ADAMA SCIENCE AND TECHNOLOGY UNIVERSITY

SCHOOL OF GRADUATE STUDIES

DEPARTMENT OF COMPUTING

A MASTER'S THESIS

ON

APPLICATION OF CASE BASED REASONING FOR CHRONIC DISEASE DIAGNOSIS AND TREATMENT

BY

SULTAN ERMIAS

February, 2017

Adama, Ethiopia.

ADAMA SCIENCE AND TECHNOLOGY UNIVERSITY

SCHOOL OF GRADUATE STUDIES

DEPARTMENT OF COMPUTING

APPLICATION OF CASE BASED REASONING FOR CHRONIC DISEASE DIAGNOSIS AND TREATMENT

BY

SULTAN ERMIAS

The Thesis Submitted to School of Graduate Studies of Adama Science and Technology University for Partial Fulfillment of the Requirements for the Degree of Master of Science in Information System

February, 2017

Adama, Ethiopia.

ADAMA SCIENCE AND TECHNOLOGY UNIVERSITY

SCHOOL OF GRADUATE STUDIES

DEPARTMENT OF COMPUTING

APPLICATION OF CASE BASED REASONING FOR CHRONIC DISEASE DIAGNOSIS AND TREATMENT

BY:

SULTAN ERMIAS

SIGNATURE PAGE (Approval sheet)

_____	_____	_____
Advisor	**Signature**	**Date**
_____	_____	_____
Chairperson	**Signature**	**Date**
_____	_____	_____
External Examiner	**Signature**	**Date**
_____	_____	_____
Internal Examiner	**Signature**	**Date**
_____	_____	_____
Head of the Department	**Signature**	**Date**
_____	_____	_____
School Dean	**Signature**	**Date**

Dedication

I would like to dedicate my thesis to my best friend Hunlagn Tesfaye who give me their support to achieve my dream and also to my father, who passed away waiting for this achievement.

Acknowledgement

First and for most I would like to thank God "*Allah*" who blessed me with the ability to finish this thesis and granted me success in this long-time endeavor "*Alhamdullelah*". Next, my especial thanks go to my beloved family for their support and encouragement throughout my study.

I gratefully acknowledge my advisor, Dr. Mesfin Abebe for his commitment and patience for reading every section of the thesis, his valuable comments, encouragement and guidance from initial to the final level of the research that enabled me to finish the research work.

I am also greatly indebted to Mr. Tagel Abonah and Mr. Nuguse Leggese who helped me in giving valuable comments and editing the paper at the time of this work.

My deep appreciation and thanks extended to Adama Referral and Medical Hospital especially the Warmline staff, particularly Dr.Tolossa Eticha, Dr.Tamiret, Dr.Aman Lenjiso, Dr.Nuru Hassen, S/r Tsion, S/r Hiwot, Mr.Mohammed jellu (HO),Mr. Feleke (HO), Mr.Birbirsa Babiso (HO) for their cooperation during the knowledge acquisition. My deeply felt gratitude also goes to the Department of Information System and its entire staff for their support throughout my stay as a graduate student. I would also like to express my appreciation to the wonderful friends I have had at the Department of Information System most of whose my group members.

Last but not least, I would like to acknowledge my staff members of Hamda Diksis secondary school especially the directors of the school Ato. Shamillis Abiraye, Ato. Tashome Regasa, Ato. Bayu for making my life prettier and making all those hardships I had to go through easier by giving me free time during my educational study.

Table of Contents

List of Tables

List of Figures

List of Acronyms

ACBRCDD	Application of Case Based Reasoning for Chronic Diseases Diagnosis
AI	Artificial Intelligence
AIDS	Acquired Immunodeficiency Syndrome
CBR	Case Based Reasoning
COPD	Chronic Obstructive Pulmonary Diseases
DBMS	Data Base Management System
DBP	Diastolic Blood Pressure
FPS	Fasting Blood Sugar
GUI	Graphical User Interface
HIV	Human Immunodeficiency Virus
IR	Information Retrieval
JCOLIBRI	Java Class Ontology Libraries Integration for Building Reasoning Infrastructure
KBS	Knowledge Based System
SBP	Systolic Blood Pressure
TB	Tuberculosis
TBP	Target Blood Pressure
TST	Tuberculin Skin Test
WHO	World Health Organization

ABSTRACT

The threats to people's health from chronic diseases increase gradually. How to decrease these threats is an important issue in medical treatment. Thus, this study is conducted with the aim to explore the potential of case-based reasoning (CBR) approach for chronic diseases diagnosis. The main goal of this research is to develop a prototype case-based reasoning system that can give decision support for chronic diseases diagnosticians at a different level of expertise. To achieve the goal of this research literatures have been thoroughly reviewed from both case-based reasoning and chronic diseases diagnosis. For the implementation of the prototype, successfully solved cases are acquired from Adama Referral and Medical Hospital. In addition, the main parameters are identified in consultation with chronic diseases experts. Then, the prototype is implemented using jCOLIBRI case-based reasoning framework. Finally, domain experts tested the system and the average relevance of retrieved cases and the average fitness of the final solution to the new case is 88% and 92% respectively. The average adequacy and clarity of the system and average interactivity of the user interface is 84% and 80% respectively. The overall performance rating by domain experts is 85.56% and the accuracy of the retrieved process is 87.8% using randomly selected test cases. Given all these results, the performance of the prototype is promising. In general, the system has basically achieved the goal since the main objective of the system was to find the potential useful similar cases efficiently for the users.

CHAPTER ONE

INTRODUCTION

1.1. Background of the Study

The growth in population ageing in most industrialized nations and the related increasing percentage of chronic diseases diffusion is posing several problems at different levels of our society (political, social, and familiar/personal levels). These diseases result from prolonged exposure to causative agents, who are associated with personal behaviors, environmental factors and genetic influence. The global burden of chronic diseases continues to grow; tackling them constitutes one of the major challenges for development in the twenty – first century. Chronic diseases are becoming the heaviest burden to health care systems worldwide [1]. [2] reported that, from 57 million global deaths in 2008, 63 percent were due to chronic diseases mainly caused by pulmonary diseases such as: heart failure, cancer, diabetes and hypertension. In the next several decades major chronic diseases will continue to increase the morbidity and mortality rate in the entire world [3]. Africa's chronic diseases burden has been strongly attributed to changing behavioral practices (for example, sedentary lifestyles and diets high in saturated fat, salt and sugar), which are linked to structural factors such as industrialization and increasing food market globalization [1]; [4].It is compounded by weak health systems that are incapable to manage with the burden of infectious and chronic diseases.

Therefore, 80% of chronic diseases deaths occur in low and middle income countries. Moreover, chronic diseases cause substantial financial burden, and can push individuals and house-holds into poverty. In low and middle income countries, middle-aged adults are especially vulnerable to chronic diseases. People in these countries tend to develop disease at younger ages, suffer longer often with preventable complications and die sooner than those in higher income countries. Therefore, Ethiopia as one of the African countries also shares the problem of chronic diseases. It is the second most populous country in Sub-Saharan Africa. A very large proportion of the population (84%) lives in the rural areas and the country experiences a heavy burden of disease mainly attributed to communicable infectious diseases [5]. Chronic tuberculosis is amongst the world's leading cause of death from a single infectious disease. The incidence of TB has more than

doubled in Africa during the last two decades. Ethiopia is ranked 8th among the 22 countries with the highest TB burden in the world [6]. According to [7], there were an estimated 220,000 (261 per 100,000) incident cases of TB in Ethiopia in 2010. According to the same report the prevalence of TB was estimated to be 330,000 (394 per 100,000). There were an estimated 29,000 deaths (35 per 100,000) due to TB, excluding HIV related deaths, in Ethiopia during the same period [8]. According to [9], tuberculosis is the second cause of death in Ethiopia. During the year 2010/11 (2003 EC), a total of 159,017 TB cases were notified in Ethiopia. Among these 151,866 (95.5%) were new cases of TB, all forms. According to [9], about 79% of HIV positive clients were screened for TB, of these 11% was found to have active TB. This unprecedented increase in TB is attributable to a number of factors, one of the most important being the large number of infectious TB patients who remain undetected and untreated. In addition to communicable TB the burden of chronic (Non- communicable) disease in sub-Saharan Africa is very high, the current density and distribution of the health workforce suggested that sub-Saharan Africa can't respond to the growing demand of chronic diseases care. We chose chronic diseases that are the most commonly occurring and with the highest morbidity and mortality, and, among these, diseases that do not typically have a specialist as the primary physician (e.g., cancer, pediatric diabetes, hypertension, COPD) have available measures of national prevalence (in the general population), and have published clinical guidelines [10]; [11]. Among the different chronic health problems; the most widespread in Ethiopia are heart disease, hypertension, stroke, cancer, chronic respiratory, diabetes, COPD, and also the infectious disease TB, where they affect nearly half of the adult population and their prevalence has increased in recent years; [12]; [13]. In Ethiopia the number of physician and nurse for 100,000 people are 3 and 6 respectively which clearly shows the shortage of qualified human power in the domain area [14].

The Ethiopian Government has designed a series of Health sector Development programs in line with the expansion and improvement of health services. The focus under the Growth and Transformation Plan (GTP) will continue towards primary health care and preventive services. Furthermore, the Government has been strengthening its measures to improve the number, skills, distribution and management of health professional through accelerated training of physicians (specialists and general practitioners). A new system for the health care service includes financial and drug management implemented to overcome the bottlenecks in the sector. Incentive package is further strengthening in order to domestically produce pharmaceutical products. These progressive strategic plans would adequately strength the potential of hospitals to provide an effective and

efficient health care service [15].

The increasing expectations of the highest quality health care and the rapid growth of evermore detailed medical knowledge leave the doctor without adequate time to devote to each case and struggling to keep up with the newest developments in his field [16]. For lack of time, most medical decisions must be based on rapid judgments of the case relying on the doctor's unaided memory. Only in rare situations can a literature search or other extended investigation be under taken to assure the doctor (and the patient) that the latest knowledge is brought to bear on any particular case. Continued training and recertification procedures encourage the doctor to keep more of the relevant information constantly in mind, but fundamental limitations of human memory and recall coupled with the growth of knowledge assure that most of what is known cannot be known by most individuals. It is the opportunity for new computer tools: to help organize, store, and retrieve appropriate medical knowledge needed by the practitioner in dealing with each difficult case, and to suggest appropriate diagnostic, prognostic and therapeutic decisions and decision making techniques. Correct diagnosis of chronic diseases at an early stage is a demanding task due to the complex interdependence on various factors. Another major challenge faced by hospital is the provision of quality services at affordable cost.

Several real world applications require capturing the evolution of the observed phenomenon overtime, in order to describe its behavior, and to exploit this information for future problem solving [17]. Man strives to augment his abilities by building tools. These tools have extended his ability to sense and to manipulate the world about him. Today we stand on the threshold of new technical developments which will augment man's reasoning, the computer and the programming methods being devised for it are the new tools to effect this change. This means that the complex interdependence factors in medical field especially in hospitals should be solved through technological advancement in which man innovates from time to time in the world. Medicine is a field in which such help is critically needed [16].

These are the motivations to develop medical diagnosis prediction system which can predict chronic diseases by processing the previous known cases. Working on chronic diseases patient’s databases is one kind of a real-life application. The databases include several factors and attributes. Therefore, CBR (Case Based Reasoning) is proposed for supporting diagnosis of chronic diseases in this study CBR has been used in various problem-solving areas such as financial forecasting, credit analysis and medical diagnosis.

1.2. Statement of Problem and its Justification

In most developing countries provision of healthcare services is a major challenge. Some of these challenges are lack of highly qualified medical human resources, financial as well as the ability to manage and transform scarce resources to meet healthcare needs [18].

Some of the studies indicate that chronic diseases are serious health problems in developing countries like Ethiopia. It is also a major cause of death and impaired quality of life [19]. The use or misuse of addictive substances, such as cigarettes, alcohol, and khat (Catha edulis Forsk) which are the primary factors for chronic diseases are increasing in the country. Due to this and other factors chronic diseases are widely spread in the country and may represent a silent epidemic in the population [19]. These conditions result in numerous adverse health outcomes, increased health care needs, and subsequently higher medical costs [20]. A financial resource for chronic healthcare in Ethiopia is also extremely inadequate [21]. In addition to this, there is no qualified experts are available in the zonal woreda due to knowledge gap chronic diseases patient's mortality rate annually increasing. As a result, chronic health services in Ethiopia are not sufficient [22].In addition, shortage of physicians or doctors and clinicians, the distribution and lack of knowledge among primary healthcare workers are the main challenges that become an obstacle to address chronic health services satisfactorily throughout the country [21]. In the majority of primary healthcare institutions, there is a shortage of qualified professionals. Some of primary healthcare workers do not have good and enough knowledge for diagnosing chronic diseases. Due to this, in most of the primary healthcare sectors, chronic diseases patients may not get the appropriate treatment. Not all clinicians are competent diagnosticians of chronic diseases [21]. The management of chronic diseases is complex and requires a well-trained health force. Furthermore, to the above mentioned problems the general practitioners have been done the following problems in their domain area that affects their decision.

- Because of the overcrowded number of patients, the patient handling capacity of the hospital and related reason enforce the general practitioner to miss-diagnosis the patient. Because of he/she uses one or two symptom for diagnosing the patients' but patients' health problem may need further investigation to identify problem.
- Because of the general practitioner or the nurse may be temporary biased or frustrated to do in a normal condition.

Consequently, due to the above mentioned reasons most of the time the general practitioner is not in

a position to make good decisions. As a result, these factors expose the patient for long waiting and additional cost in the hospital. The problem also enforces the physician to perform the same patient case by different expert and create burdens on his/her daily activity.

Although, doctors and the qualified physicians apply their skills and good experience to distinguish the diagnosis of chronic diseases for patients, the medical data collected from patients should be saved in files and kept in folders. Otherwise those huge amounts of messy medical records do not have any meaning for user [16].

