not recorded in the hypertension patient recorded file.

### **3.1.2. Knowledge Acquisition from Domain Expert**

A case based reasoning system depends on several containers of knowledge which are stored as past solved problems. In addition to the old cases, the knowledge of experts plays a vital role in structuring the knowledge base and developing the KBS. When the correctness and accuracy domain knowledge (Experts knowledge) used by the system increase, the case based reasoning system's performance will increase. As results of this,

expert on the profession (doctor, health officer and nurse) were interviewed to obtain the required knowledge for hypertension management.

The interview focuses on the concepts that the domain experts focuses during the hypertension management. The sample of semi structured interview is found in the Appendix I. As compared to the concepts acquired from hypertension cases, the domain experts consider weight and height of the patients in hypertension management. However these concepts are not found in the patient record file.

### **3.1.3. Knowledge Acquisition from Hypertension Cases**

After the knowledge acquisition from domain experts and other relevant documents, the main knowledge that is used for the case based reasoning system are collected from the hypertension complaints' card at Bole 17 Health Center and Brook Medical Service Plc. This is done with the help of the domain experts. The variables that are identified from the patient record file, as a main variable to manage the hypertension cases, are similar with the variables that are identified through semi structured interview with the domain experts and other relevant document analysis before.

The variables that are identified from hypertension patient record file, which are important to hypertension management, are Age, Sex, Systolic blood pressure, Diastolic blood pressure, Symptom, Co-morbidity, Co-medication, life style, Pregnancy. As compared to variables that are identified from the domain experts and relevant document, variables like race, type, height, and weight are not found in the patient record file.

## **3.2. Knowledge Modeling**

After the knowledge is acquired from hypertension cases, domain experts (health professionals) and other relevant documents, the next step is modeling the knowledge. The knowledge modeling involves organizing and structuring of the knowledge gathered during knowledge acquisition. This activity provides an implementation independent specification of the knowledge to be represented in the knowledge base. Knowledge Modeling is the concept of representing information and the logic for purpose of

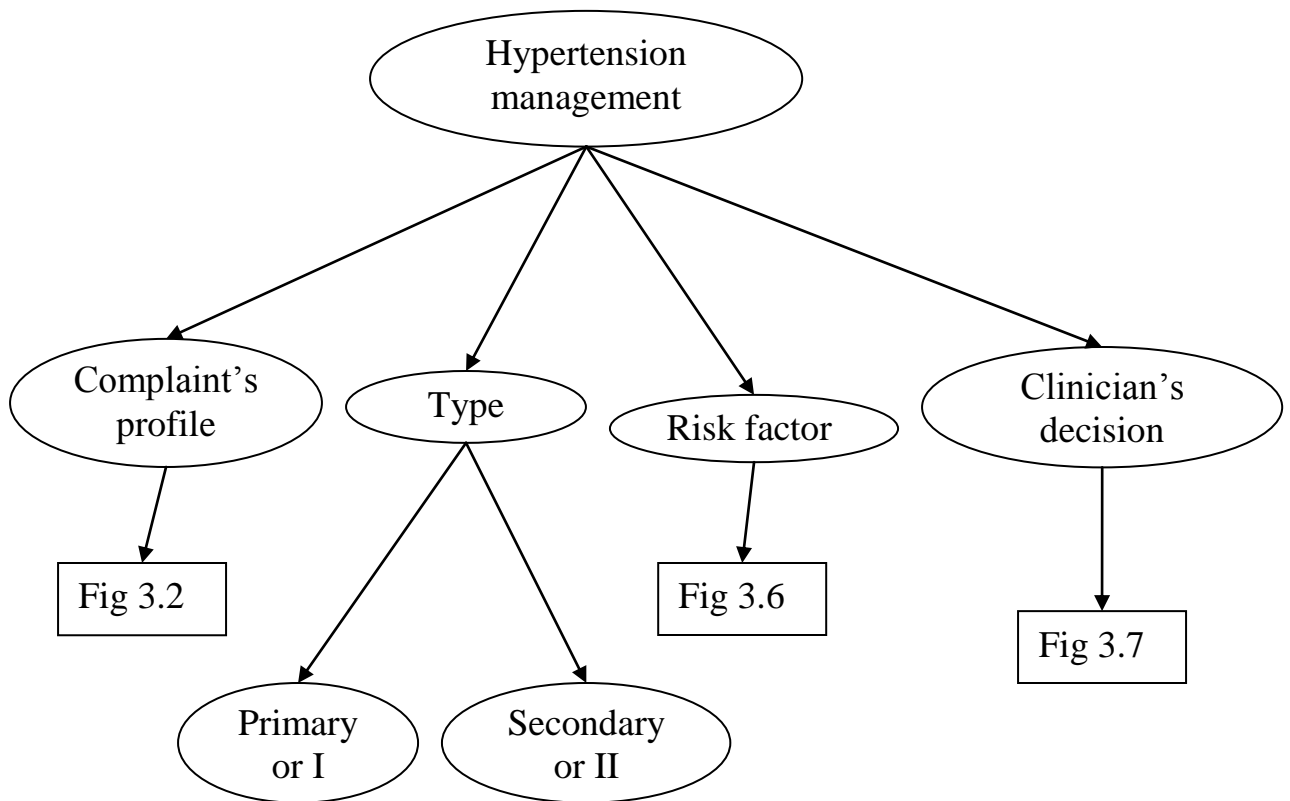
capturing, sharing and processing knowledge to simulate intelligence (Makhfi, 2011). Here, the basic concepts that reveal the main activities and decisions that are made to solve cases in the domain are modeled.

Hierarchical tree structure is used to model knowledge that is acquired through the knowledge acquisitions. The main reason that the knowledge is modeled through Hierarchical tree structure is that hierarchical tree structure can easily model concepts and clearly explain the concepts in the problem area (Lundgren-Cayrol et al. 2001; Alemu 2010; Yemisrach, 2010). Hierarchical tree structure models the knowledge in a hierarchal manner. This model puts the main concepts at the highest level of the hierarchy and other sub factors and concepts that affect the highest concept down in the hierarchy (Lundgren-Cayrol et al. 2001).

The sections below illustrate the conceptual models for hypertension management based on the gathered knowledge from the above three sources (hypertension case, domain expert and other relevant document).

### **3.2.1. Case Concept for Hypertension Management**

In management of hypertension cases, the complaints profile, the type of the hypertension and the risk factor are checked in order to diagnosis and treat the hypertension case. The general concept in managing hypertension case is found in the figure 3.1 below



*Figure 3.1 General Concept for Managing Hypertension Case*

The type of the hypertension is one of the factors that are considered in managing the disease. There are two types of hypertension these are: primary (essential) and secondary hypertension

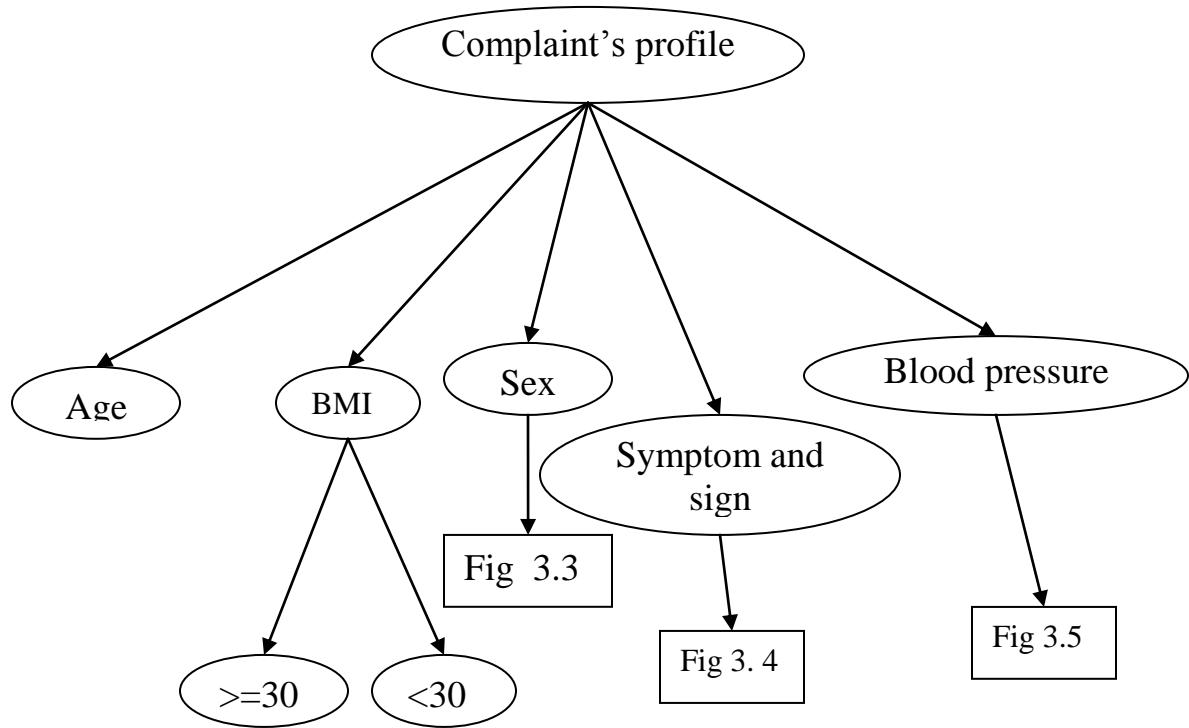
Primary hypertension: is a hypertension that the cause is not easily identifiable. Most people with primary hypertension don't have unique symptoms. It consists of more than 95 % of hypertension cases.

Secondary hypertension is a hypertension that cause is easily identifiable. In secondary hypertension a specific gene or organ can be directly responsible for the hypertension and the diagnosis the disease should consider the type.

The patients profile, the risk factor and the clinician recommendation will be discussed in the following sections.

### 3.2.1.1. Hypertension Compliant's Profile

The complaints' profile, which is needed in hypertension treatment, is depicted in figure 3.2.



*Figure 3.2 Concepts for Hypertension Compliant Profile*

Age, body mass index (BMI), sex, blood pressure records and the symptoms of compliant information are analyzed by the clinician for managing hypertension case.

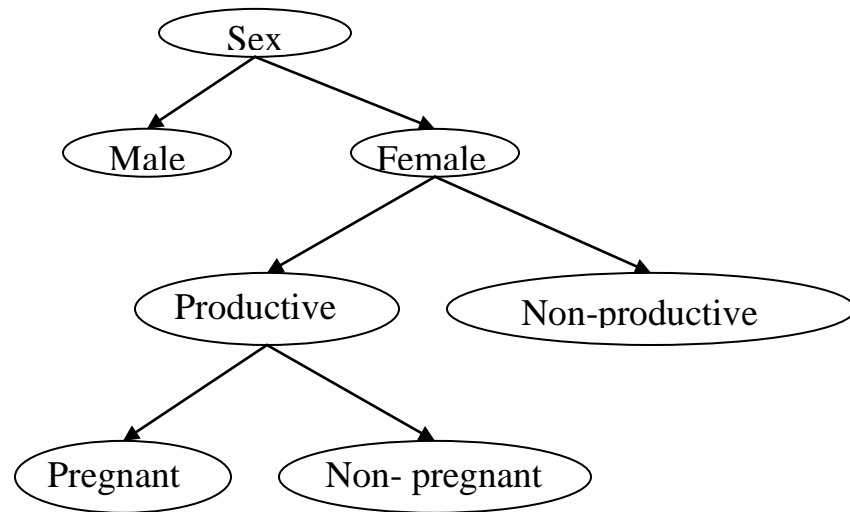
When people are getting older, they are highly prone to hypertension. As a result of this, managing hypertension cases consider the age of the compliant in order to select whether the disease should be managed with drug therapy or life modification only.

The other important factor is body mass index. Body mass index is the ratio of weight and height squared of the patient. Those of who have body mass index of 30 and above are highly prone to hypertension. One of the life modification recommendations for hypertension treatment is weight reduction, which is directly related to the body mass

index of the patient. So, clinician considers the body mass index of the patient in managing the disease.

### 3.2.1.2. Sex of Hypertension Compliant

Sex of the patient is one of the risk factor in hypertension. Research shows that the number of male hypertension patients is greater than the number of female hypertension patients. The model related to sex and pregnancy is shown in the figure 3.3 below.