Therefore, the researcher motivates to develop CBR system for chronic diseases diagnosis and treatment. Using CBR, a technique which solves a new problem by remembering a previous case and by reusing information and knowledge of that case, CBR turn those data into useful information that can help for the diagnosis of chronic diseases.

The developed automated CBR system will assist medical physician in quickly generating recommended treatment and can also serve as a teaching and training aid for junior, inexperienced medical physicists. In addition, the developed methods are generic in nature and can be adapted to be used in other CBR or intelligent decision support systems for other complex, real world, problem domains that highly depend on subjective and intuitive knowledge. The system could be used in order to assist general practitioners and primary care physicians in improving the diagnosis of chronic diseases. Presently, doctors have difficulties in determining chronic diseases in a new patient who does not have existing medical record. Therefore, those data can be used to diagnose chronic diseases for new patients who do not have existing medical records. This system is designed to assist doctor and health professionals in determining the diagnosis of patient data. Therefore, this system could help doctors and health professionals to determine the diagnosis and analysis of the patient health status. This would also reduce mortality rate of patients with chronic diseases and improve quality healthcare services.

The purpose of this study is to create a decision support system that incorporates the wealth of experience possessed by experts, by applying case-based reasoning to generate a good treatment for chronic diseases patients.

To this end, the study attempts to explore and find solutions for the following research questions:

1. What types of knowledge the human expert use in the diagnosing process of chronic diseases patients' health problem?
2. How to model, represent and implement the acquired knowledge for developing the CBR system?

3. How much is the prototype accepted by the domain experts during evaluation of the system?

1.3. Objective of the Study

1.3.1. General Objective

The main objective of this study is to design and develop a CBR system for diagnosing patients with chronic diseases in order to provide advice to domain experts and enhance services given to patients.

1.3.2. Specific Objectives

The following are the specific objectives that have been achieved in the course of the study.

- To review different global and local researches that can help for the successful completion of the study.
- To acquire chronic diseases knowledge from experts and documents and model the acquired knowledge using appropriate modeling technique
- To build a prototype CBR system for chronic diseases diagnosis
- To test the performance and user acceptance of the prototype.
- To recommend further research based on the conclusion drawn at the end of the study.

1.4. Significance of the Study

From this study, primary healthcare workers and health professionals, specifically the inexperienced chronic diseases health professionals, are the immediate beneficiaries to enhance their day to day activities. The prototype has great significance to teach primary health care workers, general physician and nurses in order to have well understanding about chronic diseases. As a result, the health workers can use the system to diagnose chronic diseases based on the symptoms given by the patient and the case retrieved from the case base.

1.4.1. Beneficiary of the Research Study

In general, the benefits and beneficiaries of the research are the following: Benefits of the research:

- The system can help physicians by providing useful information about the disease identification, diagnosis and advisement.

The beneficiaries of the proposed system are the following:

- Patients, individuals, the public in general or community

- Hospital professionals /experts
- Governmental and non-governmental institutions, health institutions, etc.

1.5. Scope and Limitation of the Study

The scope of this study focused on the application of CBR system for chronic diseases diagnosis and treatment. The focus area of this study is only Adama Referral and medical Hospital. In particular, the work concentrates on the retrieval stage of the CBR system. Hence, for this study the researcher chose chronic diseases that are the most commonly occurring and with the highest morbidity and mortality, and, among these, diseases that do not typically have a specialist as the primary physician on the basis of those identified and prioritized by domain experts at Adama Referral and medical Hospital. But it doesn‘t deal with the management of drug response to the chronic diseases patient was not included. During testing and evaluation of the prototype, only Adama Referral and Medical Hospital chronic disorder case team members are involved. The extent to which the prototype is used by the chronic disorder experts is that the experts are not recommended to use the retrieved solutions automatically; instead they have to adapt the retrieved solution in a way that fit to the new problem at hand.

1.5.1. Challenges of the Study

The proposed research encountered different limitation throughout the research process, Some of them are; constraints during knowledge acquisition process i.e. domain experts might not give full information about chronic diseases because of experts work over load and they forget the full procedure to diagnose the disease and the other constraint is lack of time to cover all chronic diseases in the research study.

1.6. Research Methodology

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

1.6.1. Literature Review

Materials such as journals, related works and manuals, which are helpful for the enhancement of the research, were reviewed. The necessary documents and tools for the development of the prototype were also reviewed.

1.6.2. Sampling Techniques and Sample Size

Purposive sampling techniques are used to collect 164 previous chronic diseases patient cases archived in Adama Referral and Medical Hospital. The selection criteria of domain experts for knowledge acquisition of the study are based on the profession or expertise, educational qualification level, year of experience and their immediate position given in the department.

1.6.3. Data Collection

To acquire the needed knowledge, the researcher has used primarily documented chronic diseases patient's history from Adama Referral and Medical Hospital. Health professionals from Adama Referral and Medical Hospital have involved eliciting the tacit knowledge. In addition, relevant literature from all possible sources, including journal articles, guideline for chronic diseases diagnosis, chronic diseases related books, thesis, the internet and related websites are reviewed as secondary sources. The acquired domain knowledge was modeled using hierarchical tree structure and implemented using case-based reasoning system. The reason for choosing hierarchical tree structure is its simplicity and ease of use to model the knowledge.

1.6.4. Data Preparation

The collected chronic diseases cases are arranged into a form that is suitable for the selected retrieval algorithm and case based reasoning tools. Chronic diseases cases that have noisy data are removed from the collected chronic diseases 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 chronic diseases 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 [23].

1.6.5. Case Features Selection and Weights

The cases represented in this study were categorized into two namely: description and solution. The description attributes are used for a user to enter the query and the solution attributes are the results displayed for a user after the similarity calculation between the query and existing case. 15 attributes has taken as a relevant description attributes based on the interview conducted with the domain experts. For solution, two attributes were selected. The description attributes are given weights according to their level of importance. The attribute which has high weight will have high similarity measure.

1.6.6. Design and Implementation Tools

In designing the case based reasoning system prototype for chronic diseases diagnosis, 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 [24]. According to [25], Nearest Neighbor algorithm has the advantage of simplicity in retrieving relevant cases. It is also suitable when there are attributes that have numeric value [26]. A voting method is used for the reuse process. This method is tested by [27] 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.

There are many programming tools used to develop different kinds of CBR systems. To develop case based reasoning system, this research used JCOLBRI framework. It is a general framework that supports many features like graphical interfaces, description logics and ontologies, textual CBR, evaluation, and so on.

1.6.7. Testing and Evaluation Techniques

The proposed CBR system has test method and evaluation method to ensure the performance of the system through test case and interface interaction with the end users. The evaluation processes tells how proposed system prototype accurate and significant to the end users. User acceptance measurements are concerned with issues how well the system addresses the needs of the user, whereas, validation measurement determines if the system performs the required task successfully. The system evaluators use visual interaction methods along with the questionnaires. Based on that evaluation criteria the user evaluate, the performance of the proposed system prototype user acceptance by using both close ended and open ended questions and also the prototype tested using test case to know about the prototype system accuracy by taking the following parameters into consideration: The users‟ feedback, regarding:

- The relevance of the information retrieved (how similar the cases stored in the case base to the case at hand).
- The relevance of the fields (attributes) in the case structure.
- The ease of use of the system, in terms of the entry and output forms layout and level of detail.

1.7. Thesis Organization

This section of the research summarizes the overall contents of the thesis.

Chapter 1: Introduction - In this chapter the researcher tries to discuss the background of the study, a statement of the problem with its justification, objective of the study, the scope of the study, the significance of the study and the methodologies used to conduct the research.

Chapter 2: Literature Review - This chapter mainly focuses on the review of literature. In this section, the conceptual discussion of CBR mainly focuses on the definition of CBR, CBR life cycles, CBR techniques, CBR tools and Application of the CBR system in the medical domain specifically in chronic disease diagnosis.

Chapter 3: Knowledge Acquisition and Modeling - This chapter presents the knowledge acquisition process, Pre-acquisition phrase, the process knowledge acquisition, general structure of chronic diseases and diagnosis of chronic diseases case structure.

Chapter 4: Design and Implementation - In this chapter the design and implementation of the prototype are realized. The case based reasoning for chronic diseases diagnosis is developed. The implementation of the system using jcolibri2.

Chapter 5: Performance Evaluation of the Prototype – In this chapter the performance of the prototype is evaluated both the performance of the system and the acceptance of the system by the users.

Chapter 6: Conclusion and Future Work - The last chapter of this thesis. In this chapter the main findings are summarized and based on the findings, recommendations are proposed for future research.

CHAPTER TWO

LITERATURE REVIEW

2.1. Introduction

The ambition for computer systems being able to support human experts during complex problem-solving task is a usual topic of AI research. In order to enable a computer system to give rational support when solving problems in a complex application domain, it is essential to provide it with specific knowledge within that domain. A number of methodologies to realize such knowledge-based systems have been developed, such as, rule-based approach. In recent years, CBR has become a very popular technique for developing knowledge-based systems that can give rational support using specific knowledge. In some real-world application domains, it has even emerged to one of the commercially most successful approaches compared to other techniques developed in AI research such as rule-based systems. Using CBR, different applications have been developed yet, to solve problems in different domains. CBR is useful for a wide variety of problem-solving tasks, including classification, planning, diagnosis, design and decision support [28]. The previous chapter to this study discussed the problem that has to be investigated by this study, and the methodology used for the successful completion of the study. This chapter provides a conceptual discussion of CBR, application of CBR in the medical domain more specifically in chronic diseases diagnosis and gives a highlight about related researches to the study.

2.2. Knowledge Based System

A knowledge-based system (KBS) is a computer program that reasons and uses a knowledge base to solve complex problems. The term is broad and is used to refer to many different kinds of systems. Every knowledge based system has two building blocks which are known as knowledge base and inference engine (IE). The IE is a software program, which infers the knowledge available in the knowledge base. The knowledge base can be used as a repository of knowledge in various forms [29].

2.2.1. Architecture of Knowledge Based Systems

According to [29], every KBS have at least five main components. These components are depicted schematically in Figure 2.1, and are explained below.

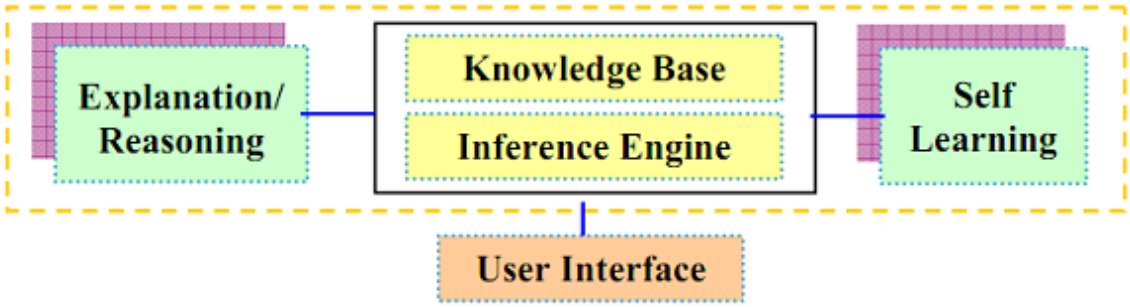

Figure 2-1: Architecture of Knowledge Base System [29]

The Knowledge Base contains all relevant knowledge acquired from domain experts. Knowledge also acquired from the user during their interaction with the system. The inference engines formulate questions and assert the answers provided by the user in a natural language form. It provides a mechanism for conveying recommendations to the end user. The explanation module provides a brief description to the user why the system arrived at a certain conclusion [30]. Self-learning is one of the elements of KBS which tries to imitate the learning capability of human beings. It is possible to update the knowledge base of the KBS either manually or automatically using machine learning algorithms [29].

User interface is a channel for communication between the KBS and the end-user. Therefore, in order for the KBS to be an interactive tool, it should include a means to show and retrieve information in a simple manner [31]. According to [29] the following are the advantages and drawbacks of knowledge based systems over traditional computer systems.

Advantages

- Overcome shortage of human experts
- Expertise can be multiplied and stored for future use.
- There are more than one experts' knowledge have grouped at one platform.
- Intelligent assistance is important for decision making.

Disadvantages

- Large volume of knowledge acquisition, representation and manipulation.

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

2.3. Types of Knowledge Based Systems

Case based reasoning and rule based reasoning are two types of reasoning techniques which are used in case based systems and rule based systems; and they use cases and rules for knowledge representation respectively. The two are discussed in the following sections.

2.3.1. Rule-Based Reasoning

A rule-based system may be viewed as consisting of three basic components: a set of rules [rule base], a data base [fact base], and an interpreter for the rules. In the simplest design a rule is an ordered pair of symbol strings, with a LHS and RHS. The rule set has a predetermined, total ordering, and the data base is simply a collection of symbols. The interpreter in this simple design operates by scanning the LHS of each rule until one is found that can be successfully matched against the data base. At that point the symbols matched in the data base are replaced with those found in the RHS of the rule and scanning either continues with the next rule or begins again with the first. A rule can also be viewed as a simple conditional statement, and the invocation of rules as a sequence of actions chained by modus ponens [32].

Reasoning Techniques

According to [33], two alternative strategies are available: Forward chaining (data driven) and backward chaining (hypothesis driven) techniques represent the fundamental reasoning approaches implemented in rule-based expert systems.

In **forward chaining,** the expert system is given data and chains forward to reach a conclusion. Forward chaining systems are data-driven. In backward chaining, the expert system is given a hypothesis and backtracks to check if it is valid. Backward chaining systems are generally goal driven or 'goal-orientated' in the sense that it tries to prove a goal or rule conclusion by confirming the truth of all of its premises [33].

Backward chaining, starts with a list of goals (or a hypothesis) and works backwards to see if there are data available that will support any of these goals. An inference engine using backward chaining would search the inference rules until it finds one which has a Then clause that matches a desired goal. If the If clause of that inference rule is not known to be true, then it is added to the

list of goals (in order for your goal to be confirmed you must also provide data that confirms this new rule) [33].

Advantages Rule Based Reasoning

Rule based reasoning approach have a numbers of good features. According to Jim [34] the major advantages of rule based reasoning in the development of knowledge based system are:

- Compact representation of general knowledge
- Homogeneity
- Independent
- Naturalness of representation
- Modularity
- Provision of explanations

Disadvantages Rule Based Reasoning

As rule based reasoning of prototype knowledge based system has many advantages. But, it has the following limitations [34].

- Knowledge acquisition bottleneck
- Brittleness/fragility of rules
- Inference efficiency problems
- Difficulty in maintenance of large rules
- Interpretation problems

2.3.2. Case-Based Reasoning

Case-based reasoning (CBR) means "adapting old solutions to meet new demands, using old cases to account for new situations, using old cases to evaluate new solutions, or reasoning from precedents to interpret a new situation" [35].CBR is more comfortable to make better decision in dynamically changing environment. People learn from their success and wrong activities to handle similar situations in the right manner and not to repeat their mistake of the past. CBR approach is more compatible to reuse previously solved problems and learning from experiences for future decision [36]. Similarly, CBR is an approach to incremental learning. Once a problem has been solved, CBR approaches use the solution to solve for future problems [37].

Cases

In CBR terminology cases are usually denotes a problem situation. A case can be defined as previously experienced situation which has been captured and learnt, is referred to as a past case, previous case, stored case or retained case. Note that the term problem solving is used with common practice in the area of knowledge-based systems. This means that problem solving is not necessarily the finding of a concrete solution to an application problem, it may be any problem put forth by the user [37].Therefore, and cases have three different aspects this can be described as follows.

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

Case Representation

The simplest way to represent a case is by using feature-value pairs. A feature value pair is used to represent a state of an entity; Features need to be identified for both problem and solution. The need for more complex representations originates from the fact that such a representation cannot entail everything we can see and that is of interest.

The Case Base

A case base is a memory; it contains a collection of cases that is used in the context of the CBR methodology for the purpose of performing a reasoning task. A case base is a data source and usually it is finite.

Case Organization

We have three main types of case organization: flat, structured, and unstructured. The flat organization is the simplest to design and implement, and most suitable for a small number of cases. There are no relationships between the cases. No one case has any relationship to another that needs to be represented, which means that the representation is complete. Cases can be organized into structures such as hierarchies and networks. The structured organizations

can be beneficial when the number of cases is very large. Cases are very commonly hidden within texts or images which are unstructured organizations.

Advantage of Case Based Reasoning

A case based reasoning approach has tremendous advantages in the development of knowledge based system. The following are main the advantages of case-based reasoning [36].

- Ability to express specialized knowledge.
- Naturalness of representation
- Modularity
- Easy to knowledge acquisition
- Self-updatability.
- Handling unexpected or missing values.
- Inference efficiency.

Disadvantage of Case Based Reasoning

Even though case based reasoning approaches have a numbers advantage. But, due to lack of sufficient cases, the construction and inference mechanism of a case-based system loss the required objective. Some of the limitation issues in case-based reasoning are [34]:

- Inability to express general knowledge
- Knowledge acquisition problems
- Inference efficiency problems and Provision of explanations

2.3.3. The Comparison of Rule Based and Case Based Reasoning

Rule based reasoning: Symbolic rules are one of the most popular knowledge representation and reasoning methods. Their popularity stems mainly from their naturalness, which facilitates comprehension of the represented knowledge.

The basic form of a rule is the following:

If <conditions>

Then <conclusion>

Where <conditions> represents the conditions of a rule, whereas <conclusion> represents its conclusion. The conditions of a rule are connected between each other with logical connectives such as AND, OR, NOT etc., thus forming a logical function. When sufficient conditions of a rule are satisfied, the conclusion is derived and the rule is said to fire (or trigger). Rules represent general knowledge regarding a domain [34]. Rules are suitable to represent general knowledge, whereas cases are suitable for representing specific situations. Rules in a rule based system have the abilities to represent experiential knowledge acquired from experts in a direct fashion. Cases are capable of representing specific historical knowledge. The problem here is that it is difficult to acquire complete and perfect knowledge in a complex domain. Cases are natural and easy to obtain. They can be collected from the historical record, repair logs or other sources [34]. CBR uses partial matching to draw a conclusion. If some of the given problem descriptions match with a given case, then the case is applicable to the proposed solution. It also tries to handle novel problems by referring previously solved cases. Rule based reasoning uses perfect matching to apply a rule for a given problem. It doesn't handle missing information and unexpected data values [38]. A case-based system can handle unexpected cases not recorded in the system or missing input values by assessing their similarity to stored cases and reusing relevant cases. It is not possible to draw conclusions from rules when there are missing values in the input data.

Thus, using the combination of both approaches, makes use of both existing knowledge and the past experiences. This integrated approach eliminates the drawbacks of each method and provides a better way to handle problems, which combine both inductive and deductive approaches [34].

2.4. Knowledge Engineering Process

Knowledge Engineering is a term that used to describe the Knowledge Based System jobs that includes the analysis and representation, and computer system design and implementation [39]. According to [40] Knowledge Engineering also defined as a process of acquiring knowledge from experts and building a knowledge base. Generally, the knowledge engineering development phases includes the knowledge acquisition, knowledge modeling, and knowledge representation, coding or knowledge implementation and finally, system testing and evaluation. All these detailed description discussed in the following section.

2.4.1 Knowledge Acquisition

Knowledge acquisition involves the extracting of knowledge from human experts, books, documents, sensors, or computer files. The knowledge may be specific to the problem domain or to the problem-solving procedures, it may be general knowledge (e.g., knowledge about business), or it may be Meta knowledge (knowledge about knowledge). (By Meta knowledge, we mean information about how experts use their knowledge to solve problems and about problem-solving procedures in general.) [41]. There exist several methods to extract human knowledge; some commonly used approaches of knowledge acquisition include interviews, questionnaires, record reviews and observation to acquire factual and explicit knowledge [29].

Knowledge Elicitation*:-* Knowledge Elicitation (KE) is the process of obtaining knowledge from a domain expert that describes how they perform a specific task and/or describes what general knowledge they have about the domain. One use of KE is obtaining knowledge from a person in order to transfer it to a computer program [42].

Interviewing:- it consists of asking the domain experts questions about the domain of interest and how they perform their tasks. Interviews can be unstructured, semi structured, or structured. The success of an interview session is dependent on the questions asked it is difficult to know which questions should be asked, particularly if the interviewer is not familiar with the domain. Its success also depends on the ability of the expert to articulate their knowledge. The expert may not remember exactly how they perform a task, especially if it is one that they perform "automatically" [42]. Interview can be classified into structure, semi structure and unstructured interview.

Structured Interviews:- A structured interview method is questioning the domain expert directly. It is goal-oriented process. It forces organized communication between the knowledge engineer and the domain expert. The structure reduces the interpretation problems inherent in unstructured interviews and allows the knowledge engineer to prevent the bias caused by the subjectivity of the domain expert [43].

Semi Structured Interview:- is an interview which has a guide that usually includes both closed ended and open-ended questions. It is more flexible than structured one. In these kinds of interview the interviewer has a chance to change the order of questions and expand the dimension of questions based on the participants' responses [43].

Unstructured Interviews:- sessions are conducted informally ways of making interview. Unstructured interview techniques provide complete or well-organized descriptions of cognitive processes. There are many reasons that enforced to applying unstructured interview. Domain the experts usually find it very difficult to express some of the most important elements of their knowledge. Through structured interview it is difficult to acquire the required knowledge. With good training and personal experience knowledge engineers can use unstructured interview to acquire relevant knowledge from domain expert [43].

Observation:-In Observation methods, the knowledge engineer observes the expert performing a task. This prevents the knowledge engineer from inadvertently interfering in the process, but does not provide any insight into why the decisions were made [44].

Document Analysis:-This technique is used to collect relevant knowledge from the existed documents of different format. These documents include professional literature, brochures, manuals, guidelines, employee handbooks, reports, glossaries, course texts, and other relevant materials [45].

Protocols:-The KE technique called Protocol Analysis involves asking the expert to perform a task while "thinking aloud." The intent is to capture both the actions performed and the mental process used to determine these actions. As with all the direct methods, the success of the protocol analysis depends on the ability of the expert to describe why they are making their decision. In some cases, the expert may not remember why they do things a certain way. In many cases, the verbalized thoughts will only be a subset of the actual knowledge used to perform the task [46].

2.4.2 Knowledge Analysis and Knowledge Modeling

Knowledge analysis is the process of making sense of knowledge acquired from domain experts and secondary sources [47]. The primary purpose of knowledge analysis is to analyze and organize the knowledge gained during the knowledge acquisition phase [48]. Similarly, knowledge modeling, also known as conceptual modeling is one of knowledge acquisition activities that help to shed light on the structure of knowledge based systems. It is the process of expressing the analyzed knowledge in an understandable and usable form such that preparing the knowledge for knowledge representation phase [47].

2.4.3 Knowledge Representation

Knowledge representation is the process of encoding human knowledge in computer understandable form. It is concerned with designing and using systems for storing knowledge [49]. The purpose of knowledge representation is to organize the acquired and modeled knowledge into a form that a knowledge based system can readily access for decision making [29]. There different knowledge representation methods are exist: heuristic rules, semantic networks, frames, objects, decision tables, decision trees, and predicate logic.

2.4.3.1 Case Representation

Case representation in CBR makes use of familiar knowledge representation formalisms from AI to represent the experience contained in the cases for reasoning purposes. A large variety of representation formalisms have been proposed. However, three major types of case representation have arisen: feature vector (or propositional) cases, structured (or relational) cases, and textual (or semi-structured) cases. In addition, specialized representations of cases for specific tasks have been used, such as plans used as cases in planning [50]. Cases can be represented in four ways discussed below.

Feature-Vector Approaches

Represent a case as a vector of attribute-value pairs, similar to the propositional representations used in Machine Learning, that support k-nearest neighbor matching and instance-based learning [51].

Structure Approaches

Originate from the episodic memory notion of cases. The representational structure itself is usually developed around a frame-based formalism. Since frames can be seen as a subset of first-order logic, cases represented as frames or some frame-like structure are examples of what ML calls relational representations. Such representations arise from gathering together clusters of relations that occur together in elementary objects. Cases, from this perspective, are clusters of relations between the kinds of elementary objects that comprise them [50].

Textual Case Representations

Go in the opposite direction by imposing only a weak structure on the cases. This allows easy exploitation of the experience captured in documents such as bug reports or FAQ's. Text fields

are usually represented as sets of linguistic items (e.g. words reduced by some stemming algorithm) similar to vector representations in information retrieval. However, the available (weak) structure of the text allows the introduction of more semantics for a better interpretation of the reusability of a case [50].

Object-Oriented Representations

Have expressiveness similar to frame representations, but have a different origin. They make use of the data modeling approach of the object-oriented paradigm, including is-a and part-of relations as well as the inheritance principle. Cases are represented as collections of objects, each of which is described by a set of attribute-value pairs [52].

Knowledge Validation

As [53] mentioned, there are four types of evaluations to be conducted on KBS. These are verification, validation, usability and usefulness.

Verification is the rightness of the developed KBS to be evaluated. It can be conducted entirely on the formal model or on the computable model whose syntax is clearly stated for their rightness to be evaluated. It assures whether the knowledge on the formal model or on the computable model does not comprise syntactical faults. A verified KBS denotes the acquired knowledge from domain experts and secondary sources rightly.

Validation is checking the knowledge base of the KBS for semantic faults that may occur during the KBS development. A validated KBS comprises the correct knowledge to perform like the domain expert in the domain area. Thus, validation searches for faults in the KBS behavior when it attempts to find a solution for a certain domain problem.

Usability is an association between the KBS and the end-user. This means whether the end-user is satisfied when he/she interacts with the KBS. Therefore, it must be evaluated before installing the KBS to the end-user.

Usefulness refers the association among the new KBS, the end-users, and the company that owns the product. The usefulness view can be noticed when the new KBS accomplishes its job. It is not possible to evaluate the new KBS if it is not functional.

2.5. CBR Cycle

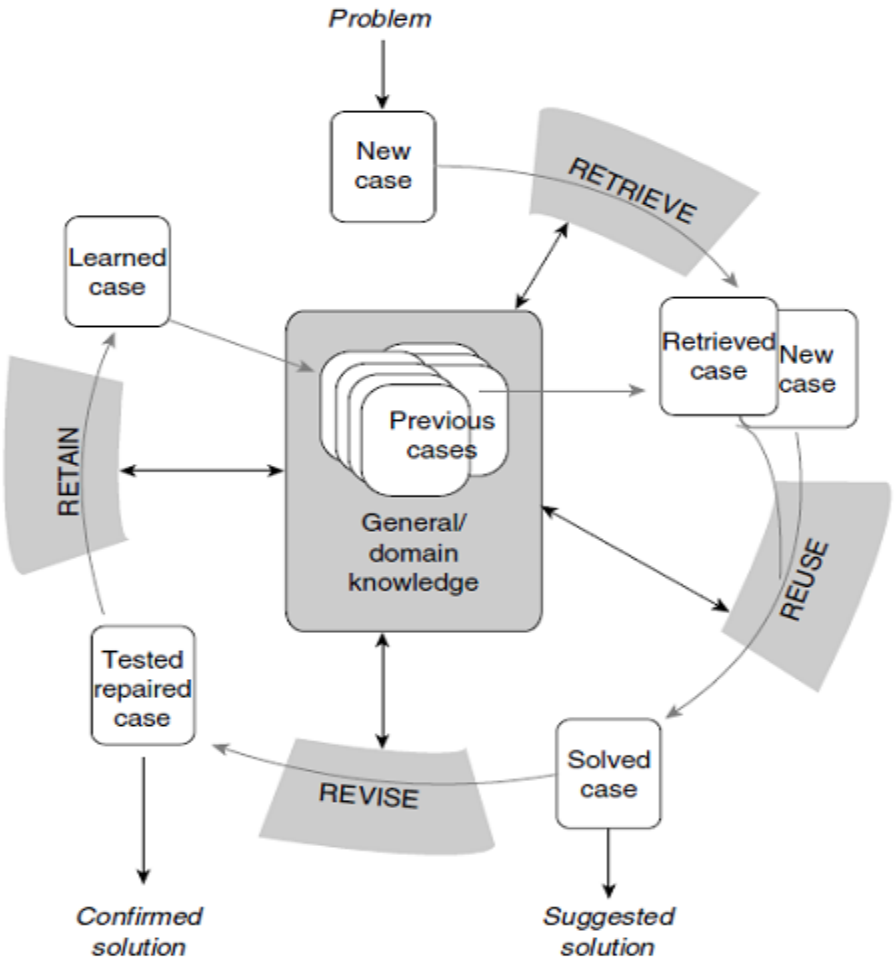

Figure 2-2: CBR Life Cycle [37]

At the highest level of generality, a general CBR cycle may be described by the following four processes:

1. RETRIEVE the most similar case or cases
2. REUSE the information and knowledge in that case to solve the problem
3. REVISE the proposed solution
4. RETAIN the parts of this experience likely to be useful for future problem solving A new problem is solved by retrieving one or more previously experienced cases, reusing the case in one way or another, revising the solution based on reusing a previous case,

and retaining the new experience by incorporating it into the existing knowledge-base (case-base).