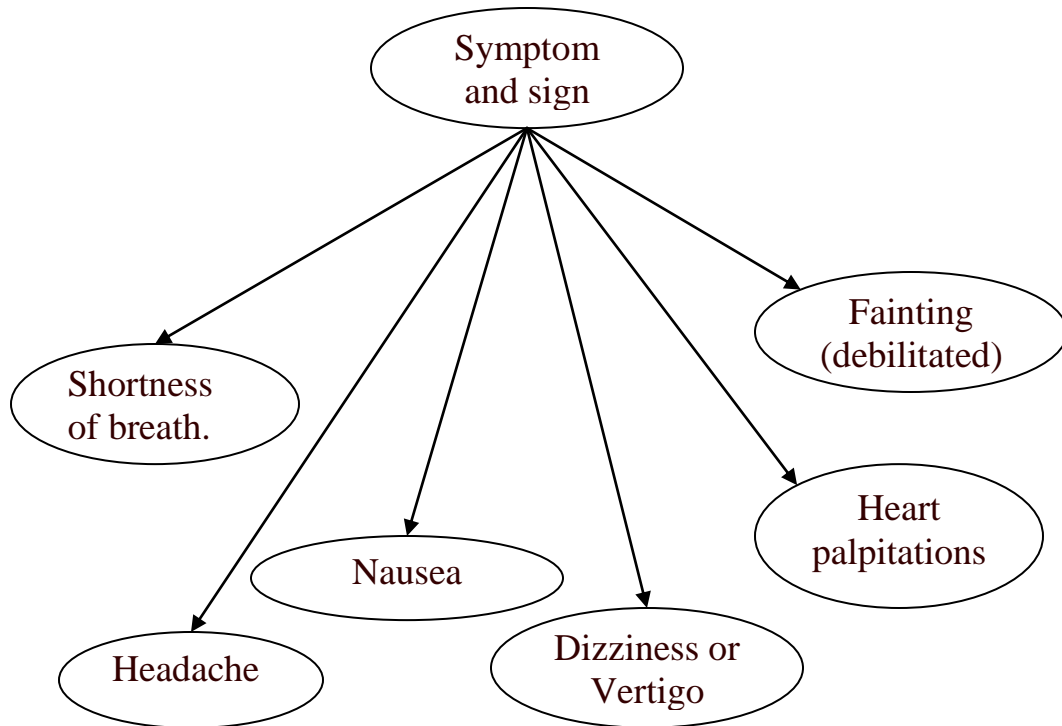


*Figure 3.3 Concept for Sex of the Compliant*

A special treatment consideration is done for pregnant patients. The reason for this is that pregnancy by itself brings hypertension which can be managed by only life style modification. However when a pregnant woman is seriously sick the drug therapy will be used by considering potential teratogenicity of the drug.

### 3.2.1.3. Sign and Symptom of Hypertension

Figure 3.4 depicts the possible common symptoms and sign that can appear with hypertension patients.



*Figure 3.4 Sign and Symptom of Hypertension Patients*

People with hypertension may not have unique symptoms. The possible symptoms of hypertension vary quite a lot from person to person. These symptoms can be symptoms of other health problems, however the following list are a few of the more common symptoms that the hypertension patients can have.

- Headaches
- Dizziness or Vertigo
- Heart palpitations
- Debilitated (unconsciousness)
- Nausea
- Shortness of breath.

### 3.2.1.4. Blood Pressure of Hypertension Compliant

The blood pressure is a key factor for diagnosis and treating of hypertension cases. Clinicians check the blood pressure of the complaints. This data allows the clinician to analyze the severity of the disease so that they can make an appropriate decision and recommendation to diagnosis and treat the disease, and it is depicted in figure 3.5. Diagnosis of hypertension is mainly based on measurement of blood pressure on three separate occasions. Systolic blood pressure (SBP) and diastolic blood pressure (DBP) measurements are optimally taken with a mercury sphygmomanometer; otherwise, a recently calibrated aneroid manometer or validated electronic device can be used.

The range of the SBP and DBP are used to classify the severity of the disease. It is one of the main factors for selecting an appropriate way of approach to treat the disease.

The range of SBP and DBP is summarized in the following tables

<b>Rang of SBP</b>	<b>The degree of SBP</b>
SBP <120	Normal
SBP 120-139	Pre-stage
SBP 140-159	Stage 1
SBP >160	Stage 2

*Table 3.1 SBP Classification in Adults*

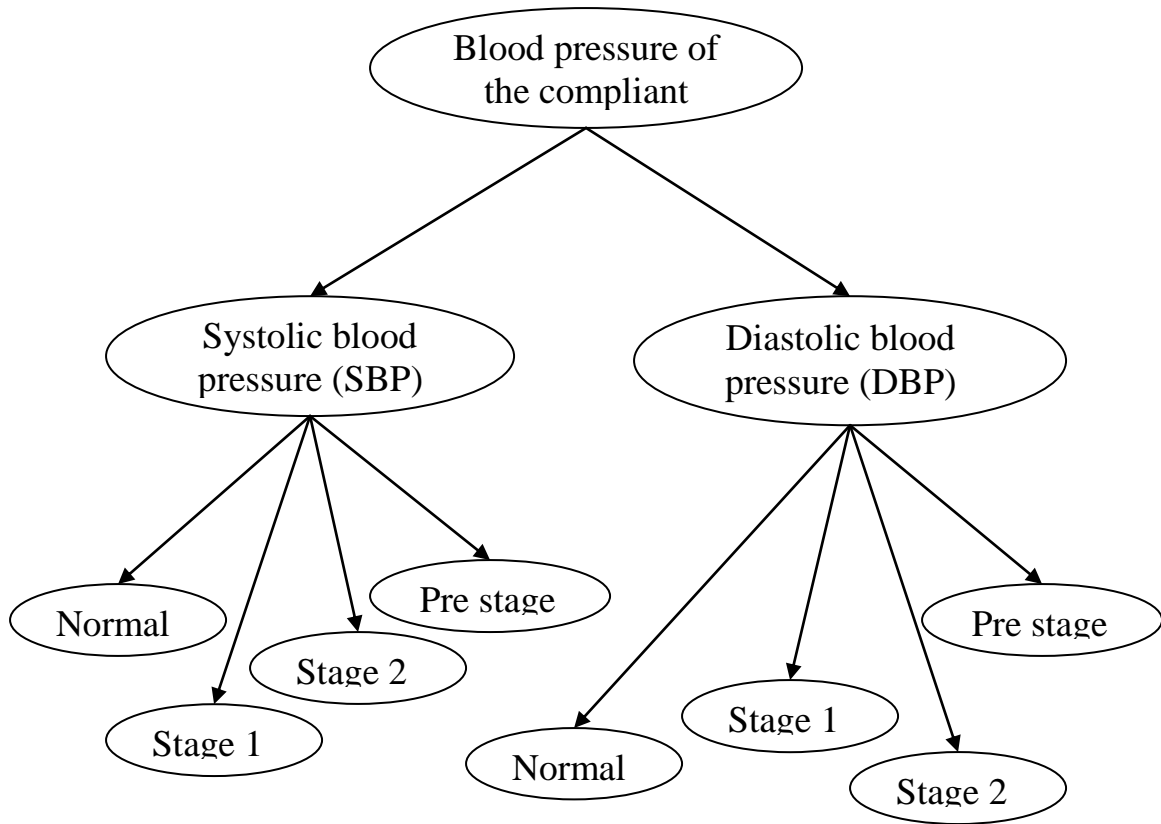


Figure 3.5 Concepts of Blood Pressure of Compliant

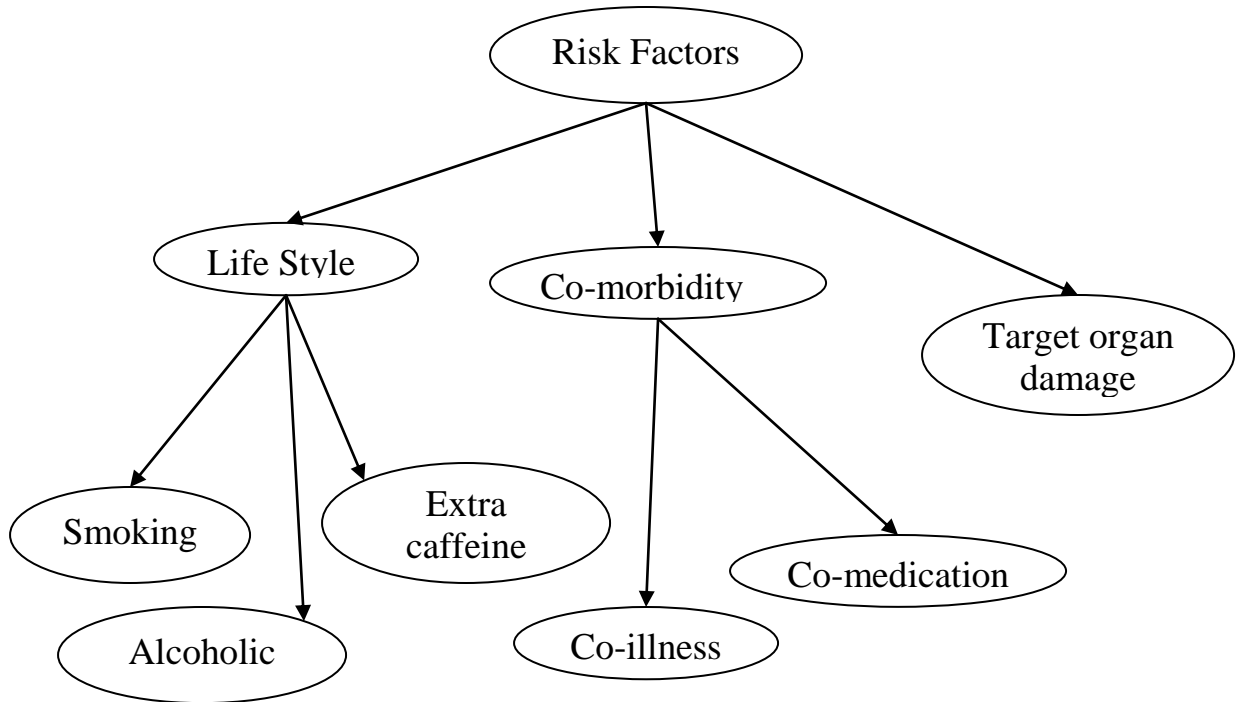
Rang of DBP	The degree of DBP
DBP <80	normal
DBP 80-89	Pre-stage
DBP 90-99	Stage 1
DBP >100	Stage 2

Table 3.2 DBP Classification in Adults

Based on the above tables the severity of the hypertension will be classified as normal, pre-stage, stage1 and stage2. However when systolic and diastolic blood pressures fall into different categories, the higher stage should be used to classify blood pressure status so that based on that the clinician will diagnosis the disease.

### 3.2.1.5. Risk Factors for Hypertension

The management of hypertension disease is not only based on the stage of blood pressure but also considers different risk factors. These risk factors are modeled in figure 3.6



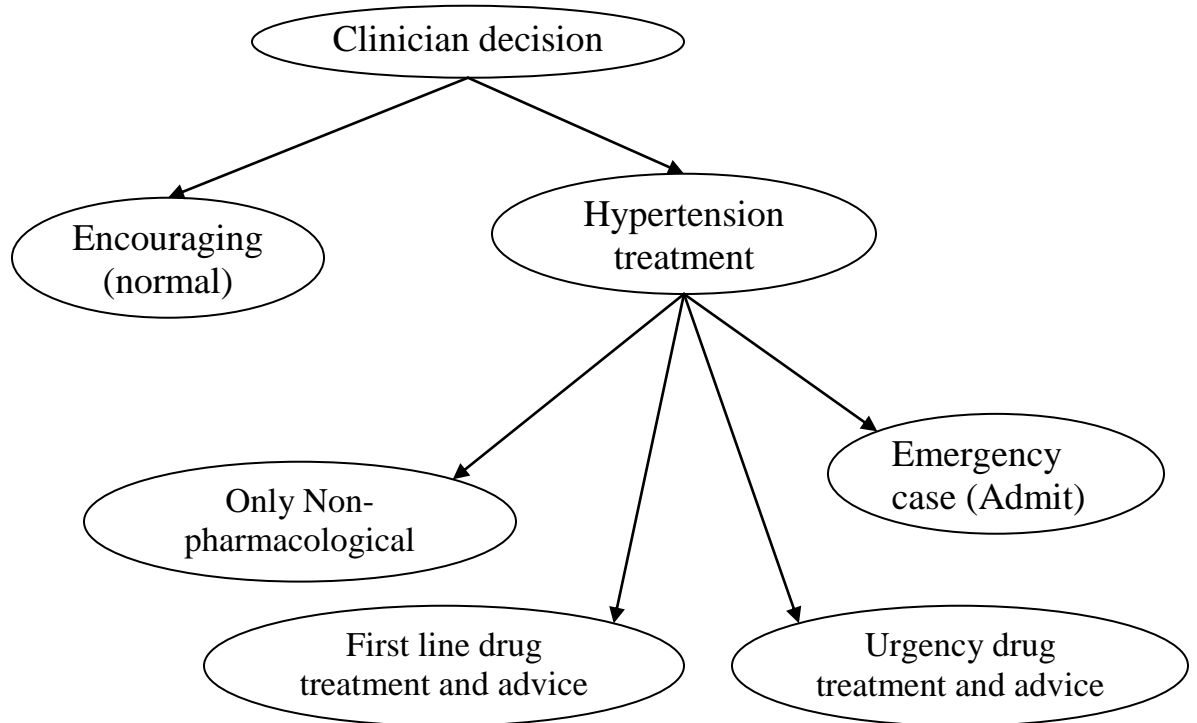
*Figure 3.6 Risk Factor of Hypertension*

Life style, co-morbidity and target organ damage are the main risk factors that cause hypertension. Clinicians analyze them in order to manage the disease properly. A person who smoke cigarettes or who is alcoholic or who take extra caffeine is highly prone to the disease due to this the management of the disease considers the life style of the patients.