Case Retrieval and Similarity Assessment

Retrieving in this context is remembering one or more similar cases. It is the main and the first step of the case based reasoning method. It takes the description of a problem as its input and provides the best matched case or set of cases as output. The quality of a case based reasoning system as a whole is highly affected on the quality of its retrieval process due to its being the base for the rest of processes.

Various researchers describe it in different ways. [37], 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, [38] 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: [54]; [55]

- 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. [56]; [55]; [54]. 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 [26]. 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: [27]

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 [55].

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

Figure 2.3 Nearest Neighbor Algorithm

Where *w* is the importance weighting of an 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 [27]; [54].

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 [55].

Case Adaptation (Reuse)

The selected case in the retrieval process can be used to understand the new situation when it is not clear by itself, to propose a solution based on the solution taken on the selected case, or to prevent from following a wrong way of solving the problem based on the outcome of the selected 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 [57].

Case Revision

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 [57].

Case Retain (Learning)

In the classic review paper by [37], retention is presented as the final step in the CBR cycle, in which the product of the most recent problem-solving episode is incorporated into the system's knowledge. To a great extent this has traditionally translated into a variety of approaches for recording the product of problem solving as a new case that can be added to the case base. Of course, there are various issues concerning how best to learn a new case and different systems record different types of information in their cases. Most, for example, simply record the target problem specification and the final solution, with the implicit assumption that the outcome was successful.

Integrating Case Based Reasoning with other Types of Knowledge

Case based reasoning can be integrated with additional types of knowledge and reasoning. The common types are Rule based reasoning and ontologies. For the purpose of this research only integration with ontologies is discussed below. According to [58] the formalization of ontologies is useful for the CBR community regarding different purposes, namely:

- Persistence of cases and/or indexes using individuals or concepts that are embedded in the ontology itself.
- As the vocabulary to define the case structure, either if the cases are embedded as individuals in the ontology itself, or if the cases are stored in a different persistence media as a data base.
- As the terminology to define the query vocabulary. The user can express better his requirements if he can use a richer vocabulary to define the query. During the similarity computation the ontology allows to bridge the gap between the query terminology and the case base terminology.
- Retrieval and similarity, adaptation and learning.
- Knowledge reuse between different CBR systems.

Case Based Reasoning and Retrieval Systems

In this section the researcher considered methods where certain solutions are retrieved from some kind of data source or base. These are database systems, information retrieval systems, and pattern recognition systems.

Database Systems What is a database? In essence a database is nothing more than a collection of information that exists over a long period of time, often many years. In common parlance, the term database refers to a collection of data that is managed by a *database management system* DBMS. The power of databases comes from a body of knowledge and technology that has developed over several decades and is embodied in specialized software called a *DBMS.* DBMS is a powerful tool for creating and managing large amounts of data efficiently and allowing it to persist over long periods of time, safely [59]. There is an obvious connection between DBMS and CBR systems. Both contain data elements of interest that one wants to retrieve if needed. Often, databases and CBR systems are combined. A major reason is that in application scenarios cases were stored in a database before one was thinking of using CBR [60].

Table 2-1: Databases versus CBR Systems [60]

Database systems	***CBR systems***
Simple search: All or nothing	Search for most similar objects
Under specification, too many answers over specification, no answers possible	Number of retrievals can be specified
No specific domain knowledge needed	Domain knowledge in the containers
Efficient and correct answers	Search can be slow, correctness not ensured
Closed world	Open world
Static when in use	Can improve incrementally
Long transactions allowed	In principle no long transactions intended
No conversions	Systems can have conversations
No adaption	Adaption intended

Information Retrieval Systems

Information retrieval (IR) is finding material (usually documents) of an unstructured nature (usually text) that satisfies an information need from within large collections (usually on local computer servers or on the internet) [61].

Information retrieval (IR) systems have in some sense the same goal as CBR systems because they are both trying to retrieve the best possible answers to queries. One can say that from the CBR process model only problem formulation and retrieval can be found in IR. There is, in particular, no reuse and adaptation [60].

Table 2-2: CBR systems versus IR systems [60]

Information retrieval system	***CBR systems***
Indexing of internet	Specific for a context
Can search on the Web	Search restricted to parts of the Web
Indexing by keywords	Indexing by features
No specific domain knowledge needed	Domain knowledge necessary
Heavily based on statistical knowledge	User domain and background knowledge
Query expansion	Conversational CBR
No adaptation	Adaptation intended

Case Based Reasoning and Machine Learning

Machine learning is a branch of Artificial Intelligence that focuses on the study of methods and techniques for programming computers to learn. Learning is the science of getting computers to act without being explicitly programmed [62].

Table 2-3: Machine Learning versus CBR Systems [60]

Machine learning systems	***CBR systems***
General rules and laws are generated	Specific solutions are generated

Results are not precise or certain	Results are not precise or certain
Degree of falsehood cannot be grasped	Similarity controls falsehood
Unsupervised learning possible	Unsupervised problem solving cannot be done

2.6. CBR Tools and Techniques

As stated by [37], CBR has various techniques. The most common and famous once are the following: Case representation, Indexing, Storage, Retrieval and Case adaptation.

There are different types of tools that can be used for developing a case base reasoning system. Most of these tools are commercials. The following case based reasoning tools are indicated on the paper of [63]:

Table 2-4: Case Based Reasoning Tools

Tools	***Platform***	***Representation***	***Retrieval***	***Interface***
ART Enterprise	PC, Workstation	Flat attribute: value pairs supporting a full range of variable types	Nearest-neighbor	Fully featured GUI builder
Case Advisor	PC ,Window	Flat records supporting text and weighted questions	Nearest-neighbor and Knowledge-guided	Use Netscape
CBR_3	PC, Window	Flat records supporting text and weighted questions	Nearest-neighbor and Knowledge-guided	Case point available as a DLL or API and CGI scripts
Eclipse	Any ANSI C environment	Flat attribute	Nearest-neighbor	No interface, only supply as a C library
ESTEEM	PC Windows	Case can be nested	Nearest-	Simply form-based

			neighbor with inductive weight generation	GUI builder
KATE	PC Windows and UNIX	Hierarchical Cases	Nearest-neighbor and induction	Toolbox interface can be customized
Caspian	Dos, Mac, or UNIX	Attribute-Value for feature representation	Nearest-neighbor	Can be integrated with a GUI
CASUEL	Portable	Frame-Like language for storing and exchanging descriptive models as ASCII file	Not Applicable	Not Applicable
CBR-Works	Ms Windows, Mac, or UNIX	Flat records supporting, text and weighted questions	Nearest-neighbor with support of feature weights	Fully featured GUI
MyCBR	Open source code written in java	Objects of the class described by its attributes	Query Model	User can customize the GUI and handle most of things

jCOLIBRI

jCOLIBRI is a technological evolution of COLIBRI and it is an object-oriented framework in Java which is designed for building CBR systems. It is a java-based and uses JavaBeans technology for case representation and automatically generation of user interface. This framework is developed by the GAIA artificial intelligence group in Complutense University in Madrid. The framework is built in two hierarchical levels- upper and lower. The lower level

consists of library of classes (Software modules) for full 4REs CBR cycle, also for definition of cases, attributes and connectors for access to outer databases. The upper level is “black box” – graphical interface, which allows non-complicated user CBR application generation based on lower level’s modules. jCOLIBRI supports full CBR cycle. At the retrieve stage the nearest N cases are retrieved. At reuse stage several methods for adaptation are available (direct proportion and also in ontology). At revise stage methods for revision of cases are realized, as well methods for new indices generation and methods for decision making (preference elicitation). At retain stage there are methods for query retaining to the case base for future use. jCOLIBRI allows retrieval from clustered and indexed case bases and submits program interfaces (connectors) to access text and XML files, as well standard and descriptive logic databases. These interfaces can be used for diagnostic systems database access. There are lots of CBR applications, developed on jCOLIBRI based: additional shells (abstract levels) for distributed CBR systems, statistical CBR systems, multi-agent supervisor systems for text file classification, and lots of CBR recommender systems.

After comparing the tools as shown in the table 2.4 the tools used for this research had to be chosen. Finding the right tools for the case based reasoning component was a challenging task. Open source tools were given priority because custom code or domain specific algorithm’s had to be coded as part of the CBR cycle. The researcher has good experience in programming .NET. So, one option was to start from scratch and build a custom CBR system. The obvious problem of this approach was the great deal of time it required and also becomes re-inventing the wheel. The other problem encountered was finding relevant, open source tools based on .NET, most were simple projects done by low reputation programmers, a good program was not found. Then the research tried to find tools based on the Java language. In Java a number of better tools such as CBRWorks and Orenge were considered. CBRWorks is developed for e-commerce applications but can be used for other purposes also. It contains elements from all knowledge containers and can perform the full CBR cycle. Orenge is a further development and has a more powerful retrieval engine.

After comparing those specified tools and others, finally the “white box” or open source version of JColibri framework was selected. It is a general framework that supports many features like graphical interfaces, description logics and ontologies, textual CBR, evaluation, and so on.

JColibri 2 [58] has added a number of features and is becoming more and more a reference tool for teaching and research purposes.

2.7. Application of CBR

CBR can be applied to any problem domain, in which information can be suitably encoded as cases. The underlying requirement is that similar cases should have similar solutions. It is particularly useful, when problem solving depends on expert experience or intuitive reasoning or when the underlying theory is difficult to encode or not very well understood. One of the main domains of CBR is in decision support systems. Historically, one of the earliest fields of interest in CBR was the field of law, which naturally deals with information encoded in the form of cases precedents. HYPO [64] is a CBR system that models how attorneys review past cases or precedents for trade secret disputes and infer legal arguments. [65] provide a survey of legal systems using CBR. A review of legal CBR systems with respect to argument schemes can be found in [66].

An overview of commercial CBR systems, ranging from customer-support help desk applications to engineering problem solving can be found in [67]. A variety of CBR applications has been reported in the literature. The use of CBR in electronic commerce has also started emerging recently. [68] describes his experience of using CBR for a Virtual Travel Agency. CBR also finds use in diagnostic systems, both in health care and in the industries. The oil and gas industry has used CBR system to reduce drilling costs and to increase safety. An interesting overview of CBR in drilling operations can be found in [69]. [70] also gives a good overview of CBR systems, in general.

2.8. Applications of Case Based Reasoning in Medicine

Many researchers are working on medical CBR with many diverse applications, ranging from psychiatry and epidemiology to clinical diagnosis. Most of them aim for a successful implementation of CBR methods to enhance the work of health experts, and to improve the efficiency and quality of healthcare. Early CBR in medical publications appeared in the late 1980‟s [71]; [72]; [73]; [74]. The applications of CBR in medicine focus mainly on diagnosis, classification, planning and tutoring. Many systems have been applied; some of these applications are discussed in the next.

Two milestone papers exemplify CBR coupled with evidence-based decision-making and CBR in a clinical setting [75]. In the first paper, [74] applied CBR to provide evidence-based medical decision-support over the Internet. They integrated automatic processing of medical practice guidelines and clinical pathways with CBR. It was called Care Partner and it used cases and prototypical cases to represent the variety and complexity of knowledge. This knowledge system was interactive and allowed physicians to select and organize diagnoses, treatment, follow-up actions and make recommendations according to the latest validated medical expertise. In the second paper, [76] successfully implemented CBR in the complex domain of intensive care medicine for antibiotic therapy decision support, where physicians are accountable for their decisions.