A hypertension patient who has diabetics, renal failure or other kind of cardio vascular diseases will be treated in special way as compared to other compliant. Due to this clinician investigates co-illness and co-medication in hypertension management.

### 3.2.1.6. Clinician Decision for Hypertension Management

The clinician decision on hypertension treatment is shown in figure 3.7 below.



*Figure 3.7 Concepts for Clinician Decision for Hypertension Case*

The clinicians investigate risk factors and patient profile such as age, pregnancy, systolic blood pressure, diastolic blood pressure..., which have been discussed in the earlier parts of this section, then the clinician manage the disease. The clinician will check for other diseases if the compliant is free from hypertension. When the compliant has hypertension then the clinician critically selects how the hypertension should be treated.

The main decision that the clinician make on hypertension case is that weather the disease should be treated with drug therapy or life modification only by analyzing the severity of hypertension, patients profile, risk factors and other factors of the patient. The main reason behind this idea is that once the drug therapy is started, it is rare to stop the drug. Most of the time, the drug will be given for life.

As a result of this the clinicians select either the hypertension patient should be treated with life modification (advice) only or with drug therapy. The drug therapy can be first line drug, mostly Thiazide type diuretic, or urgency of drugs like Captopril, or emergency drug like Hydralazine,.

### 3.2.2. Goal of Antihypertensive Treatment

The ultimate goal of preventing and effectively controlling hypertension is to reduce morbidity and mortality by the least intrusive means possible. The primary focus of treatment is in achieving the target systolic blood pressure and it is depicted in the figure 3.8 below.

Treatment to lower levels may be useful, particularly to prevent stroke, to preserve renal function, and to prevent or slow heart failure progression. The targeted blood pressure (TBP) is  $< 130/80$  mm Hg for patients with diabetes, renal insufficiency. Blood pressure control can be achieved by lifestyle modifications and as necessary, pharmacologic treatment.

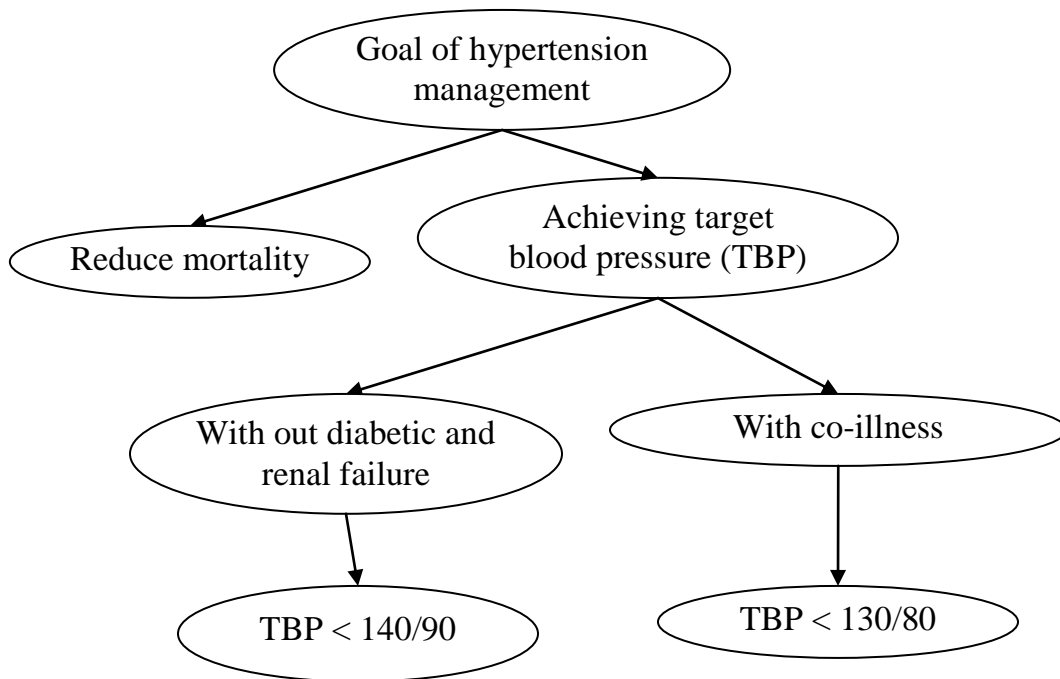


Figure 3.8 Goal of Antihypertensive treatment

### 3.3. Hypertension Case Structure

As it was presented in chapter two a case for case based reasoning has problem description and solution parts. And the hypertension cases that are developed for the study have these two parts.

**Situation/Problem description:** this is a part of the case structure that is made up of attributes (variable) which describe the hypertension cases so as to be solved.

**Solution:** this part of the case structure provides how the hypertension cases should be diagnosed treated.

The researcher with the help of domain expert study the hypertension cases in order to define the attribute that are needed for building the problem description and solution parts of the cases. The case structure is developed from patient record of hypertension patients.

In general, the attributes (variables) that are needed for managing hypertension case are many. However due to shortage of time and scarcity of resource, the researcher used those attributes that have direct impact on hypertension management and which are recorded in the hypertension complaints' history card. The case structure for hypertension management that shows the problem description and solution attributes for this study is found in table 3.3 below.

After the case structure is constructed, the cases that build the case base are collected from the hypertension compliant card histories. The compliant card histories are converted in to cases though the definition of attribute-value pairs in order to make the case base of the system. This conversion is a very hard work and done with the help of domain experts (doctors, health officer, and nurse).

Attribute	Parameter of case
Age	Description
Sex	Description
Systolic blood pressure	Description
Diastolic blood pressure	Description
Pregnancy	Description
Fainting	Description
Smoking	Description
Alcoholic	Description
Co-illness	Description
Co-medication	Description
Palpitation	Description
Headache	Description
Dizziness	Description
Recommendation	Solution

*Table 3.3 Case Structure for Hypertension Management*

The descriptions of the attributes that are used for building the case structure are presented as follows:

**Age:** it is one of major risk factor and it is defined in years. The chances of having high blood pressure increases as a person get older. Old people are highly prone to hypertension.

**Sex:** it define the categories of the compliant as male or female. It is one of the risk factors for the disease. The number of male hypertension patients is greater than the number of female hypertension patients

**Systolic blood pressure:** It is one of critical attribute that helps clinician to identify hypertension. It is the amount of pressure that blood exerts on vessels while the heart is

beating. It is usually measured by mercury sphygmomanometer. In the patient card, blood pressure of the patient is usually written in the form of A/B, where A represent systolic blood pressure and B represent diastolic blood pressure.

**Diastolic blood pressure:** it is the amount of the pressure in blood vessels between heartbeats (when the heart is resting). Like systolic blood pressure, diastolic blood pressure is the major important attribute that helps the clinician to manage the disease. The value of the systolic blood pressure and diastolic blood pressure of the compliant should be based on the measurement of blood pressure on three separated occasions.

**Pregnancy:** it is an attribute that shows whether the compliant is pregnant or not. Pregnancy by itself can cause hypertension. Clinicians treat hypertension of pregnant woman by making special consideration not to affect embryonic development.

**Fainting:** it is an attribute that show whether the compliant is conscious or not. It is a sign that can be shown when the hypertension patient is at a very higher stage or when the hypertension cause an end organ damaged.

**Smoking:** This attribute that shows whether the compliant is a smoker or not. Smoking is one of the life style (habit) that is considered as risk factor for hypertension. Smokers are highly exposed to hypertension compared to non-smokers.

**Alcoholic:** Like smoking, alcoholic is a life style that is considered as a risk factor for hypertension. From non-alcoholic person, alcoholic persons are prone to hypertension disease.

**Co-illness:** this attribute shows whether the compliant has other kind of disease such as diabetics, renal failure or CHF. Co-illness is one of a major risk factor that affects the hypertension management.

**Co-medication:** It is the medication given to the compliant for his/her co-illness.

**Palpitation:** It is an attribute that determine whether the patients have the symptom of palpitation or not. Palpitation is a rapid irregular heart beat.

**Headache:** It is one of the symptoms is shown in the hypertension patients. This attribute used to determine whether the patient has a headache or not.

**Dizziness:** This attribute used to determine whether the patient has dizziness or not.

**Recommendation:** all the above attributes are used to construct the description parts of the hypertension cases. The attribute “Recommendation” is the solution part of the hypertension case, and it is the final decision that the clinician make after he/she analyze all the patient profile, life style and risk factors.

According to the domain expert the major decision that the clinician makes for hypertension treatment is whether to start the drug therapy or not. The reason behind this idea is that mostly if the drug therapy is started it will be taken for life; the probability to stop the drug is rare. According to STGLGH (2010) and domain expert, the decisions that are made for hypertension cases are five as listed below:

- To check for other disease if the compliant is free from hypertension,
- To advice for the patient in order to make life medication only by analyzing severity of the disease, life style and other risk factors of the patient,
- To give first line drug treatment and advice for the patients by analyzing the severity of the disease, life style and other risk factors of the patient,
- To give urgency drug treatment and advice by analyzing the severity of the disease, life style and other risk factors of the patient,
- To admit the patient (emergency) by analyzing the severity of the disease, life style and other risk factors of the patient.

# CHAPTER FOUR

## IMPLEMENTATION

### 4.1. Designing of the Prototype

After case acquisition and its transformation into useful knowledge, the next steps are to codify the knowledge using suitable representational format, and implement the prototype with an appropriate algorithm and tool.

In this research, Nearest Neighbor retrieval algorithm is used to design the retrieval task. The reason that the researcher used is that it retrieves cases which match partially with the new case and it has the advantage of simplicity in retrieving relevant cases (Mishra and Sahu, 2011; Salem et al, 2005; Martin, 1995). It is also suitable when there are attributes that have numeric (continuous) value (Fang and Songdong, 2007). In this research, attribute value for age, systolic blood pressure and diastolic blood pressure are numeric.

The reuse task of the prototype is designed by using voting method. While the revise and retain process is designed by using feedback of domain expert and incremental learning respectively.

Feature-vector case representation approaches are used to build the case base. The reason that the researcher used a Feature-vector approach of case representation is that this approach supports Nearest Neighbor retrieval algorithm and it represents cases in a easy way(Bergmann et al, 2005; Tran and Schönwälder, 2008).

The tool that is used in this research is python 2.6.4. Even though there are different tools that could be used for designing case based reasoning, which had been discussed in section 2.6 of chapter two, some of the tools are proprietary and can not be easily accessible by the researcher. Even though JCOLIBRI is open source software it doesn't allow voting automatically for the reuse task (Alemu, 2010). Python is open source software (Ascher, 2002). As the researcher is familiar with python programming and python is open source software, the researcher uses python 2.6.4.

## 4.2. A CBR Framework for Hypertension Management by Using Nearest Neighbor Retrieval Algorithm

The framework for case based reasoning hypertension management designed in this research is depicted in the figure 4.1 below. The framework deals with retrieval and reuse process of case based reasoning. The explanation of the framework is as follows.

The knowledge engineer prepare the main part of the case based reasoning, case base, with the help of domain experts. In constructing the case base the knowledge engineer, with the help of domain experts, extract attributes that have a direct impact for managing hypertension cases. Based on the selected attributes hypertension cases are collected. A total of 45 cases are used to build the case base of the prototype.

The case base in this research are stored in a form of plaintexts, which is arranged in N columns representing case attributes ( $\{A_1, A_2, A_3, A_4, \dots, A_N\}$ ), and M rows representing individual cases ( $\{C_1, C_2, C_3, C_4, \dots, C_M\}$ ). Consequently, each attribute have a sequence of possible values associated with the case,

$$\text{i.e.} \quad \begin{array}{l} \{V_{11}, V_{12}, V_{13}, V_{14} \dots V_{1N}\} \\ \{V_{21}, V_{22}, V_{23}, V_{24} \dots V_{2N}\} \\ \dots \quad \dots \quad \dots \quad \dots \quad \cdot \\ \{V_{M1}, V_{M2}, V_{M3}, V_{M4} \dots V_{MN}\} \end{array}$$

To implement the case base reasoning system by using Nearest Neighbor algorithm, the weight of the attribute should be assigned in order to differentiate the importance of the attributes. The importance of the attributes (weights) for most problem domains is context dependent. The weights which are assigned to each attribute of the case informs how much attention to pay to the attribute, when computing the similarity measure of a new case with the old cases from the case base (Salem et al, 2005).

In this research, the importance values of the attributes are assigned by consulting domain experts, as domain experts know which attribute is very important in managing hypertension cases. The weight of the attributes range from 1.0 to 0.1. The experts gives the highest values to the systolic blood pressure and diastolic blood pressure. Next higher value is given to the risk factors, pregnancy and fainting. The rest of the attributes have

got the least value. All the list of the case attributes with their weight are shown in the table 4.1

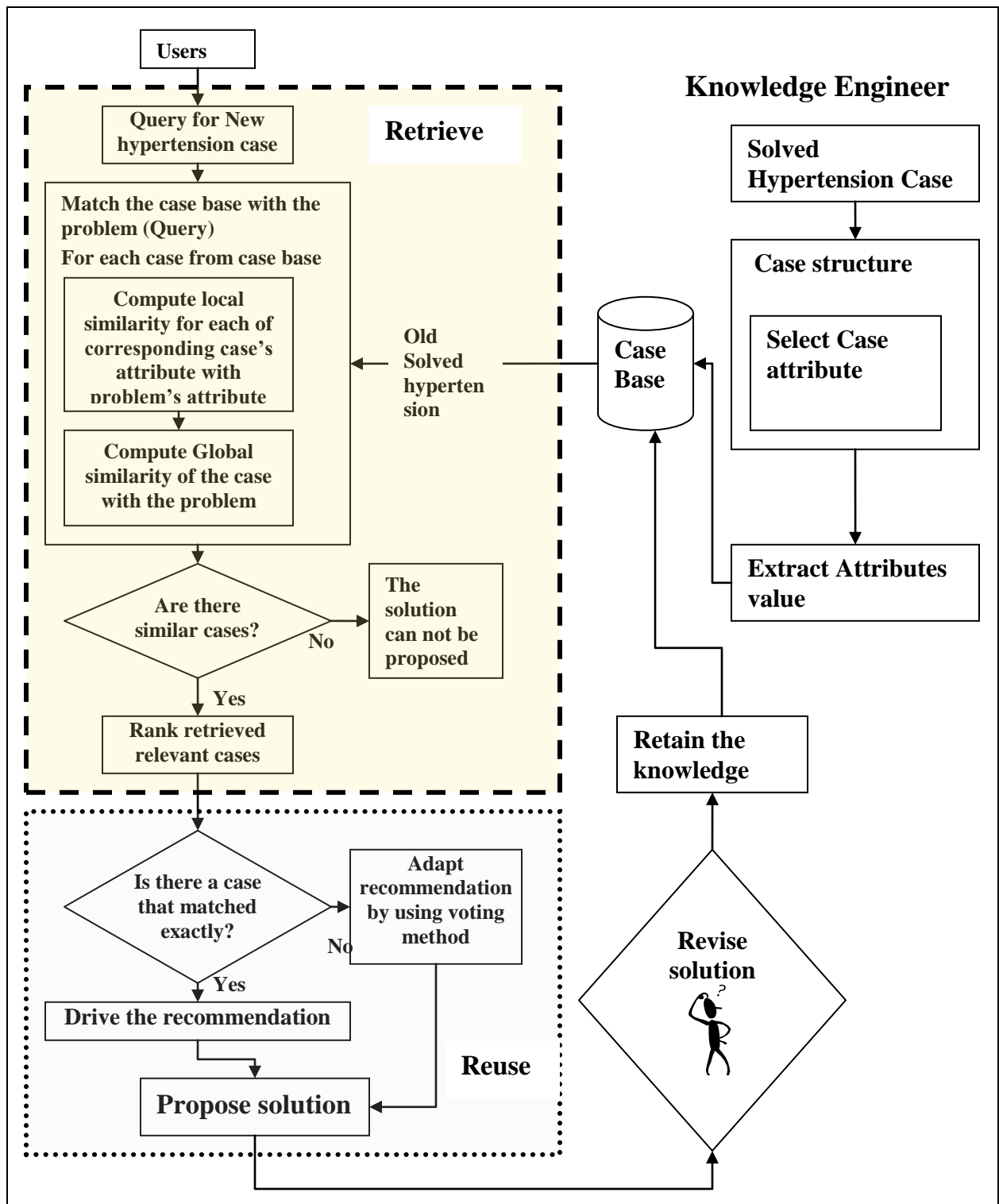


Figure 4.1 Architecture of Case Based Reasoning for Hypertension Management.

The prototype developed in this research works in following manner. As the new case (not solved) is entered, the prototype of the system matches the new case with solved case from the case base of the system by using similarity measurement. If relevant cases from the case base are found, then the prototype rank the relevant retrieved cases based on their global similarity. Next, the prototype proposes a solution. The proposed solution can be derived directly from a retrieved case that matches exactly with the problem of the new case. However when there is uncertainty, i.e. the most relevant retrieved case doesn't have global similarity of 1.0; the solution that will be proposed is selected by using voting technique from the retrieved relevant cases. When the newly solved case is new to the case base, it will be evaluated by domain expert based on the consequence of the proposed solution. Finally, if the newly solved case is valid (verified by the domain experts), it will be add to the case base for future use, i.e the system learn from the successfully solved case.

### **4.3. Designing the Retrieval Process**

To retrieve relevant cases by using Nearest Neighbor algorithm, similarity measurement are computed between a new case and old cases from the case base. Then based on the similarity value the cases are ranked. The next sub sections discuss about similarity measurement and how relevant cases from the case base are ranked.

#### **4.3.1. Similarity Measurement**

Similarity measurement determines how the new case is similar with old case stored in case base. The similarity measurement in Nearest Neighbor case base reasoning is calculated by local and global similarity functions.

##### **4.3.1.1. Local Similarity Measurement**

The local similarity is used to measure the similarity of attributes of the new case with their corresponding attribute value of old case from the case base. The local similarities that are used in this research are shown in table 4.1 below with their corresponding attributes.

<b>Attribute</b>	<b>Data types</b>	<b>Local Similarity</b>	<b>Importance value (Weight)</b>
Age	Integer	Interval	0.2
Sex	Boolean	Equal	0.2
Systolic blood pressure	Integer	Interval	1.0
Diastolic blood pressure	Integer	Interval	1.0
Pregnancy	Boolean	Equal	0.2
Fainting	Boolean	Equal	0.2
Co-illness	String	Equal	0.4
Co-medication	String	Equal	0.4
Smoking	Boolean	Equal	0.2
Alcoholic	Boolean	Equal	0.2
Palpitation	Boolean	Equal	0.1
Headache	Boolean	Equal	0.1
Dizziness	Boolean	Equal	0.1

*Table 4.1 Attributes Description of the Case Structure*

Age, systolic blood pressure and diastolic blood pressure have number value and they show similar characteristics with some interval. For the attribute which has a value of number the interval local similarity function is preferable (Salem et al, 2005; Armengol and Plaza, 2001). As a result of this the local similarity of age, systolic blood pressure and diastolic blood pressure that is used in this research is interval. The algorithm and the sample code in python for interval similarity are shown below.

Algorithm for similarity of interval (Salem et al, 2005; Armengol and Plaza, 2001)

*Step 1 Compute the difference of the two given value of attributes and place it in absolute value.*

*Step 2 Compute the difference of the maximum and minimum values of the given attribute from the case base*

*Step 3 Divide the result of the 1<sup>st</sup> step by the result of the 2<sup>nd</sup> step*

*Step 4 Subtract the result of step 3 from value 1*

*The output of step 4 is the local similarity value of the given attribute of the new case with the old case.*

The algorithm for interval local similarity is summarized in the formula below.

$$\text{Sim}(f_i^l, f_i^R) = 1 - (|f_i^l - f_i^R| / |f_{\max} - f_{\min}|)$$

Where  $f_{\max}$  is the maximum value of the attribute  $i$  in the case base and

$f_{\min}$  is the minimum value of the attribute  $i$  in the case base

$f_i^l$  is the value of attribute  $i$  of the input case (new case)

$f_i^R$  is the value of attribute  $i$  of the case in the case base

As indicated in table 4.1 above the local similarity for Sex, Pregnancy, Fainting, Smoking, Alcoholic, Co-illness, Co-medication, Palpitation, Headache and Dizziness attribute is equal functions. The reason for this is that these attributes needs exact match for determining the similarity between their corresponding values and the value of these attributes are string or Boolean. Equal function is preferable for local similarity of an attribute that has a string value (Salem et al, 2005).

Local similarity Algorithm for equal function (Salem et al, 2005)

*If the two value are equal*

*return 1*

*Else*

*return 0*

#### **4.3.1.2. Global Similarity Measurement**

The global similarity measurement allows measuring the similarity between two cases, i.e. an old case from the case base with new case from the user. This measurement computes the similarity by using all local similarity of attributes which make up the case. It is used to select the most relevant cases.

The algorithm for this measurement is as follows: (Salem et al, 2005; Watson & Marir, 1994).

*Step 1, Find the local similarity for all attributes of the case which make up the case*

*Step 2, Multiply the result of the local similarity of attributes with their corresponding attribute importance value (weight)*

*Step 3, add the value of all attribute results of step 2*

*Steps 4, add all weights of attributes that represent the importance value of the attributes*

*Step 5, divide the result of step 3 by the result of step 4 and the result of this step is the global similarity that represents the degree of match of the old case to the input.*

#### **4.3.2. Ranking Relevant Cases**

Ranking relevant cases is done in order to place the most relevant cases at the first position. As it was indicated in section 4.3.1.2, the sorting is done by the value of global similarity of the old case with the input case. Since Selection sort is easy to implement, it is used in this research.

The algorithm for ranking relevant cases in descending order is as follows:

- 1. Find a case which has a maximum global similarity value from the relevant cases*
- 2. Swap it with the case in the first position for unsorted cases*
- 3. Repeat the steps above for the remainder of the relevant case list (starting at the second position and advancing each time)*

## 4.4. Designing the Reuse Process

As it is indicated in chapter two, the second cycle next to retrieval is reuse. Reuse is a process in case based reasoning that proposes a solution for problem (new case). It proposes recommendation for the new case directly from the old solved case which match exactly. However, when there is uncertainty, i.e. when there are no relevant retrieved cases that match exactly with the new case, adaptation is used. In this research, a voting method is used for the adaptation. This adaptation method is tested by Salem et al (2005) in their research for supporting diagnosis of heart diseases and it proposes solutions correctly.

The algorithm for the reuse is as follows:

- *If there is a solved case which exactly match with the new case then directly use the recommendation of the solved case as a solution for the new case*
- *However if there is uncertainty then voting should be applied and it is done in following manner*
  - *Identify and store if different recommendations have been retrieved from the most closest cases*
  - *Compute how many times each recommendation occurs in the retrieved relevant cases*
  - *Propose the mostly occurred recommendation as a solution*

## 4.5. Designing the learning mechanism for prototype

Case based reasoning systems aims to learn from the solved new case for future use. As it was mentioned in chapter 2, it is done by the revise and retain task of the case based reasoning model. The revise task of the case based reasoning evaluates how good the proposed solution is for the given query (new cases) while the retain task decides whether to merge the new solved case with existing knowledge base or not. In this study, in order to enable the KBS prototype for hypertension management learn from the newly solved cases, a way of getting feedback for the new solved cases from domain experts and an

incremental learning are used to design the revise and the retain task of the prototype. The main aim of the revise and retain tasks of the prototype is to learn successfully solved new cases and use it for the future.