Table 2.5: Medical CBR Systems [75]

Task type	***Citation***	***System Name***	***Application Domain***
Diagnosis and decision support systems	[71]	*SHRINK*	Psychiatry
	[72]	*CASEY*	Heart Failure
	[73]	*MEDIC*	Dyspnoea
	[77]	*BOLERO*	Pneumonia
	[78]	*FLORENCE*	Healthcare Planning
	[79]	*MNAOMIA*	Psychiatry
	[80]	*SCINA*	Detection of Coronary Heart Diseases
	[74]	*CAREP-ARTNER*	Stem Cell Transplantation
	[81]	*AUGUSTE*	Alzheimer‟s disease
Classification Systems	[82]	*PROTOS*	Audiological
	[83]	*MACRAD*	Image Analysis
	[84]	*IMAGECREEK*	Image Analysis
	[85]; [86]	*PHYSYST*	Phylogenetic Classification
	[87]	*CTS*	Image Analysis
Planning Systems	[88]	*ALEXIA*	Hypertension
	[89]	*ROENTGEN*	Radiation Therapy
	[90]	*CAMP*	Daily Menu Planning

	[91]	*T-IDDM*	Diabetes treatment
Tutoring Systems	[92]	*ICONS*	Antibiotics therapy for intensive care
	[93]	*CADI*	Cardiac auscultation

Researchers who have contributed substantially to CBR in medicine include Gierl, Schmidt, and their colleagues, who focused on a range of applications including children dysmorphic syndromes, antibiotics therapy advising for intensive care, and monitoring emerging diseases [92]; [76]. Notable is their *ICONS* system [92], first applied to the determination of antibiotic therapy treatment for intensive care, and then to the prognosis of kidney function defects. For this latter application, *ICONS* learned prototypes associated with graded levels of severity through temporal abstraction [92] and matched new cases with these prototypes to predict the severity of a renal disease. Bichindaritz, Sullivan, Seroussi, Kansu, Potter, and their colleagues focused on hypertension diagnosis and therapy, psychiatric diagnosis, treatment and follow-up, stem cell transplantation long-term follow-up decision support, as well as phylogenetic classification [88]; [85]; [79]; [74]; [86]. In *ALEXIA* [88], CBR used a deep pathophysiological model of hypertension to assess the similarity between cases favoring the most important attributes in the model. *MNAOMIA* [79] is a CBR system that can adapt to the medical task at hand, namely diagnosis, treatment planning, and clinical research assistance. Later refined in *CARE-PARTNER* [74], this CBR system proposed a cooperation framework between cases and clinical practice guidelines in the domain of stem-cell post-transplant care.

Portinale, Bellazzi, Montani, Stefanelli, and their colleagues focused on diabetic patients‟ management, therapy and haemodialysis treatment [91]; [94]. Their system highlighted the importance of temporal abstraction to facilitate similarity assessment between cases, and thus retrieval [91].

Marling, Petot, Sterling, Whitehouse, and their colleagues [90]; [81], focused on designing individualized therapy plans (i.e., for nutrition menus, patient care for Alzheimer's, and cardiac and pulmonary disease). In particular, *AUGUSTE* [81] demonstrated the utility of CBR for supporting treatment planning in a medical domain lacking a strong domain theory. [87] focused on biomedical image interpretation, segmentation and similarity analysis for computed tomographies and microscopic images.

2.9. Related Works

Here in this section, the researcher discusses different research papers that are related to case based reasoning in different areas conducted locally in Ethiopia which helps to understand on how to solve and develop case based reasoning for diagnosis of chronic diseases.

In the domain of education, Biazen [95] develop a prototype case-based recommender system to advice students in field of study selection at higher education in Ethiopia. He used 105 cases which for building and testing the prototype and he achieves average performance of 77.2% and 80.2% by the domain experts and students respectively.

In the domain of health, Alemu [96] and Henok [14] developed a prototype knowledge based system using CBR techniques. Alemu [96] developed a CBR system for adverse drug reaction antiretroviral drug cases consultancy service. For building and testing the prototype Alemu [96] used 51 typical adverse drug reactions cases. Using 10 test cases, he achieved average recall and precision 72% and 63% respectively. Alemu [96] faces different challenges such as extracting features from free text, weighting importance of the extracted features and the adaptation system to be used. To overcome these problems, he proposed investigating natural language processing to extract features, using machine learning algorithms for feature weighting and investigating different adaptation techniques. Likewise, Henok [14] developed a prototype knowledge based system using CBR technique for hypertension management. He used 45 hypertension patient cases for building and testing the prototype. Using 7 test cases from the case base he registered 86.1% average recall and 60% average precision. The challenges that Henok faces and the recommendations proposed are almost similar to Alemu [96] except Henok proposed a hybrid knowledge based system using rule-based and CBR systems to increase the performance of the prototype. Similarly, Getachew [97] developed a prototype case-based reasoning system that can give decision support for anxiety disorder diagnosticians at a different level of expertise. Getachew performed testing of the prototype from two sides. The first one is testing in terms of precision and recall and registered 71% and 82% respectively. In addition to this, the average solution similarity using methods leave one out evaluator and hold out evaluation achieved performance of 73% and 75.5% respectively. The second one is the performance of the system is evaluated by the potential users of the system and achieved 83.2% performance.

Generally, the common features of all the systems are the case representation method and the retrieval algorithm the researchers used are similar. All the researchers used nearest neighbor retrieval algorithm, feature-value case representation method and the standard measures of performance in information retrieval to evaluate the effectiveness of their system (Biazen [95];2013; Alemu [96]; 2010;Henok [14], 2011;Getachew [97], 2012). But, Henok [14]; (2011) is unique from others by using accuracy performance evaluation to evaluate the reuse capability of his system. Thus, [98] enhances a rule based expert system with an object oriented database, and so it shows how traditional problem solving methods such as databases can be successfully enhance the performance of a knowledge based system. Among the future works recommended by Biazen [95] one is providing significant information about fields to students. Therefore in this research databases were used to store important facts about patient history consequently the steps of the case based reasoning cycle were enhanced using the database.

A numbers of local researches have been done research in the medical domain. The main intension of some researchers is to investigate the applicability of knowledge based system using CBR technique in the medical domain area. The researcher employs similar approaches such as CBR techniques to develop a knowledge based system. In addition most of the researches are implemented for specific domain area. One related work: Henok [14] used a case based reasoning system to experiment their hypothesis. Among the future works recommended by Henok [14], one is conducting future works on more chronic diseases so as to solve the shortage of domain experts. Therefore in this research includes four chronic diseases was implemented within one system rather than doing on a specific disease. Consequently, this helps domain experts to diagnosis the four chronic diseases.

Decisions about the management of chronic disease patients should not only take blood pressure levels, fasting blood sugar, TST 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 chronic diseases Whitworth [99]. And in Addis Ababa chronic diseases is widely spread Tesfaye et al. [19]. Therefore the main aim of the study is to investigate the applicability of cases based reasoning technique in designing KBS for chronic diseases diagnosis. This work will also be used as an input for further development of works in the medical domain area. Related works are summarized in the table below.

Table 2.6: Related Work Summarizations

Author	*Research title*	*Method /technique*	*Algorithm*	*Programing Tool*	*No. of case used*	*Retrieval*		*Accuracy*
						Recall	*Precision*	
Ethiopia (2002)	Application of CBR for Amharic Legal Precedent Retrieval	Feature value representation	NNR	jCOLIBRI	39	95.05%	82%	Not evaluated
Yemisrach (2010)	Application of CBR System in Legal Knowledge-Based System	Feature value representation	NNR	jCOLIBRI	40	87.5%	41.01 %	Not evaluated
Alemu (2010)	CBR for adverse drug reaction Antiretrovial Drug	Feature value representation	NNR	jCOLIBRI	51	72%	63%	Not evalua

	Cases Consultancy Service							ted
Biazen (2013)	CBR Field of Study Selection	Feature value representation	NNR	jCOLIBRI	105	85%	55%	Not evaluated
Henok (2011)	CBR Hypertension management	Feature value representation	NNR	Python	45	86.1%	60%	88.89 %
Getachew (2012)	CBR for anxiety disorder diagnosis	Feature value representation	NNR	jCOLIBRI	50	82%	71%	Not evaluated

CHAPTER THREE

KNOWLEDGE ACQUISITION AND MODELING

3.1. Knowledge Acquisition

One of the most important knowledge based system development activity is knowledge acquisition. Knowledge acquisition is formally defined as "the transfer and transformation of problem solving expertise from some knowledge source to a computer program" [100]**.** The process of knowledge acquisition typically involves a knowledge engineer gaining "procedural and declarative knowledge" from a human expert in the subject domain [100]**.** 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 [101]**.** 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: [102]**.**

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 [102]**.**

Therefore, the main objective of this chapter is to collect chronic diseases patient's cases and model how chronic diseases patients are treated. The main features regarding to chronic diseases diagnosis and the fundamental concepts are identified about the diagnosis procedures that help to

develop the proposed prototype CBR system. Generally, this chapter covers the knowledge (cases) acquisition phase and the conceptualization process. The basic attributes and concepts, which are crucial in the diagnosis process, are identified.

3.2. Pre-Acquisition Phase

Before the actual knowledge acquisition process, the researcher has conducted a preliminary investigation in the domain area. During the preliminary investigation a numbers of stakeholders from different domain, i.e., medical director, general practitioner, nurses, specialist and medical students are participated. These stakeholders are used to understand the dimensions of the domain area. During this phase, the researcher has tried to conduct pilot study and informal discussion with the domain experts to understand the dimension of the problems. In addition, secondary sources of knowledge are reviewed on each proposed disease. Based on the finding of pilot study, the researcher constructed semi-structured interview questions in a way that helps to acquire the required knowledge from the domain experts.

Therefore, the main objective of this chapter is to interact with domain experts in order to collect, gather and analyze the required knowledge. The knowledge content covers issues such as how general practitioner diagnosis patients based on their pillar symptoms. The acquired knowledge is then represented using case based reasoning.

3.3. The Process of Knowledge Acquisition

For the purpose of this research, the process of knowledge acquisition includes some basic activities such as interview of domain expert's and review of relevant sources of information. The objective of knowledge acquisition is to gather the required knowledge, interpreting the acquired knowledge, analyzing and validating the knowledge content. Based on the acquired knowledge, the proposed knowledge base system is designed using hierarchal tree model. Therefore, knowledge acquisition process of this thesis is based on domain expert interviewing and reviewing of related documents. This section discusses the detail of knowledge acquisition techniques as follows.

3.3.1. Interviewing Domain Experts

Primary sources of knowledge are collected from human experts in the domain area at Adama Referral and Medical Hospital. To gather the required knowledge semi-structured interview

technique is used. Since one of the main focuses of this research is eliciting relevant tacit knowledge from the domain experts, four (4) domain experts were selected. During the interview phase, the main challenge is willingness of domain expert to share their expertise and experience. To handle these challenge domain experts were selected based on their interest. The domains of interview with expert covered issues such as how the expert interact with patients, what are the techniques used to identify the pillar symptoms, the procedures of diagnosing and what are the possible treatments.

During the extensive discussion, the researcher tries to acquire the relevant tacit knowledge which is significant to generate the proposed CBR system. These domain experts are practically participated throughout the research work and they were consulted to confirm the correctness of the acquired knowledge. During face to face communication, the acquired knowledge from domain experts has been recorded manually by using pen and paper sheet. In addition, secondary source of knowledge has been gathered from the internet, standard treatment guideline (STG) for Ethiopians health professional, manuals, research papers and journal articles. But, the primary data source is patient history of Adama Referral and Medical Hospital of chronic diseases patients. Profiles of domain experts participated in the interview process are presented in the Table 3.1 below.

Table 3-1: Domain Experts' Profiles

No	*Educational level*	*Specialization*	*Area interviewed*	*Role*	*Quantity*
1	MD	Medicine	Medical Category	General	3
2	MD	Medicine	Medical Category	Medical director	1

As indicated in the Table 3.1 above, the health professionals are devoted to providing health care services in the institution. In addition, they participate in conducting research and teaching of medical students in the university. According to the domain experts, the investigation of patient health problem starts by collecting some relevant information such as location, age, and family history of the patient. History of patient is important attribute to determine the types of health problem; because some diseases have unique associations with geographical location, age factors and family inheritance case. The actual diagnosing task of the physician started using a medical instrument which is called vital sign. This instrument is used to investigate particular part of

patient body affected by certain diseases or unknown feeling. According to the domain expert vital sign are classified into four main categories. Each category of vital sign is discussed as follows:

1. ***Pulse rate:*** the physician check the pulse rate of the patients. There are different parameters which are used to check the normal health condition of patient is usually between "60-90" beats per minutes. But, Patient pulse rate above or below of the given interval shows that he/she is not at normal condition. This measurement helps the physician to understand the health status of the patient related to their pulse rate.
2. ***Respirator rate:*** The normal condition of respirator rate is usually between"12-24" breaths per minutes. Respirator rate of an individual is between the above intervals shows that the patient respiratory rate is at regular pattern. But, if the patient register respirator rate below or above this interval, it indicates that the patient has some complication in relation to his/her respirator rate.
3. ***Temperature measure:*** usually the normal condition of human body temperature is between"36.5 - 37.2" degree C, or Celsius. Above and below this normal condition shows the abnormal health status of the patients. Therefore, the increment and decrement in temperature helps the physicians to identify underlying causes that affect the normal condition.
4. ***Blood pressure:*** blood pressure is used to measure the flow of blood in our body circulation system. The normal range of human blood pressure is 120/90 mmHg, the upper threshold can increase from 90 to 140 mmHg and the lower thresholds are "between 65 and 85" mmHg which is considered as normal state.

Therefore, human expert uses the above four tools and laboratory test results as inference mechanism in addition to the pillar symptoms collected from the patient during chronic diseases examination.

To select a sample of the proposed disease the researcher considers different factors during the knowledge acquisition process. Firstly, the most widely occurring and principal chronic diseases types are identified. Secondly, diseases which are not prevalence to the research area are determined. Thirdly, the diseases with frequently changed sign and symptoms are identified; since they can create difficult during the diagnosing process. The proposed knowledge base

system is developed based on the pillar symptom of each disease in the case base. Therefore, the diseases are selected purposively by considering the above factors.

3.4. General Structure of Chronic Diseases

Disease can be classified as acute or chronic. An acute disease starts quickly and lasts for a short period the patient either recovers or dies. A chronic disease lasts for a long time and weakens the patients. In addition, WHO defines chronic diseases as" diseases of long duration and generally slow progression", and the United States Centers for Disease Prevention and Control defines them as "conditions that are not cured once acquired…are considered chronic….Additionally, other conditions must have been present 3 months or longer to be considered chronic" [103].Disease are said to be infectious(will spread from one person to another) or noninfectious (will not spread from one person to another).Even though, most of chronic disease are noninfectious(Non-communicable disease) there are also disease that are chronic infections such as TB, HIV AIDS, etc. For this study, with discussion of domain experts the researcher includes TB as one of chronic infectious disease. Because, in Ethiopia TB is the most commonly occurring with the highest morbidity and mortality rate.