The algorithm used for designing learning mechanism for this prototype is as follow:

- ❖ *If there is a case in the case base that exactly match with the new solved case, then the new case will not be retained in the case base as it is already there*
- ❖ *However if there is no a case from the case base that exactly match with the new solved case (if the solved case is new to the case base)*
  - *The new solved hypertension case will be propose to the domain experts so as to be evaluated based on the consequence of the solution*
  - *If the domain expert verify that the new solved case is a valid one, then the new solved hypertension case will be retained in the case base for future use*

The sample code for the implementation is attached in the appendix VI. The next chapter presents experiments that are conducted to test and evaluate the prototype of the knowledge based system

# CHAPTER FIVE

## TESTING AND EVALUATION

### 5.1. Testing and Evaluation of the Prototype

After the prototype is implemented, experimentations are conducted to test and evaluate the prototype. This chapter presents case similarity testing and evaluation of the prototype. The case similarity testing is conducted to investigate how new cases with cases from the case base are matched and evaluation is conducted to investigate the applicability of the prototype for hypertension management.

### 5.2. Case Similarity Testing

An experiment is made to know how new cases are matched with the cases from the case base. For this experiment, the researcher uses three experimental groups. The first group is made up of cases from the case base. The second group consists of cases which are made by modifying an attribute value of the case from the case base, while the third group is made up of cases which have two modified attribute values. Each test case is presented to the system individually to evaluate the performance of the similarity measures. Table 5.1 below shows the sample of queries that are used in this experiment with their description.

Case 19

*A compliant is a 37 is years old woman and she doesn't have any kind of cardiovascular disease and diabetics. She doesn't have a symptom of headache, dizziness and palpitation. She is conscious. She doesn't smoke cigarette and drink alcohol. However her systolic blood pressure and diastolic blood pressure is 180 and 90.*

Case 44

*A compliant is 39 years old man and don't have any kind of cardio vascular disease and diabetics. He is not a smoker and drinker. He is conscious. He has a symptom of headache. His systolic blood pressure and diastolic blood pressure shows the same reading in different check up. His systolic blood pressure diastolic blood pressure registers 160 and 105 respectively.*

Query	Description of query	With respect to Case
Query1	Same value for all attributes	Case19
Query2	values of attribute “ <b>headache</b> ” is changed	Case19
Query3	values of attributes “ <b>headache</b> and <b>sex</b> ” are changed	Case19
Query4	Same value for all attributes	Case44
Query5	values of the attribute “ <b>co-illness</b> ” is changed	Case44
Query6	values of the attribute “ <b>smoking</b> and <b>co-illness</b> ” are changed	Case44

*Table 5.1 Sample of Query for Case Similarity Testing and their Corresponding Cases*

After the query is provided to the system the similarity of the query with respect to the case is shown in the table 5.2 below.

Query	Degree of Similarity	With respect to case
Query1	1.0	Case19
Query2	0.97	Case19
Query3	0.92	Case19
Query4	1.0	Case44
Query5	0.91	Case44
Query6	0.86	Case44

*Table 5.2 Query Similarity with their Corresponding Cases from the Case Base*

The case similarity test results of this experiment shows that when the test case has attributes value the same as a case stored in case-base, the degree of similarity(global similarity) become 1.0 (exact match) as in query1 and query 4 in Table 5.2. On the other hand, the degree of similarity decreases when there is a change on one or more attributes value of the test case as compared to a case from the case base. When attribute value that has higher importance value (weight) is changed, the degree of similarity decreases highly.

### **5.3. Evaluation of the Retrieval and Reuse process by Using Statistical Analysis**

The statistical analysis evaluation uses the 45 hypertension cases that have been collected from Brook Medical Service Plc and Bole 17 Health Center and a leave-one-out cross validation testing proportion. The leave-one-out cross validation provides an almost unbiased estimate of generalization performance (Cawley and Talbot, 2008), i.e it allows the researcher to test for all cases by making one of the cases as a testing data and the rest of the cases as a training data (case base). Leave-one-out cross validation is common evaluation strategy in case based reasoning (Jagannathan, et al, 2010). The statistical analysis is done for both retrieval and reuse processes.

#### **5.3.1. Evaluation of the Retrieval Process**

The first task of case based reasoning hypertension management is to retrieve cases that are relevant to the new hypertension case, so as to enable users to manage the new hypertension case by analyzing the retrieved cases. In this research, the effectiveness of the retrieval process of the case based reasoning hypertension management is measured by using recall and precision. According to McSherry (2001) precision and recall are useful measures of retrieval performance in case based reasoning. Recall is the percentage of relevant cases for the query (new case) that are retrieved, whereas precision is the percentage of retrieved cases that are relevant to the query (Junker et al, 1999; Losee, 2000; McSherry, 2001).

To do this evaluation, for each test case the relevant hypertension cases from the case-base should be known. Due to this, test cases are given to the domain experts in order to assign possible relevant cases from the case base to each of the test cases. The domain expert uses the value of recommendation attribute of the hypertension case as the main concept to assign the relevant case to the queries, i.e. hypertension cases that have similar solution (recommendation) are relevant to each other. Based on this concept recall and precision are calculated.

Table 5.3 below shows sample test cases with their corresponding relevant hypertension cases that are assigned by the domain experts from the case base.

<b>Test case</b>	<b>Relevant case from the case base</b>
Case21	case1, case3, case8, case10, case17, case19, case29, case32, case34, case36, case40
Case22	case2, case7, case14, case18, case31, case30, case39
Case 23	case16, case35, case33, case42, case44
Case24	Case4, case5, case13, case28,
Case25	case6, Case9, case37, case43
Case26	case38, case45
Case27	case12, case11, case15, case20, case41

*Table 5.3 Relevant Cases Assigned by Domain Expert for the Sample Test Case*

After the relevant cases are assigned the next step is to calculate the precision and recall value of the retrieval performance of the case based reasoning system with a threshold interval.

As to the researcher there is no standard threshold for degree of similarity that has been used for retrieving relevant cases in case based reasoning. Abdel-Badeeh et al (2005) on their research, which is entitled on “A Case Based Expert System for Supporting Diagnosis of Heart Diseases”, uses a threshold similarity of [1.0, 0.5), i.e. the cases that have the global similarity of greater than 0.5 are retrieved. On the other hand, Yemisrach (2010) on her thesis uses a threshold interval of [1.0, 0.8) after she investigate the result

of the retrieval process in three threshold intervals, i.e [1.0, 0.8), [0.8, 0.5), [0.5, 0). As the result of these, the threshold interval for this study is selected by testing two threshold intervals with sample test cases. The two thresholds are [1, 0.8) and [1, 0.5). The result that is found with the thresholds of [1, 0.8) is more better in precision and shows a small reduction in recall value as compared to the result that is found with in thresholds of [1.0, 0.5). Due to this the threshold interval of [1.0, 0.8) is used for this study.

Forty five experiments are conducted to measure recall and precision by using a leave-one-out cross validation testing proportion and [1.0, 0.8) threshold interval. The recall and precision results for the sample test queries are shown in the table 5.4 below

<b>Test cases</b>	<b>Retrieval Performance measure</b>	
	<b>Recall</b>	<b>Precision</b>
Case21	0.82	0.53
Case22	0.86	0.75
Case23	1.0	0.36
Case24	1.0	0.67
Case25	0.75	0.5
Case26	1.0	1.0
Case27	1.0	0.71

*Table 5.4 Recall and Precision Result for the Sample Test Case*

Both the average recall and precision of the experiment register more than average that shows a promising result. The average recall and precision for the prototype in percent is shown in the table 5.5 below.

<b>Recall</b>	<b>Precision</b>
86.1 %	60 %

*Table 5.5 Average Recall and Precision Value of the Retrieval process*

### 5.3.2. Evaluation of the Reuse Process

As the goal of reuse process in this research is to recommend correctly for hypertension cases, i.e. to solve the problem correctly, the performance of the reuse process is measured by using accuracy. Accuracy is one of the useful measurements in case based reasoning (McSherry, 2001). This measurement had been used by Salem et al (2005) on their research. Accuracy is defined as the percentage of the number of correctly recommended cases (Junker et al, 1999).

Since the research uses a leave-one-out cross validation testing proportion, 45 experiments are conducted to evaluate the performance of reuse process of the case based reasoning system prototype for hypertension management. The result shows that the reuse process also registers above average, which is a promising result.

After the reuse task of the system is tested for the 45 cases, summery of the result is shown in the table 5.6 below

<b>Total number of tested case query</b>	<b>Total number of correctly recommended cases</b>	<b>Accuracy</b>
45	40	88.89%

*Table 5.6 Accuracy Value of the Reuse Process*

### 5.3.3. Comparison of the Prototype Performance with Previous Case Based Reasoning System

The performance of the hypertension management prototype system is compared with the previously conducted research in health area. The previous research focuses on retrieval phases of the case based reasoning system and the performance of the reuse phase is not evaluated. Thus, the recall and precision value of retrieval performance of the systems are compared and shown in the table 5.7 below.

Domain area	Tool	Retrieval task		Reuse task
		Recall	Precision	Accuracy
AIDS	JCOLIBRI	72%	63%	Not evaluated
Hypertension	Python	86.1 %	60 %	88.89%

*Table 5.7 Comparison of the Prototype with the Previous CBR system*

The above table 5.7 shows that the result of the precision value of the case based reasoning system of hypertension management is nearly the same with the precision value of system developed for AIDS by Alemu (2010), while the value of recall shows an improvement. The reduction in the precision value could be caused by the composition of the attribute for the case structure of hypertension case. As it was indicated in chapter three of this thesis, the patient record for the hypertension patient does not include attributes like height, weight and race which affect the management of hypertension. As a result of this, these attributes are not included in the case base of the prototype, and it could affect the performance of the prototype.

The other reason that lowers precision value of the hypertension management system could be the tool and its threshold specification that are used in the research. Alemu used JCOLIBRI that retrieve N Nearest cases for a given query. This is to mean that the threshold is not by the degree of similarity of the cases rather it is based on the number of nearest cases to be retrieved. As the number of relevant cases for queries varies, assigning a number for the number of cases to be retrieved could affect the performance.

One of the main objectives in this research is to investigate the applicability of the case based reasoning system in recommending the appropriate solution to the hypertension cases. This deals with the reuse task. And more than an average accuracy is registered which is a promising result. This task is not evaluated in the Alemu (2010) thesis research.

## 5.4. Testing the learning mechanism

The main aim of the learning mechanism of this prototype is to learn from solved new hypertension cases and use it for managing other hypertension cases. This part of the prototype is tested with a case that is collected from Bole 17 Health Center. For the experiment a new hypertension cases is provided to the prototype. The sample of hypertension case that is used for the testing purpose is as follows:

*A compliant was 43 years old man. He was free from diabetics and any cardiovascular diseases. He had a symptom of bad headache. He was conscious. His systolic blood pressure and systolic blood pressure showed the same reading in different check up. The systolic blood pressure was 135 and the diastolic blood pressure is 85. He was a smoker and drinker.*

*The recommendation by the domain expert to the compliant was to make only life style modification. After the compliant comes in other day for check up he become healthy*

After the problem description of the above new hypertension case is fed for the prototype, the prototype computes the similarity of the new hypertension case with the old cases from the case base. The prototype retrieves cases which are considered as a relevant case in a rank order by the retrieval process. Then the prototype proposes a solution to the new case from the retrieved relevant cases. The proposed solution for the new hypertension case is “to make life style modification only”. As this solved hypertension case is new for the KBS prototype of hypertension management, the revise process of the prototype propose the new solved hypertension cases so as to be verified by domain experts for its consequence. The revise process is depicted in the figure 5.1.

```

=====
The recommednation for the case by the system is:

      Control it with Lifestyle Modification only
=====

*****

For the new hypertension case, which is a compliant with
the Age of 43 , Sex m , SBP 135 , DBP 85
who is conscious, without diabetics and cardio vascular disease
who is smoker, who is drinker , who feels headace
The recommendation by prototype is:
      Control it with Lifestyle Modification only

Does the recommendation for the new case is good and has to be retained?
      (yes/no)  :

```

*Figure 5.1 Revise Process for the newly Solved Case*

The domain expert knows the consequence of the proposed recommendation. As the domain expert confirms that the new solved hypertension case is valid, the new solved hypertension case is retained for future use by the retain process of the prototype. This is depicted in the figure 5.2 below.

```

*****

For the new hypertension case, which is a compliant with
the Age of 43 , Sex m , SBP 135 , DBP 85
who is conscious, without diabetics and cardio vascular disease
who is smoker, who is drinker , who feels headace
The recommendation by prototype is:
      Control it with Lifestyle Modification only

Does the recommendation for the new case is good and has to be retained?
      (yes/no)  :yes

      The new solved case has been retained in the case base
      with caseid: case46 for future use

*****

```

*Figure 5.2 Retaining Process of the newly Solved Case for the future use*

In order to check the learning mechanism, the problem description of the above hypertension case is again fed to the prototype. The prototype propose a solution for the case, but the revise task doesn't propose the solved hypertension case to be verified, this shows that the KBS prototype for hypertension management learn from successfully solved hypertension cases (previous experiment) and uses it in solving other hypertension cases.

## **5.5. User Evaluation**

For the user evaluation, domain experts are selected from Bole 17 Health Center and Brook Medical Service Plc by using purposive sampling. The selection is based on the academic qualification in the domain area, work experience, willingness and participation in this research. The domain experts are grouped into two groups comprising five experts each. The first group consists of evaluators who have participated in the knowledge acquisition and different phases of the research activities. The second group consists of experts from the same domain area; however, they did not participate in the research before. Each group consists of 1 medical doctor, 3 health officers and 1 nurse. As a result of this a total of ten domain experts have participated in the user evaluation.

All the participants feed hypertension cases (query) to the case base reasoning system and look up the result of the system for the given queries. Then they evaluate the performance of the system in terms of adequacy and clarity of advising, relevancy of the retrieved cases in the decision making process, rank of the retrieved relevant cases, fitness of the final solution to the new case at hand, ease of use, relevance of the attributes in representing the hypertension case, and speed of the system. These evaluation parameters of the case based reasoning hypertension management system are adapted from Buchanan and Forsythe (1991), Lee et al (2008), Alemu (2010) and Yemisrach (2010) with some modification within the context of this study. The domain expert evaluation form is found in appendix II.

The domain experts assign values (excellent=5, very good=4, good=3, fair=2, and poor=1) to the parameters used in the evaluation of the system. The summary of domain

expert evaluation of group one (who had been participated in the knowledge acquisitions) is shown in the table 5.8 below.

<b>Criteria of evaluation</b>	<b>Poor</b>	<b>Fair</b>	<b>Good</b>	<b>Very good</b>	<b>Excellent</b>	<b>Average</b>
Adequacy and Clarity of advising			2	3		3.6
Relevancy of the retrieved cases in the decision making process			2	3		3.6
Does the method propose relevant cases in a proper rank				3	2	4.4
Fitness of the final solution to the case at hand			1	2	2	4.2
Easy of use				2	3	4.6
Relevance of the attributes in representing the hypertension case				4	1	4.2
is the method efficient in time			1		4	4.6
	Average					4.17

*Table 5.8 Performance Evaluation by Group one Domain Experts*

As shown in the table 5.8 above, 60% of the evaluators, who have prior knowledge of the system, rate the parameters of ‘Adequacy and Clarity of advising’ and ‘relevancy of the retrieved cases’ as very good, while the remaining of them rate both parameters as good. On the other hand, 60% and 40% of the evaluators evaluate the parameter of ‘the rank of the retrieved relevant cases’ as very good and excellent, respectively.

In addition 40 % of the evaluators respond as very good for the parameters of both ‘Fitness of the final solution’ and ‘easy of use’, while 60% of respondents rate ‘ease of use’ as excellent. 40% of respondents rate the parameter of ‘Fitness of the final solution’ as excellent and the remaining of them (20%) rates it as good. Finally 80% of the evaluators rate the parameters of ‘relevance of the attributes’ and ‘efficiency in time’ as very good and excellent respectively, while the remaining of them rate the parameters of

‘efficiency in time’ and ‘relevance of the attributes’ as good and excellent respectively. The over all performance of the prototype as evaluated by group one is 4.17 out of 5.

In addition, table 5.9 below shows the performance of the prototype as evaluated by the domain experts who did not participate in the thesis research (group two).

<b>Criteria of evaluation</b>	<b>Poor</b>	<b>Fair</b>	<b>Good</b>	<b>Very good</b>	<b>Excellent</b>	<b>Average</b>
Adequacy and Clarity of advising			3	2		3.4
Relevancy of the retrieved cases in the decision making process		1	2	2		3.2
Does the method propose relevant cases in a proper rank			1	4		3.8
Fitness of the final solution to the case at hand				3	2	4.4
Easy of use			1	3	1	4
Relevance of the attributes in representing the hypertension case		1	1	3		3.4
is the method efficient in time				2	3	4.6
	Average					3.82

*Table 5.9 Performance Evaluation by Group two Domain Experts*

As shown in table 5.9 above, 40% of the evaluators, who have no prior knowledge of the system, rate the parameters of ‘Adequacy and Clarity of advising’ and ‘relevancy of the retrieved cases’ as very good, while the remaining of them rate the parameter of ‘Adequacy and Clarity of advising’ as good. 40% and 20% of the evaluators rate the parameter of ‘relevancy of the retrieved cases’ as good and fair, respectively. On the other hand, 80% and 20% of the evaluators rate the parameters of ‘the rank of the retrieved relevant cases’ as very good and good respectively.

Moreover, 60 % of the evaluators rate the parameter of ‘Fitness of the final solution’ as very good, while the remaining of them (40 %) rates the parameter as excellent.

Similarly 60% of the respondents respond both parameters of 'Easy of use' and 'relevance of the attributes' as very good, while 20% of evaluators rate both parameters as good. Besides, 20% of the respondents rate the parameters of 'easy of use' as excellent, while the same percent of evaluators (20%) rate parameters of 'relevance of attribute' as fair. Finally, 40% of the evaluators respond to the parameter of 'efficiency of time' as very good while the rest of them (60%) rate the parameter as excellent. The overall performance of the prototype as evaluated by group two domain experts is 3.82 out of 5.

The performance evaluation results of the two groups show a little variation, the cause of which could be due to the differences in the awareness of the experts about how the hypertension prototype system works, how the attributes are selected and weighted, and how the case for the system is constructed. Domain experts in group one have better awareness about the prototype compared to domain experts in group two, as they have participated in knowledge acquisitions and other activities of the thesis research.

As compared to the user acceptance of previously conducted research in health area, the prototype which is developed in this study achieves an encouraging result. For example; Alemu jorgi conducted a research for HIV/AIDS. The result of user evaluation for the prototype is nearly very good when evaluated by domain experts who have prior knowledge about the system, while the system register an encouraging result, which is more than good, when evaluated by domain experts who did not have prior knowledge about the system.

In general, the domain experts (evaluators) assign more than average value for all parameters. This shows that the hypertension management prototype achieved promising user acceptance.

## **5.6. Discussion**

The case similarity testing shows that when the query is made up of attribute values that have the same value with the case from the case base, the result of the global similarity becomes 1.0. But when there is a difference in the attribute values of the query and the

case in the case base, the global similarity value decrease. Therefore adding cases in the case base improve the performance of case based reasoning system in solving problems (new cases).

The average recall and precision values for the retrieval performance of the case based reasoning system for hypertension management are 86.1% and 60% respectively. This indicates that the prototype provides a high percentage of the relevant cases for query which enable the user to diagnose and treat the new hypertension case. Similarly the reuse part of the prototype for hypertension management registers 88.89% accuracy. This is an encouraging result as the prototype system has given a high percentage of appropriate recommendations for the new hypertension cases. This shows that a promising result is achieved in the study.

The reasons that the prototype couldn't achieve 100% retrieval and reuse performance could be due to the data and the algorithm used to develop the prototype. As it was indicated in chapter three, attributes like race, weight and height, which are considered in hypertension management are not recorded in the complaints' record file and are not included in the case base. This could affect the performance of the prototype. The Nearest Neighbor algorithm, which is used to develop the retrieval process of the prototype, uses distance to compute the similarity between the query and cases by representing the cases in N dimension vector. However the recommendation for the hypertension cases doesn't have clear boundaries as it has subjectivity and depends on the experience of the domain experts. In addition, the importance value that are assigned to the attributes of the case structure are done manually with the help of the domain experts, as there is no research that is conducted for the importance value of the attributes in hypertension management. This could affect the result of the retrieval and reuse performance of the prototype.

The performance of the retrieval process and reuse process of the prototype can improve, if all the attributes are included in the research or a way of mechanism that assigns an importance value to the attribute is integrated to the prototype. Since most of the time weight, height and other related attributes are not recorded in the complaints' record file, a hybrid of rule based and case based reasoning can be applied for the future. Adding

other solved hypertension cases to the case base can also improve the performance of the prototype, as it was suggested for the case similarity testing above.

The over all user evaluation for the case based reasoning hypertension management prototype is very good. This shows that the prototype achieved an encouraging result from the perspective of domain experts.

In general, the case based reasoning approach in designing hypertension management system shows an encouraging result for retrieving relevant cases and proposing solution so as to diagnosis and treat new hypertension cases. It also attain promising user acceptance as it is evaluated by the domain experts.

## **5.7. Sample of the Experiment**

A compliant comes with the age of 50 to Brook Medical Service Plc. He has been treated for diabetics with a drug of glibonclimid. He has a symptom of bad headache but he is conscious. His systolic blood pressure and systolic blood pressure shows the same reading in different check up. The systolic blood pressure is 170 and the diastolic blood pressure is 100. He does not smoke cigarette and drink alcohol.

Question. Is the compliant a hypertensive patient? If he is a hypertensive patient what kind of hypertension treatment should be applied?

From the above case description the following facts are extracted and fed as the hypertension case query to the prototype. The query form is found in the Appendix III

- The complaint's age is 50 and sex male
- He didn't show a sign of unconsciousness
- The systolic blood pressure is 170 and diastolic blood pressure is 100
- The compliant has co-illness of diabetics and take drug of glibonclimid
- He is not smoker and alcoholic
- He has a symptom of headache

**Recommendation by the system:** After the above facts of the new case are fed for the prototype, the prototype computes the similarity of the hypertension case with the old cases from the case base. The prototype display cases which are considered as a relevant case by the retrieval process of the prototype in a rank order. Then the prototype proposes a solution to the new case from the retrieved relevant cases. The prototype decide that compliant is a hypertension patient and the patient has to make life modification as well as start first line hypertension drug by considering his co-morbidity.

The recommendation proposed by the system is correct as compared to the physician recommendation for the compliant at Brook Medical Service Plc. The results of the retrieval and reuse process are presented in the appendix IV and appendix V respectively.

## CHAPTER SIX

### CONCLUSION AND RECOMMENDATIONS

#### 6.1. Conclusion

Case based reasoning is a new technology which enables us to design an intelligent agent that reasons and makes decision from the past solved cases. A research on such kind of approach is important in any domain particularly medical area, as the experts use their experience to solve new cases. As compared to rule based reasoning, case based reasoning can work with new cases that match partially to the case from the case base. However, rule based reasoning can not solve a problem that doesn't exactly match with the rule of the system. This shows that rule based reasoning works in closed assumption where every fact are known and represented. As a result of this rule based reasoning performs weakly in area where experts use their experience in solving problems.

This research is conducted in order to investigate the case based reasoning approach for hypertension management, particularly in providing relevant cases and proposing solutions to the new hypertension case from already solved hypertension cases.

For the purpose of the study the required knowledge is acquired from domain experts, previous solved hypertension cases and other relevant documents through semi-structured interview and document analysis. The domain experts are selected from Bole 17 Health Center and Brook Medical Service Plc by using purposive sampling. Hierarchical tree modeling technique is employed in order to model the acquired knowledge. Then case structure, which includes attributes that have direct impact on the hypertension management, is formulated to build the case base. The case base is constructed with 45 hypertension cases from Brook Medical Service Plc and Bole 17 Health Center.

The prototype is developed by using python 2.6. The retrieval and reuse task of the prototype uses Nearest Neighbor retrieval algorithm and voting method respectively. Feedback of domain expert and incremental learning are used to design the revise and retain task of the prototype respectively.

The case representation approach that is used in the study is features vector approach. The prototype is tested and evaluated in order to investigate its performance in providing relevant case and proposing solution for the problem (new hypertension case). The evaluation is conducted through statistical analysis (recall, precision and accuracy) and user evaluation.