Therefore, the management of chronic diseases is complex and requires a well-trained health force. Decisions shall be made within a time-sensitive environment, with limited manpower and information (response of the patient). Especially, the situations become more difficult for those physicians who do not have a medical diagnosis experience. To handle such complexity, general practitioners are required to possess specialized knowledge as well as experience with a wide range of illness and injuries.

For the purpose of this research, Chronic diseases are classified into four categories namely, *Cardiovascular diseases, Endocrinal condition, Infectious diseases, Respiratory diseases*. From each category the researcher select a sample of disease based on the domain expert's recommendation. According to domain experts those selected diseases are most significant disease in the domain area and affect human health as they observed during clinical diagnosis process.

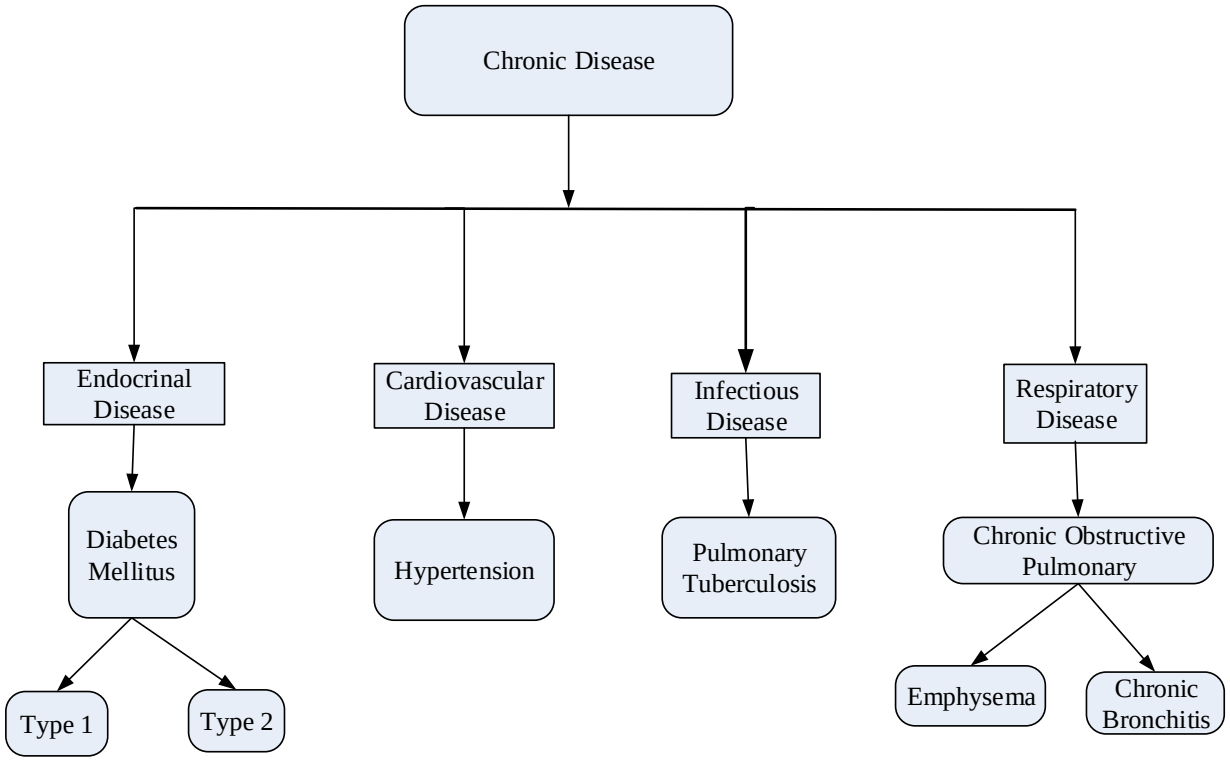

Figure 3-1: Block Diagram of General Structure of Chronic Diseases

3.4.1. The Categories of Chronic Diseases

Figure 3.1 above block diagram of chronic diseases are categorized into Cardiovascular, Endocrinal, Infectious, Respiratory diseases. For this study sample of diseases (cases) are selected from each category. For example, from Cardiovascular Disease (Hypertension) is selected. From the Endocrinal condition (Diabetes mellitus) is selected. From Infectious Disease (Pulmonary TB) is selected. Finally, from Respiratory Disease (chronic obstructive pulmonary) is selected. In addition, diabetes mellitus and chronic obstructive pulmonary have two sub category each of them. For this research a total of 164 cases are used to build the case base with 25 hypertension cases, 53 diabetes mellitus cases, 46 pulmonary TB cases, 40 COPD cases are collected. Therefore, this section discusses the characteristics, symptoms and possible treatment of each disease in their respective category.

1. **Endocrinal Condition**

It is signs and symptoms of disease relating to endocrine gland dysfunction, e.g. diabetes mellitus, Cushing's disease, Addison's disease, thyrotoxicosis, Graves' disease. Endocrine gland

is a gland that secretes a substance (a hormone) into the bloodstream. We chose chronic diseases that [104] are the most commonly occurring and with the highest morbidity and mortality, and, among these, disease that [11] do not typically have a specialist as the primary physician that is diabetes mellitus.

Diabetes Mellitus

Diabetes is a long-lasting disease that happens when the human body has problem in an accurate manner governing the amount of glucose in the blood. It is a chronic diseases in which the human body's cells either do not respond properly to insulin or insulin production becomes insufficient. To know diabetes, it is important to know the role of glucose acts in connection with the body, and what can occur when governing of glucose goes wrong and the amount of blood glucose level comes to be hazardously low or high [105]. Insulin plays a key role to open the cells, letting them accept and use the existing glucose [106]**.** Cells take glucose from the blood with the help of insulin. The amount of glucose level decreases as glucose moves out from the blood and moves into the cells. It can be assumed that insulin serves as a channel for transporting glucose between the blood and cells. It is essential to know when the amount of insulin level increases, the amount of glucose level in the blood decreases. Since the amount of glucose moves into the cells are going to be used for energy by the cells of the human body, the body is considered to control and buffer the amount of sugar in the blood stream to keep up a stable amount to satisfy the desires of cells. One of the human body's organ, pancreas generates, stores and discharges into the blood to get the amount of glucose back. However, individuals living with diabetes either cannot produce this hormone in the right manner or their body cannot properly use it for normal body functions. Therefore, this creates the accumulation of glucose in the blood and when the accumulated glucose is not used by the body cells, then signs and symptoms of diabetes start to emerge [105] .

As noted in [105], the following are the common signs and symptoms of diabetes.

- Polydipsia or excessive thirst
- Polyuria or frequent urination
- Polyphagia or excessive hunger
- Unusual weight loss

- Tiredness
- Blurred Vision
- Dizziness

If the amount of glucose in the blood becomes very high and left untreated, then it brings unconsciousness and probably passing away. The signs and symptoms of diabetes could differ depending on the types of diabetes. Domain experts usually recognize diabetes by the appearance of at least the above first three basic symptoms, i.e. polyuria (frequent urination), polydipsia (excessive thirst) and polyphagia (excessive hunger).

There are two main types of diabetes [106]**.** These are Type_1 and Type_2 diabetes. Sometimes, Type_1 diabetes is called insulin-dependent diabetes.

Type1 Diabetes is induced by an inappropriate immune response of the body, where the body's immune system begins to destroy β cells by eliminating production of insulin in the pancreas. Hence, the body cannot bring about the required insulin. The disease can have an effect upon individuals of any age, but most of the time it begins in individuals having the age less than 35. Individuals having this type of diabetes require injections of insulin daily so as to control the amounts of glucose in their blood. In the absence of insulin, individuals with type1 diabetes will pass away. Type_1 diabetes rapidly develops and can comprise symptoms like [107].

- Polydipsia or excessive thirst
- Polyuria or frequent urination
- Polyphagia or excessive hunger
- Unusual weight loss
- Tiredness
- Blurred vision
- Slow-healing wounds
- Recurring infections

Type_2 Diabetes is the most common type of diabetes. It ordinarily happens in adult individuals, but is progressively realized in youngsters. In type_2 diabetes, the body is capable to yield insulin. However, it is either insufficient or it has difficulty in using it properly leading to an accumulation of glucose in the blood. Individuals with type_2 diabetes may stay unconscious of their disease for a prolonged period of time since symptoms may require years to come out or

be realized, during which time the body is being discredited by surplus blood glucose. According to International Diabetes Federation [107] the reasons for developing type_2 diabetes are still not understood though there are some essential risk factors that are mentioned below.

- Obesity or more than average fatness
- Poor nutrition
- Physical inactivity or sedentary lifestyle
- Increasing age
- Genetic background of diabetes
- Ethnicity
- Poor diet during gestation affecting the developing child

When we compare individuals with type_1 diabetes and type_2 diabetes, most individuals with type_2 diabetes do not ordinarily need injections of insulin daily basis. However, they may be dictated to take insulin plus a healthy nutrition and increase physical exercise to control their condition. The number of individuals with type_2 diabetes is increasing very quickly throughout the world. According to National Diabetes Information Clearinghouse [108], blood tests are applied for diagnosis of diabetes and pre-diabetes since at first patients having type_2 diabetes may not show symptoms. Blood tests for all types of diabetes include taking blood at offices of medical institution and sending the blood sample to a laboratory for examination. The following is one of the most common types of laboratory tests that the domain experts should know that will help them in order to make decisions:

A. **Fasting Blood Sugar (FPS) test:** It is applied to diagnose and identify diabetes and pre-diabetes. It is the best screening test applied for diagnosing diabetes since it is easier and faster to perform, more convenient and acceptable to patients, and less expensive. This concept is shown in table 3.2 as follows.

Table 3-2: Concepts of Lab Test Results

Attribute Name	***Values in mg/dL***	***State***
Lab result using FPS test	126 or above	Diabetes
	100 to 125	Pre-diabetes
	99 or below	Normal

Since there are two types of diabetes, the domain expert also has another concept of diagnosis to know in which type the patient may fall. Age is the main concept of the domain experts to classify the patient's diabetes type.

B. **Concepts of Age**

The third concept of domain experts in diagnosing patients is age of patient. Domain experts first gather the general symptoms from patient and order a laboratory test. If the lab test result indicates diabetes, the next question that they have to raise is asking age of patient. Knowing age of patient helps domain expert in trying to classify in which diabetes type the patient be under. Most domain experts said that types of diabetes can be classified by age of patients. But in rare cases this does not work. The domain expert's concept and decision passed is summarized in table 3.3 below.

Table 3-3: Concepts of Age

Attribute Name	***Values***	***Type of diabetes***
Age	≤ 15	Type 1
	> 15 to 35	Type 1 or Type 2
	> 35	Type 2

After the concept of age is analyzed, the domain experts have another concept that can help them for diagnosing the new patient. This concept is family history of the patient related to diabetes.

C. **Concepts of Family History of Diabetes**

The domain experts have a concept of family history of diabetes. The family history of diabetes may cause the patient to be diabetic. By considering this the domain expert asks the patient for his/her family history of diabetes, then the domain expert can decide the patient has Type2 diabetes, especially for patients with age between 15 and 35 years. This concept is shown in the table 3.4 as follows.

Table 3-4: Concepts of Family History of Diabetes

Attribute Name	***Values***	***Age***	***Decision***
Family history of diabetes	Exists	> 15 to 35	Type 2 diabetes
	Doesn't exist	> 15 to 35	?

After this concept the next concept that domain experts used is the concept of ketone and/or autoantibodies existence in the blood during laboratory testing.

D. **Concepts of Ketone and/or Autoantibodies**

This concept is the most useful concept in order to conclude whether the patient is diabetic

Type_1 or other diabetes type. If the age of the patient are between 15 and 35 years old and did not exist family history of diabetes. Summary of result of this concept is shown in table 3.5 as follows.

Table 3-5: Concepts of Ketone and/or Autoantibodies

Attribute Name	***Values***	***Age***	***Family history of diabetes***	***Decision***
Ketone and/or autoantibodies	Exists	> 15 to 35	doesn't exist	Type_1 diabetes
	Doesn't exist	> 15 to 35	doesn't exist	Other diabetes types

2. Cardiovascular Disease

Diseases affecting the heart or blood vessels include arteriosclerosis, coronary artery disease, heart valve disease, arrhythmia, heart failure, hypertension, orthostatic hypotension, shock, endocarditis, diseases of the peripheral vascular system, congenital heart disease. We chose chronic disease that [104] are the most commonly occurring and with the highest morbidity and mortality, among these, disease that [11] do not typically have a specialist as the primary physician that is hypertension.

Hypertension

It is a chronic disease, which is 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. 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 [109].

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.

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. The range of SBP and DBP is summarized in the following table:

Table 3-6: Level of Hypertension in Adult

Range of SBP in mmHg	***Range of DBP in mmHg***	***Level of hypertension***
SBP 140 – 159	DBP 90 – 99	Mild
SBP 160 – 179	DBP 100 – 109	Moderate
SBP > 180	DBP≥ 110	Severe

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. 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 and (TBP) is < 140/80 mmHg for patient without diabetes and renal failure.

Blood pressure control can be achieved by lifestyle modifications and as necessary, pharmacologic treatment.

3. Infectious Diseases

Infectious diseases are caused by pathogenic microorganisms, such as bacteria, viruses, parasites or fungi; the diseases can be spread, directly, from one person to another.

Pulmonary Tuberculosis (TB)

Pulmonary Tuberculosis (TB) is a chronic bacterial infection caused by Mycobacterium tuberculosis. It is spread through the air and usually infects the lungs, although other organs and parts of the body can be involved as well. Most people who are infected harbor the tuberculosis bacterium without symptoms. This is known as latent tuberculosis. If the body's resistance is low because of aging, malnutrition, infections such as HIV, or other reasons, the bacteria may break out of hiding and cause active tuberculosis.

Domain expert suggested that common symptoms such as long term cough (more than two weeks), fever, night sweat and weight loss are the pillar symptom of pulmonary TB disease.

Since considering only the symptoms of pulmonary TB are not enough to conclude that whether the patients have pulmonary TB or not, laboratory test result is the key part and the most important concept for the decision.