The evaluation results show that the case based reasoning prototype for hypertension management is encouraging as retrieval performance of the prototype registers an average value of 86.1% recall and 60% of precision, while its reuse performance register an average value of 88.89% accuracy. The domain experts (evaluators) assign more than average value for all parameters that are used in the user evaluation form for the prototype. This shows that the prototype achieves an encouraging result from domain expert side in retrieving a ranked order of relevant cases, as well as in proposing a solution to new hypertension cases. More over the prototype achieved promising result for its speed and easiness to use from the perspective of domain experts. Therefore integrating case based reasoning approach in medical services for hypertension management can helps to improve the hypertension management and solves the shortage of domain expertise in the country.

The research demonstrates the case based reasoning approach as the primary reasoning technique for the purpose of diagnosing and treating hypertension cases. This may motivate others to investigate a case based reasoning based approach to the management of other chronic diseases. This will help countries and societies in solving the shortage of experts in the medical domain.

In general, the study achieves its objective by developing the prototype with promising performance and user acceptance, and demonstrating case based reasoning approach in designing knowledge based system for hypertension management.

## 6.2. Recommendations

The study achieves its objective by demonstrating case based reasoning approach in designing knowledge based system for hypertension management. However, there are problem areas that need further investigation and the researcher recommends the following issues as a future research direction based on this study.

- One of the challenging tasks in developing case based reasoning is to construct the case structure and store the case in the case base. In this research the case structure and the case base are constructed manually from texts with the help of domain experts. A future research direction can be to investigate the applicability of natural language processing (NLP) in constructing the case base for the case based reasoning approach.
- In this research, the importance values of the attributes are assigned manually with the help of domain experts. The performance of the retrieval process and reuse process can be improved if it is done with the machine learning algorithms automatically. A possible future work direction can be to investigate and integrate machine learning algorithm for assigning weight for the attributes.
- This study investigated a pure case based reasoning approach for hypertension management. Further researches can be conducted by integrating other approach like rule based reasoning with the aim of improving the performance of the knowledge based system.
- In this study the management of drug response to the hypertension patient was not included. Further investigation can be conducted in order to improve the applicability of the prototype on hypertension management. Generally, medical domain is an active area that should be investigated by case based reasoning, as domain experts use their experience in managing diseases. A future work can be conducted in other chronic diseases so as to solve the shortage of domain experts.

## References

- Aamodt, A. and Plaza, E. (1994).** Case-Based Reasoning: Foundational Issues, Methodological Variations, and System Approaches. *AICom: Artificial Intelligence Communications*, vol. 7, issue 1, pp. 39-59
- Abdelhamied, K., Hafez, S., Abdalla, W., Hiekal H. and Adel A. (1988).** A Rule-Based Expert System for Rapid Problem Solving in Crowded Outpatient Clinics in Egypt. *IEEE*, vol. 3, pp. 1419 – 1420
- Alemu, J. (2010).** A Case-Based Approach for Designing Knowledge-Based System for Addis Resource Center (Arc): The Case of Warmline Clinician Consultation Service. Msc Thesis, Addis Ababa University, Ethiopia
- Anteneh, W. (2010).** Design and Development of Prototype Knowledge-Based System for Antiretroviral Therapy. Msc Thesis, Addis Ababa University, Ethiopia
- Armengol, E. and Plaza, E. (2001).** Similarity Assessment for Relational CBR. In *Case-Based Reasoning Research and Development (ICCBR)*, Lecture Notes in Artificial Intelligence, vol. 2080, pp. 44-58
- Ascher, D. (2002).** The Python Software Foundation: A Primer. *The Perl Review*, vol. 0, issue 2, pp. 5-11
- Ashraf, M.H. and Iqbal, N. (2006).** Evaluation of JCOLIBRI, Msc Thesis, Malardalen University, Sweden. Available at  
[\[http://www.codeproject.com/KB/architecture/commonkadsx.aspx\]](http://www.codeproject.com/KB/architecture/commonkadsx.aspx) accessed on March 3, 2011
- Bergmann, R., Kolodner, J. and Plaza, E. (2005).** Representation in Case-Based Reasoning. *The Knowledge Engineering Review*, vol. 20, pp. 1–4

**Binjie, L. (2010).** Research of Knowledge Based Expert System Used in Maternity Diagnosis. International Conference on Computer Application and System Modeling. vol. 11, pp.405-408

**Brooks, R.A. (1991).** Intelligence without Representation. Artificial Intelligence, vol. 47, pp. 139–159

**Buchanan, B.G. and Duda, R. (1982).** Principles of Rule-Based Expert Systems. In M. Yovits (Ed.), Advances in Computers, vol. 22, New York: Academic Press, Stanford university

**Burge, J.E. (1998).** Knowledge Elicitation for Design Task Sequencing Knowledge. Msc Thesis, Worcester Polytechnic Institute, Available at

[\[http://www.wpi.edu/Pubs/ETD/Available/etd-101399-23113/unrestricted/jburge.pdf\]](http://www.wpi.edu/Pubs/ETD/Available/etd-101399-23113/unrestricted/jburge.pdf) accessed on March 10, 2011

**Carson, E.R. and Cramp, D.G. (2002).** The Role of Health Care Modelling in the Development of Knowledge-Based Systems for Chronic Disease Management. Engineering in Medicine and Biology Society, Proceedings of the 15th Annual International Conference of the IEEE, pp. 606-607

**Clancey, W.J. (1984).** Knowledge Acquisition for classification expert systems. Proceedings of ACM Conference on Computer Science, pp. 11-14

**DeTore, A.D. (1989).** An Introduction to Expert Systems. Journal of insurance Medicine, vol. 21, pp. 233-236

**Evans, S. (1990).** The PACE Nursing Consultant System: A System for Decision Support and Clinical Care Planning. Proceedings of the Annual Symposium on Computer Application in Medical Care, pp. 103-104

**Fang, H. and Songdong, J. (2007).** Case-Based Reasoning for Logistics Outsourcing Risk Assessment Model. Proceedings of International Conference on Enterprise and Management Innovation, pp.1133-1138

**Fogel, D.B. (2005).** Evolutionary Computation: Toward a New Philosophy of Machine Intelligence. Third Edition, New Jersey, John Wiley & Sons, Inc

**Forsythe, D.E. and Buchanan, B.G. (1991).** Broadening Our Approach to Evaluating Medical Information Systems. Proceedings of the Annual Symposium on Computer Application in Medical Care, pp. 8-12

**Fredlund, M.D., Sillers, W.S., Fredlund, D.G., Wilson, G.W. (1996).** Design of a knowledge-based system for unsaturated soil properties. Third Canadian Conference on Computing in Civil and Building Engineering, pp. 659-677

**Getahun, W., Gedif, T. and Tesfaye, F. (2010).** Regular Khat (Catha Edulis) Chewing is Associated with Elevated Diastolic Blood Pressure Among Adults in Butajira, Ethiopia: A Comparative Study. BMC Public Health , vol. 10

**Holt, A. and Benwell, G.L. (1999).** Applying Case-Based Reasoning Techniques in GIS. International Journal of Geographical Information Science, vol. 13, issue 1, pp. 9-25

**Jones, P.H. (1989).** Knowledge Acquisition: Knowledge Engineering in Agriculture. ASAE Monograph, ASAE, St. Joseph, MI, No.8, Available at <https://engineering.purdue.edu/~engelb/abe565/knowacq.htm> accessed date, March 21, 2011

**Junker, M., Hoch, R. and Dengel, A. (1999).** On the Evaluation of Document Analysis Components by Recall, Precision, and Accuracy. Proceedings of the Fifth International Conference on Document Analysis and Recognition, pp.713-716

**Kolodner, J. (1992).** An Introduction to Case-Based Reasoning. Artificial Intelligence Review, vol. 6, pp. 3-34.

**Kyung-Ho L. and Dongkon, L. (1999).** An Approach to Case-Based System for Conceptual Ship Design Assistant, Expert Systems with Applications, vol. 16, pp. 97-104

**Lee, K.C., Kirlidog, M., Lee, S. and Lim, G.G. (2008).** User Evaluations of Tax Filing Websites. *Online Information Review*, Vol. 32, pp. 842-859

**Lekoubou, A., Awah, P., Fezeu, L., Sobngwi, E. and Kengne, A.P. (2010).** Hypertension, Diabetes Mellitus and Task Shifting in Their Management in Sub-Saharan Africa. *International Journal of Environmental Research and Public Health*, vol. 7, pp. 353-363

**Lopez, R., Mcsherry, D., Bridge, D., Leake, D., Smyth, B., Craw, S., Faltings, B., Maher, M., Cox, M., Forbus, K., Keane, M., Aamodt, M. and Watson, I. (2006).** Retrieval, Reuse, Revision, and Retention in Case-Based Reasoning. *The Knowledge Engineering Review*, vol. 20, issue 3, pp. 215-240

**Losee, R.M. (2000).** When Information Retrieval Measures agree about the Relative Quality of Document Rankings. *Journal of the American Society for Information Science*, vol. 51, pp. 834-840

**Lotfy- Abdrabou, E.M. and Salem, A.M. (2010).** A Breast Cancer Classifier based on a Combination of Case- Based Reasoning and Ontology Approach. *Proceedings of the International Multiconference on Computer Science and Information Technology*, pp. 3–10

**Luan, X., MacArthur, R.D., Cohn, J.A., Crane, L.R., Lin, F. and Barth-Jones, D.C (2005).** A fuzzy Discrete Event System for HIV/AIDS Treatment. *IEEE International Conference on Fuzzy Systems*, Departments of Electrical and Computer Engineering, Wayne State University Detroit, MI 48202, USA

**Lundgren-Cayrol, K., Paquette, G., Miara, A., Lundgren-Cayrol, K., Paquette, G., Miara, A., Bergeron, F., Rivard, J., Rosca, I. (2001).** Explor@ Advisory Agent: Tracing the Student's Trail. *Proceeding of WebNet'2001, World Conference of the WWW and Internet*, pp. 802–808

**Martin, B. (1995).** Instance-Based Learning: Nearest Neighbour with Generalisation. Msc Thesis , University of Waikato, Available at

[<http://www.cs.waikato.ac.nz/pubs/wp/1995/uow-cs-wp-1995-18.pdf>] accessed on March 27, 2011

**Masizana-Katongo, A.N., Leburu-Dingalo, T.K., Mpoeleng, D. (2009).** An Expert System for HIV and AIDS Information. Lecture Notes in Engineering and Computer Science, Newswood and International Association of Engineers, vol. 2176, issue 1, pp. 88-92

**Matsopoulou, G.K., Koulouliasa, V., Asvestasa, P., Mouravlianskya, N., Delibasis, K. and Demetriades, D. (2004).** MITIS: a WWW-based Medical System for Managing and Processing Gynecological— Obstetrical— Radiological data. Computer Methods and Programs in Biomedicine, vol. 76, pp. 53-71

**Makhfi, P. (2011).** Introduction to knowledge modeling and neural network. Available at [http://www.makhfi.com/KCM\\_intro.htm](http://www.makhfi.com/KCM_intro.htm) accessed date, March 27, 2011

**McSherry, D. (2001).** Precision and Recall in Interactive Case-Based Reasoning. In Case-Based Reasoning Research and Development (ICCBR), Lecture Notes in Artificial Intelligence vol. 2080, pp. 392–406

**Miller, A.W. (2009).** Problem Detection for Situation Assessment In Case-Based Reasoning For Diabetes Management. Msc Thesis, The Russ College of Engineering and Technology, Ohio University, Available at

<http://etd.ohiolink.edu/send-pdf.cgi/Miller%20Wesley%20A.pdf?ohiou1245162949> accessed date, March 16, 2011

**Mishra, D. and Sahu,B., (2011).** Feature Selection for Cancer Classification: A Signal-to-noise Ratio Approach. International Journal of Scientific & Engineering Research, vol. 2, Issue 4, pp. 1-7

**Montani, S. and Bellazzi, R. (2001).** Integrating Case Based and Rule Based Reasoning in a Decision Support System: Evaluation with Simulated Patients. Proceedings of the Annual Symposium of American Medical Informatics Association, pp. 887-891

**Prentzas, J. and Hatzilygeroudis, I. (2003).** Integrations of Rule-Based and Case-Based Reasoning. Proceeding of the International Conference on Computer, Communication and Control Technologies, vol. IV, pp. 81-85

**Prentzas, J. and Hatzilygeroudis, I. (2007).** Categorizing Approaches Combining Rule-Based and Case-Based Reasoning. Expert Systems, vol. 24, pp. 97-122

**Recio-Garcia, J.A. and Daz-Agudo, B. (2004).** An Introductory User-Guide to JCOLIBRI 0.3. Department of Informatics, university of Madrid, Spain, Technical Report No 144/2004