The examination laboratory test of pulmonary TB patient consists of a skin test, possibly a blood sample and a chest x-ray for persons greater 15 years old. One of methods domain experts identify whether the patient has pulmonary TB or not is using tuberculin skin test (TST).

The major symptoms of pulmonary TB include:

- Productive Cough
- High Fever
- Night Sweat
- Hemoptysis
- Weight Loss
- Amount of fluid increases dramatically
- Compress the lung
- Shortness of breath

- Chest pain

Treatment of pulmonary tuberculosis: Tuberculosis is usually treated with tablets for six months. To be cured it is vital to follow the treatment and completed in the correct manner as it prescribed by doctor.

Concepts of Family History of TB disease

The domain experts have a concept of family history of pulmonary TB. The family history of pulmonary TB may cause the patient to be pulmonary TB affected. By considering this the domain expert asks the patient for his/her family history of pulmonary TB, then the domain expert can decide the patient has pulmonary TB disease, especially for patients with pulmonary TB can easily transmitted to other by contacting and sneezing on the air.

4. Respiratory Disease

Chronic respiratory diseases (CRDs) are diseases of the airways and other structures of the lung. This disease could be characterized by a high inflammatory cells recruitment and/or destructive cycle of infection. Some of the most common are asthma, chronic obstructive pulmonary disease, or acute respiratory distress syndrome . CRDs are not curable, however, various forms of treatment that help dilate major air passages and improve shortness of breath can help control symptoms and increase the quality of life for people with the disease [110].

Chronic obstructive pulmonary diseases (COPD)

Chronic obstructive pulmonary disease is characterized by airflow limitation that is not fully reversible**.** COPD may include diseases such as emphysema, chronic bronchitis or a combination of these disorders. COPD is generated by the abnormal inflammatory response of the lungs to noxious particles or gases. Environmental exposures and host factors are risk factors of COPD, and the most severe risk factor is cigarette smoking. There are three primary symptoms of COPD: cough, sputum production, and dyspnea on exertion**.**

A. Emphysema

Emphysema is a respiratory disease. In this condition, millions of the lungs' tiny air sacs (alveoli) stretch out of shape or rupture. As these thin, fragile air sacs become damaged or

destroyed, the lungs lose their natural elasticity. Air stays trapped inside the lungs because the damaged lung sacs can't empty.

The major symptoms caused by emphysema include:

- Wheezing
- Shortness of breath
- Tightness feeling in the chest

Treatment

No treatment can reverse or stop emphysema. But treatment can help to:

- Relieve symptoms
- Treat complications
- Minimize disability

Doctors' top treatment advice is to quit smoking. This is the single most important factor for maintaining healthy lungs.

Medical treatment

Medications may include:

- Bronchodilators Inhalers
- Corticosteroids
- Antibiotics
- Oxygen therapy
- Pulmonary rehabilitation
- Lung volume reduction surgery
- Lung transplant

B. Chronic Bronchitis

Chronic bronchitis is defined as a cough that occurs every day with sputum production that lasts for at least 3 months, two years in a row. Bronchitis is an inflammation (or irritation) of the airways in the lungs. Airways are the tubes in your lungs that air passes through. They are also called bronchial tubes.

Major Symptoms of Chronic Bronchitis

- Production of mucus (sputum), which can be clear, white, and yellowish-gray or green in color-rarely, it, may be streaked with blood.
- Frequent respiratory infections
- shortness of breath

Treatment

- Quitting smoking
- Inhaled medicines
- Antibiotics
- Oxygen therapy machine
- Surgery

3.4.2. Knowledge Acquisition from Chronic Diseases Cases

After the knowledge acquisition form the main knowledge that is used for the case based reasoning system are collected from chronic diseases complaints 'card from Adama Referral and Medical Hospital. 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 diagnosis the chronic diseases 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 chronic diseases patients' record file are: Age, Sex, Disease Categories, Present Illness1, Present Illness2, Premonitory Symptoms, Fasting Blood Sugar, TST, Systolic blood pressure, Diastolic blood pressure, Co-morbidity, Co-medication, Smoking and Drinking Status, Family History of a patients. As compared to variables that are identified from the domain experts and relevant document, variables like race, height, and weight are not found in the patient record file.

3.5. Knowledge Modeling

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 [111] . 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 the knowledge. 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 [112]; [96]; [113]; [14]. The next sections illustrate the conceptual models for chronic diseases diagnosis and treatments.

3.5.1. Case Concept for Chronic Diseases diagnosis and treatment

In diagnosis of chronic diseases cases, the complaints profile, the disease category and the risk factors are checked in order to diagnosis and treat the chronic diseases cases. The general concept in diagnosis of chronic diseases case is found in the figure 3.2 below.

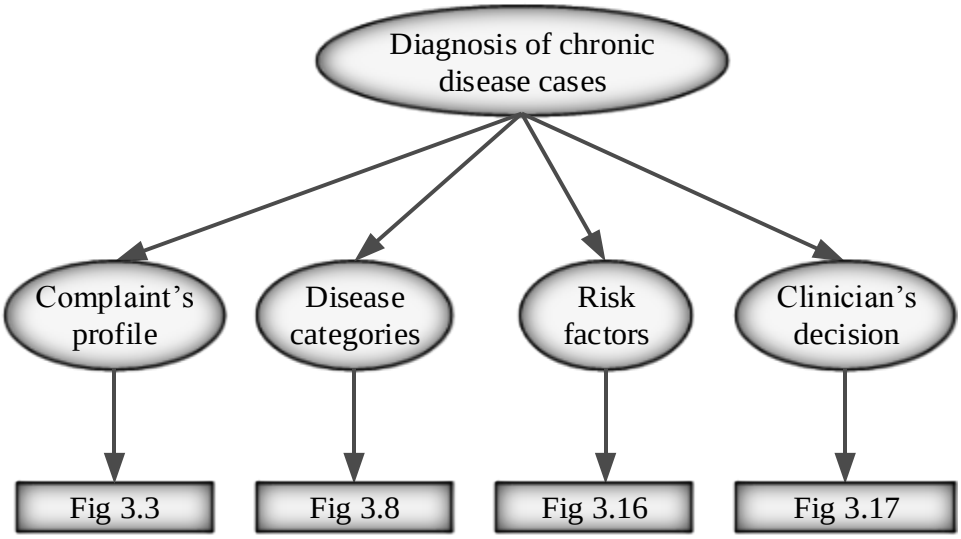

Figure 3.2. General Concept for Diagnosis of chronic disease cases

The patient profile, disease categories, risk factors and clinician's recommendation of chronic diseases are the main factors that are considered in diagnosing chronic diseases.

3.5.1.1 Chronic disease Complaint's Profile

Complaints' profile, which is needed in chronic diseases treatment, is depicted in figure 3.3.

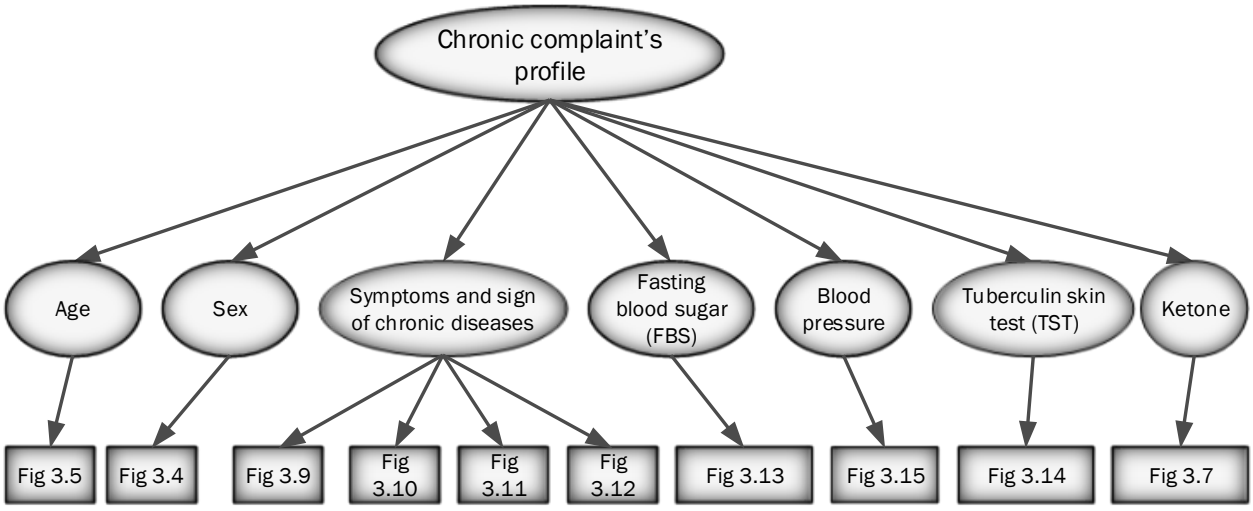

Figure 3.3. Concepts for Chronic Diseases Patient Profile

Age, sex, blood pressure records, TST, FBS, family history and the symptoms of compliant information are analyzed by the clinician for the treatment of chronic diseases cases.

1. Sex of Chronic Diseases Complaint's

Sex of the patient is one of the risk factor in chronic diseases especially for hypertension and pulmonary TB cases. 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 3.4 below.

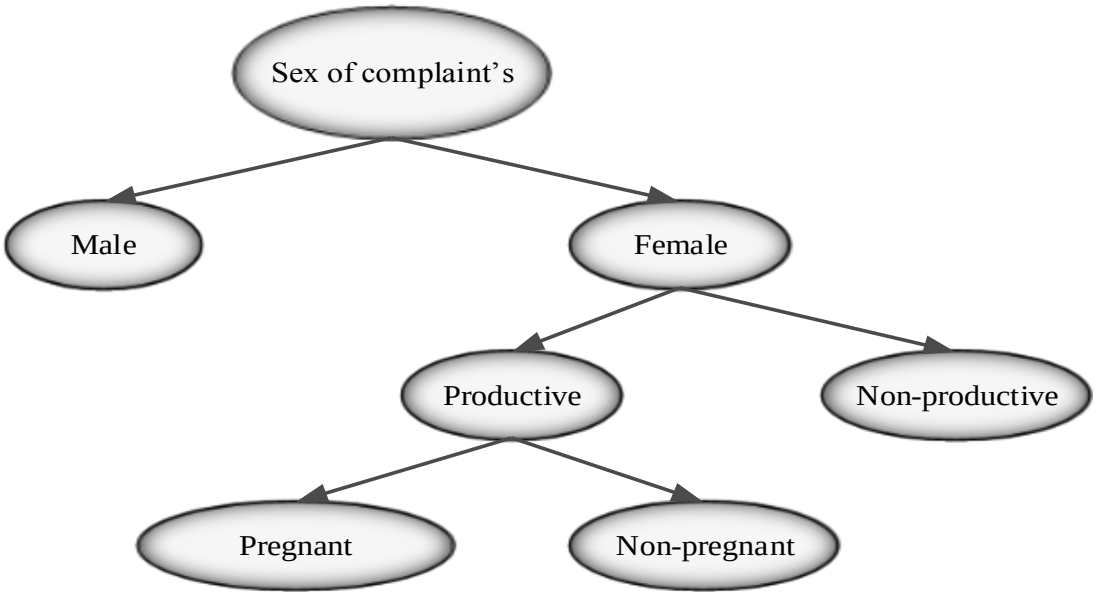

Figure 3.4. Concepts for Sex of the Compliant

2. Concepts of Age, Family history and Ketone for chronic disease Diagnosis

The concepts of age, family history and ketone for chronic disease diagnosis will be discussed in the following sections.

A. Concepts of Age of a complaint's

Domain experts first gather the general symptoms from patient and order a laboratory test. If the lab test result indicates diabetes, the next question that they have is asking the age of the patient. Knowing age of patient helps domain expert in trying to classify in which diabetes type the patient be under.

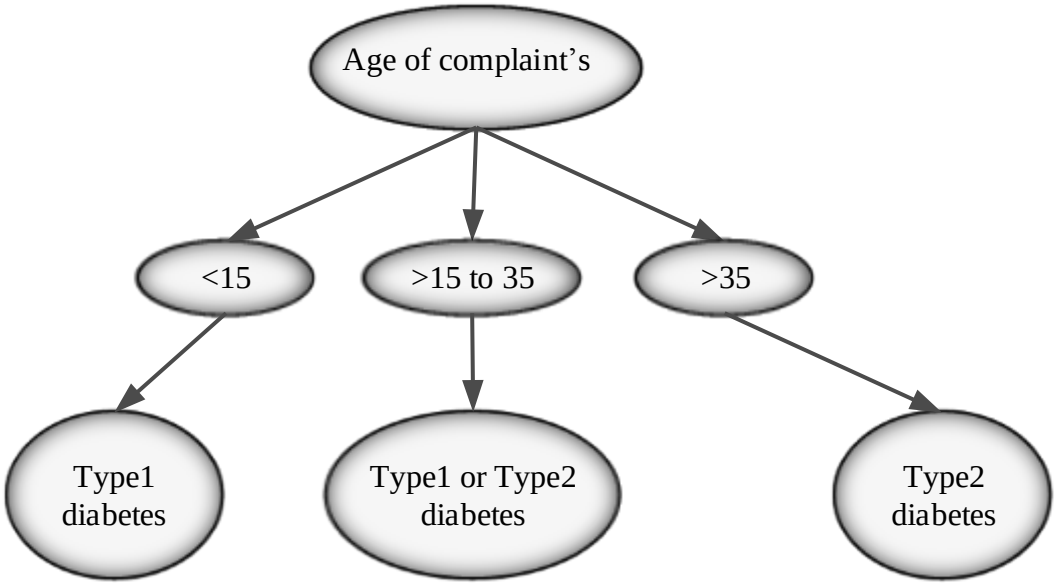

Figure 3.5.Concept of Age of a compliant

After the concept of age is analyzed, the domain experts have another concept that will help them for diagnosing the new patient. This concept is family history of diabetes will help to decide for. Similarly, family history is also important for chronic diseases diagnosis.