**Redit, A. (2006).** Design and Development of a Prototype Knowledge Based System for HIV Pre-Test Counseling. Msc Thesis, Addis Ababa University, Ethiopia

**Russell, S.J. and Norvig, P. (2003).** Artificial Intelligence: A Modern Approach. 2nd edition, New Jersey, prentice hall, Inc

**Sagheb-Tehrani, M. (2009).** A Conceptual Model of Knowledge Elicitation. Proceedings of the Conference on Information Systems Applied Research, vol. 2, Washington DC

**Sajja, P.S. and Akerkar, R. (2010).** Knowledge-Based Systems for Development. Advanced Knowledge Based System: Model, Application & Research, vol. 1, pp.1-11

**Salem, A.M. (2007).** Case-based Reasoning Technology for Medical Diagnosis. World Academy of Science, Engineering and Technology, vol. 31, pp. 9-13

**Salem, A.M., Roushdy, M. and HodHod, R.A. (2005).** A Case Based Expert System for Supporting Diagnosis of Heart Diseases. International Journal on Artificial Intelligence and Machine Learning, vol. 5, pp.33-39

**Schmidt, R., Montani, S., Bellazzi, R. and Gierl, L. (2001).** Cased-Based Reasoning for Medical Knowledge-Based Systems. International Journal of Medical Informatics, Vol. 64, pp.355–367

**Shadrick S.B., Iussier, J.W. and Hinkle, R. (2005).** Concept Development for Future Domains: A New Method of Knowledge Elicitation., U.S Army Research Institute for the Behavioral and Social Sciences 2511, Technical Report No 1167

**Singh, Y., Bhatia, P.K. and Sangwan, O. (2007).** A Review of Studies on Machine Learning Techniques. International Journal of Computer Science and Security, vol.1, issue 1, pp. 70-84

**SOH, (2011).** Syptom-Of-Hypertension. Available at [\[Http://Www.Symptoms-Of-Hypertension.Com\]](http://www.Symptoms-Of-Hypertension.Com) accessed on February 16, 2011

**Speel, P.H., Schreiber, A.T., Joolingen, W.V., Heijst, G.V. and Beijer, G.J. (2001).** Conceptual Modeling for Knowledge-Based Systems. In the Encyclopedia of Computer Science and Technology, vol. 44, pp. 107–132

**STGLGH. (2010).** Standard Treatment Guide line for General Hospitals. Drug Administration Control Agency, Ethiopia

**Tan, C. (2008).** A Prototype of Knowledge-Based System for Fault Diagnosis in Automatic Wire Bonding Machine. Turkish Journal of Engineering and Environmental Sciences, vol 32, pp. 235 – 244

**Tesfaye, F., Byass, P. and Wall, S. (2009).** Population Based Prevalence of High Blood Pressure among Adults in Addis Ababa: Uncovering a Silent Epidemic. BMC Cardiovascular Disorder, Vol. 9

**Tran, H.M. and Schönwälder, J. (2008).** Fault Resolution in Case-Based Reasoning. PRICAI: Trends in Artificial Intelligence, vol. 5351, pp. 417-429

**Watson, I. & Marir, F. (1994).** Case-Based Reasoning: a Review. The Knowledge Engineering Review, vol. 9, issue 4, pp 327-354

**Whitworth, J.A. (2003).** World Health Organization (Who)/International Society of Hypertension (Ish) Statement on Management of Hypertension. Journal of Hypertension, vol. 21, pp. 1983-1992

**Yemisrach Hailemariam. (2010).** Application of Case Based Reasoning in Legal Knowledge Based System: A Prototype on Children Criminal Cases In Ethiopia. Msc Thesis, Addis Ababa University, Ethiopia

# Appendixes

## Appendix I

### The main Interview Questions used for domain experts

After the interviewer introduces the objective of the study and confirms the willingness of the experts, the interviewer asks them semi structured questions that are useful for the study. The interviewer recorded the answer of the experts by using paper and pen. The following are sample questions that were used for the interview with domain experts.

1. What is hypertension disease?
2. How many types of hypertension disease are found and which one of them is the most prevalent?
3. Does the hypertension disease have stages? If it has, what are they and by what measurement they are differentiated?
4. What attributes are considered by the clinician in order to identify whether the compliant has the disease or not?
5. If the compliant is a hypertension patient, what things are considered by the clinician in order to manage the disease?
6. What are the main risk factors for hypertension disease?
7. What are the main decisions that the clinician make in hypertension treatment?
8. Which attribute are the most important attribute in diagnosing the disease that the clinician should focus?
9. What are the symptoms for hypertension disease and which one of them are relatively common?
10. What is the goal of the hypertension management?

## Appendix II:

Evaluation form found below is a form used in the study to evaluate the performance of the prototype by the domain expert

### **Domain Expert Evaluation Form**

This is an evaluation form to be filled by health professionals in order to assess the applicability of the case based reasoning system in hypertension management. I thank you in advance for your willingness and valuable time.

Description of the parameter values are as follows

Performance value	1	2	3	4	5
Description	Poor	Fair	Good	Very Good	Excellent

**Instruction:** Please, tick on the appropriate value for the corresponding parameter of the case based reasoning system in hypertension management.

Parameter	Performance value				
	1	2	3	4	5
Adequacy and Clarity of advising					
Relevancy of the retrieved cases in the decision making process					
Does the method propose relevant cases in a proper rank					
Fitness of the final solution to the case at hand					
Easy of use					
Relevance of the attributes in representing the hypertension case					
is the method efficient in time					

### Appendix III:

Query form for new hypertension case

```
#####  
#  
#          CASE BASED REASONING HYPERTENSION MANAGEMENT          #  
#                                     Copy Right @ 2011, H.B #  
#  
#####  
  
Enter age of the compliant in year:   50  
  
Enter sex of the compliant:    m  
  
Is the compliant unconscious/deblitaded (yes/no):    no  
  
Enter the co-illness if the compliant has(diabetics or CVC):  diabetics  
  
Enter co-medication:  glibonclimid  
  
Is the compliants a smoker (yes/no):  no  
  
Is the compliants an alcholic (yes/no):    no  
  
Has the compliant a headache (yes/no):    yes  
  
Does the compleant feel dizzy (yes/no):    no  
  
Has the compliant a symptom of palpitation (yes/no):  no  
  
Enter the compliant's syastolic blood pressure:    170  
  
Enter the compliant's diastolic blood pressure:    100
```



## Appendix V:

Retrieved relevant case and the proposed recommendation by the prototype for sample query.

```
With case39
Has similarity of 0.815411117737
A person with the Age of 76 , Sex f , SBP 140 , DBP 90 ,
who was conscious, who had diabetics and took a comedication of glibonclimid
who was not smoker , who was not drinker , who felt headace diziness
The phycician decision was
    Start first line drug by considering compliants' comorbidity and
    make life style modification

With case4
Has similarity of 0.815146094216
A person with the Age of 50 , Sex m , SBP 110 , DBP 80 ,
who was conscious, who had diabetics and took a comedication of glibonclimid
who was not smoker , who was not drinker , who felt headace
The phycician decision was
    The compliant is at normal stage
    However; the compliant is at higher risk for hypertension
    Check for other disease

With case31
Has similarity of 0.810706950242
A person with the Age of 69 , Sex m , SBP 160 , DBP 90 ,
who was conscious, who had CHF and took a comedication of digoxin
who was not smoker , who was not drinker , who felt headace palpitation
The phycician decision was
    Start first line drug by considering compliants' comorbidity and
    make life style modification

With case28
Has similarity of 0.804710793083
A person with the Age of 50 , Sex m , SBP 115 , DBP 75 ,
who was conscious, who had diabetics and took a comedication of glibonclimid
who was not smoker , who was not drinker , who felt headace
The phycician decision was
    The compliant is at normal stage
    However; the compliant is at higher risk for hypertension
    Check for other disease

=====
The reconmednation for the case by the system is:

    Start first line drug by considering compliants' comorbidity and
    make life style modification
=====
```

## Appendix VI:

Sample code in python

Sample code for local similarity measurement is as follows:

```
# calculating the interval of maximum and minimum value of systolic blood
pressure
size=len(indexterms)
minsbp=int(indexterms[0][3])
maxsbp=int(indexterms[0][3])
for i in range(size):
    if int(indexterms[i][3])<minsbp:
        minsbp=int(indexterms[i][3])
    if int(indexterms[i][3])>maxsbp:
        maxsbp=int(indexterms[i][3])
sbpinterval=maxsbp-minsbp
return sbpinterval
```

where *indexterms* is a nested list that holds all the case from the casebase

*size* is the number of case in the casebase

*indexterms[i][0]* is the value of systolic blood pressure for case *i*

# calculating local similarity of systolic blood pressure

```
def simsbp(csbp,psbp,sbpinterval):
    if int(csbp)<int(psbp):
        ssbp=1-((int(psbp)-int(csbp)+0.0)/sbpinterval)
    if int(csbp)>=int(psbp):
        ssbp=1-((int(csbp)-int(psbp)+0.0)/sbpinterval)
    return ssbp
```

where *csbp* is the systolic blood pressure a case from case base

*psbp* is the systolic blood pressure of new case (the problem)

*sbpinterval* is the interval of maximum and minimum interval of systolic blood pressure from the casebase.

Sample code in python for local similarity of sex attribute:

```
def simsex(csex,psex):  
    if csex==psex:  
        return 1  
    if csex!=psex:  
        return 0
```

where *csex* is the sex of a case from the case base

*psex* is the sex of the new case(patient)

Sample code in python for global measurement is as follows:

```
sage=simage(int(indexterms[i][1]),int(page),int(ageinterval))  
ssex=simsex(indexterms[i][2],psex)  
ssbp=simsbp(indexterms[i][3],int(psbp),int(sbpinterval))  
sdbp=simdbp(indexterms[i][4],int(pdbp),int(dbpinterval))  
spregnancy=simpregnancy(indexterms[i][5],ppregnancy)  
sfainting=simfainting(indexterms[i][6],pfainting)  
scoillness=simcoillness(indexterms[i][7],pcoillness)
```

```
uppersimcase=wage*sage+wsex*ssex+wsbp*ssbp+wdbp*sdbp+wpregnancy*spr  
egnancy+wfainting*sfainting+wcoillness*scoillness+wcomedication*scomedicat  
ion+wsmoking*ssmoking+walcoholic*salcoholic+wheadache*sheadache+wdizzi  
ness*sdizziness+wpalpitation*spalpitation
```

```
lowersimcase=wage+wsex+wsbp+wdbp+wpregnancy+wfainting+wcoillness+w  
comedication+wsmoking+walcoholic+wheadache+wdizziness+wpalpitation  
simcase=uppersimcase/lowersimcase
```

Sample code for ranking relevant case in python

```
if len(listrcase)>0: #  
    for i in range(len(listrcase)):  
        for j in range(i+1,len(listrcase)):  
            if listrcase[i][15]<listrcase[j][15]:  
                temp=listrcase[i]  
                listrcase[i]=listrcase[j]  
                listrcase[j]=temp
```

Sample code for reuse

```
if listrcase[0][15]==1.0: # if there is exact matching  
    print "The recommendation for the problem is ", listrcase[0][14]  
  
else: # when there is uncertainty on the retrieved relevant cases  
  
#identifying and storing different recommendation from the most closest  
retrieved case  
  
lrecom=[]  
vrecom=[]  
for i in range(len(listrcase)):  
    if not lrecom.__contains__(listrcase[i][14]) and listrcase[i][15] == listrcase[0][15]:  
        lrecom.append(listrcase[i][14])  
  
  
# Computing for how many times each recommendation occurs in the retrieved  
relevant cases and store it in a nested list  
  
  
for i in range(len(lrecom)):  
    v=0  
    for j in range(len(listrcase)):  
        if lrecom[i]==listrcase[j][14]:
```

```
    v=v+1
vrecom.append([])
vrecom[i].append(lrecom[i])
vrecom[i].append(v)
```

*#sorting recommendation by their vote score (number of occurrence in the retrieved relevant cases)*

```
for i in range(len(vrecom)):
    for j in range(j+1,len(vrecom)):
        if vrecom[i][1]<vrecom[j][1]:
            temp=vrecom[j]
            vrecom[j]=vrecom[i]
            vrecom[i]=temp
```

# Declaration

I declare that the thesis is my original work and has not been presented for a degree in any other university.

---

Henok Bekele

July 2011

This thesis has been submitted for examination with my approval as university advisor.

---

Gashaw Kebede (PhD)