B. **Concepts of Family History of Chronic diseases**

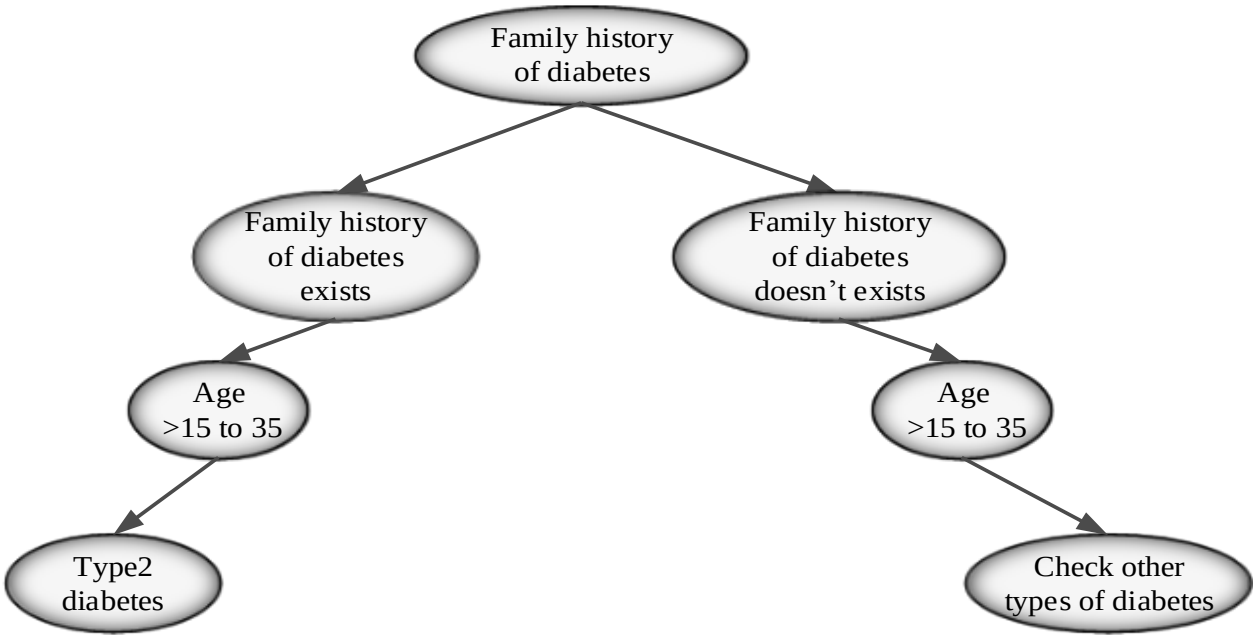

Figure 3.6.Concept of family history of diabetes mellitus

As shown in the above figure 3.6 the concept of family history of diabetes may cause the patient to be diabetic. If family history of diabetes exists, age of the diabetic patient is between 15 and 35 years old it is type_2 diabetes. Similarly, if family history of diabetes doesn't exist, age of the diabetic patient is between 15 and 35 years old the domain experts check other types of diabetes. After this concept the next concept that domain experts used is the concept of ketone and/or autoantibodies existence in the blood during laboratory testing.

C. Concepts of Ketone and/or Autoantibodies of diabetic complaint's

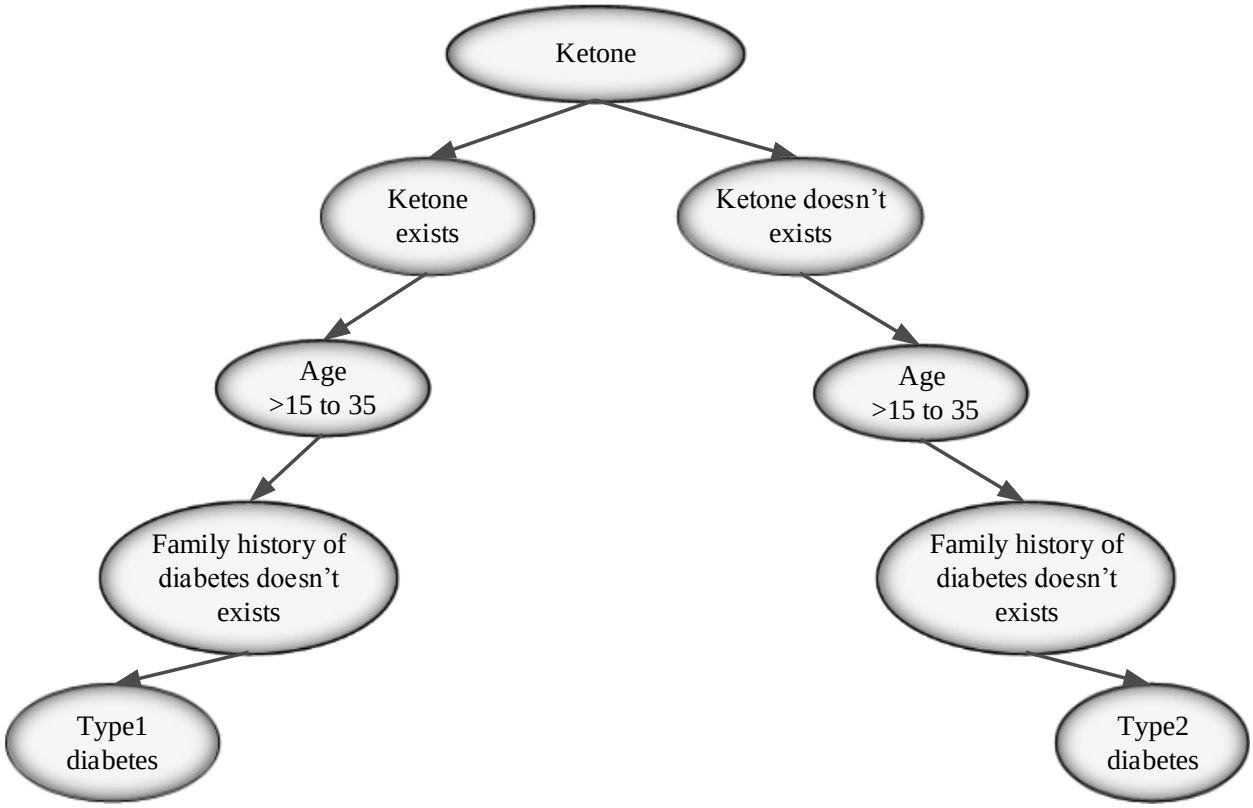

Figure 3.7.Concept of ketone of a compliant

As shown in the above figure 3.7 the concept of ketone is the most useful concept in order to conclude whether the patient is diabetic Type_1 or other diabetes type. If ketone exists, age of the diabetic patient is between 15 and 35 years old and family history of diabetes doesn't exist it is type_1 diabetes. Similarly, if ketone doesn't exist, age of the diabetic patient is between 15 and 35 years old and family history of diabetes doesn't exist it is type_2 diabetes.

3.5.1.2. Concept for Disease Categories of Compliant's

Disease categories of compliant are analyzed by the clinician for diagnosis of chronic diseases cases.

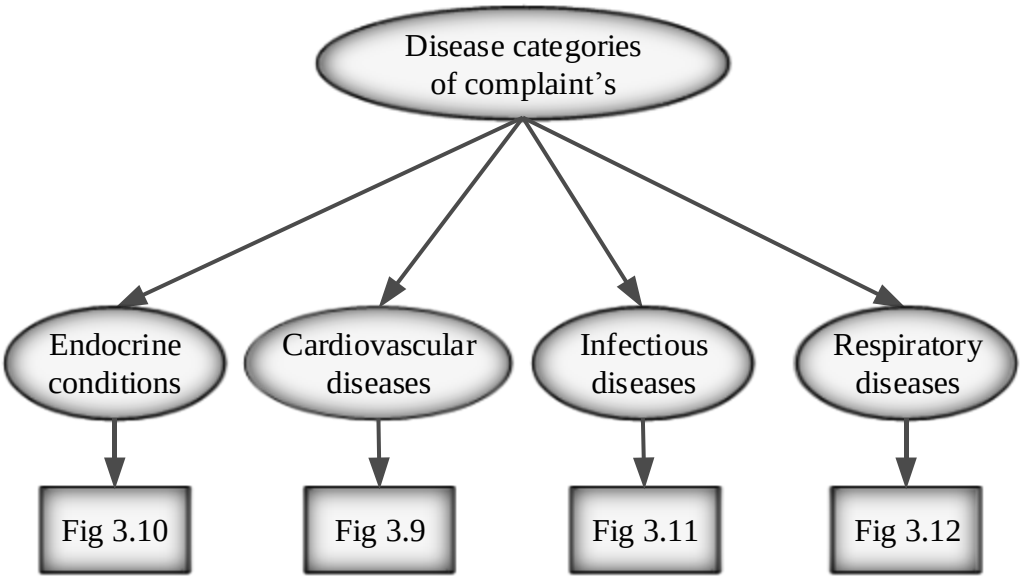

Figure 3.8.Concept for disease category of a compliant

As show in the above figure 3-8 category of chronic diseases are endocrine conditions, cardiovascular, infectious, respiratory diseases. According to domain experts those selected diseases categories are most significant diseases in the domain areas and affect human health that they observed during clinical diagnosis process. From those categories the researcher had taken as sample diseases those are shown by figure under each category that are common diseases that mostly occurred in the domain areas.

3.5.1.3 Concepts of Symptoms and sign of Chronic Diseases

The concepts of symptoms and sign of each type of chronic diseases will be discussed in the following sections.

A. Concept of Symptoms and Sign of Hypertension

Figure 3.9 depicts the possible common symptoms and sign that can appear with hypertension patients.

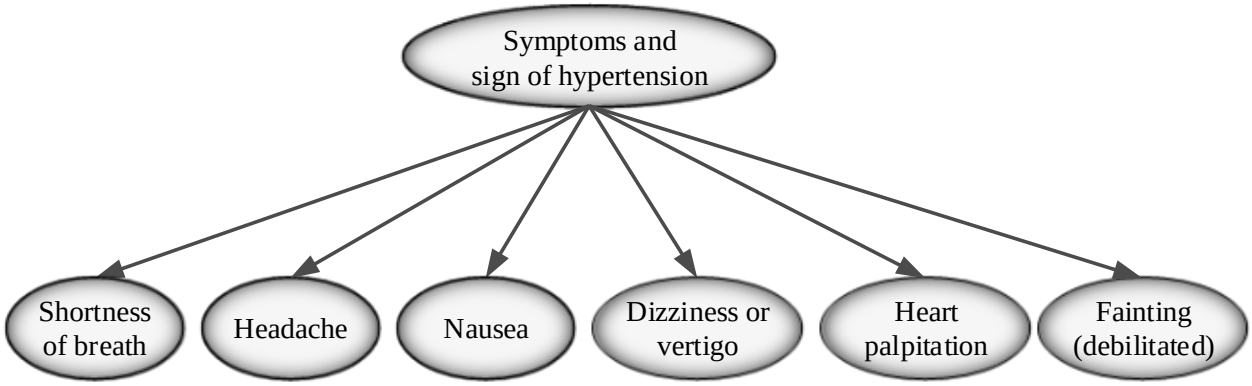

Figure 3.9.Sign and Symptom of Hypertension patients

The possible symptoms of hypertension vary quite a lot from person to person. These symptoms can be symptoms of other health problems, however as shown in above figure 3.9 a few of the more common symptoms that the hypertension patients can have. After we check the common symptoms of hypertension then we send for laboratory test that is shown in the figure 3.15 and then domain experts check risk factors of a complaint's is shown in the figure 3.16.

B. **Concept of Symptoms and Sign of Diabetes Mellitus**

In the diagnosis of diabetes, the domain experts have a concept of symptoms that is used to differentiate the related symptoms of diabetes and the non-related symptoms of diabetes. During process of knowledge gathering through interview, the domain experts (physicians) explained that the following are the symptoms.

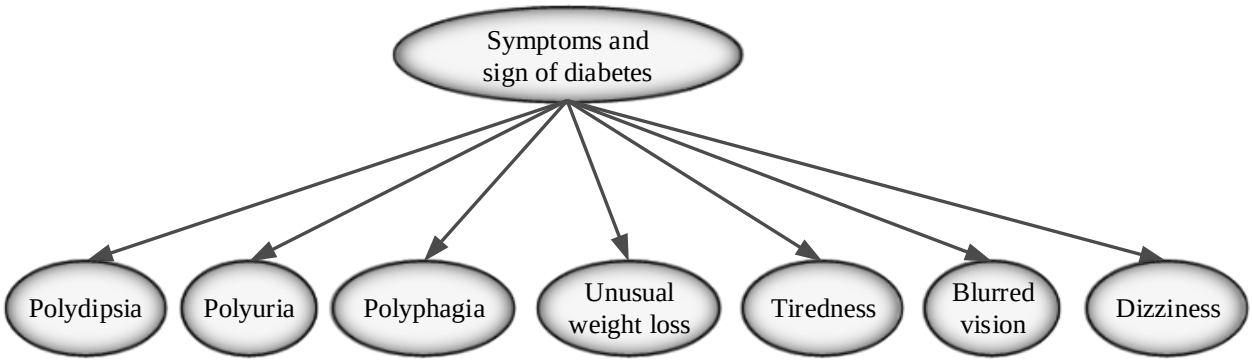

Figure 3.10. Sign and Symptom of diabetes mellitus patients

Domain experts usually recognize diabetes by the appearance of at least the above first three basic symptoms, i.e. polyuria (frequent urination), polydipsia (excessive thirst) and polyphagia (excessive hunger).

After checking symptoms in the patient, the next concept that the domain experts used is the concept of laboratory test result that is shown in the figure 3.13 and figure 3.15 then domain experts check risk factors of a complaint's that is shown in the figure 3.16.

Concept of Symptoms and Sign of Pulmonary Tuberculosis (TB)

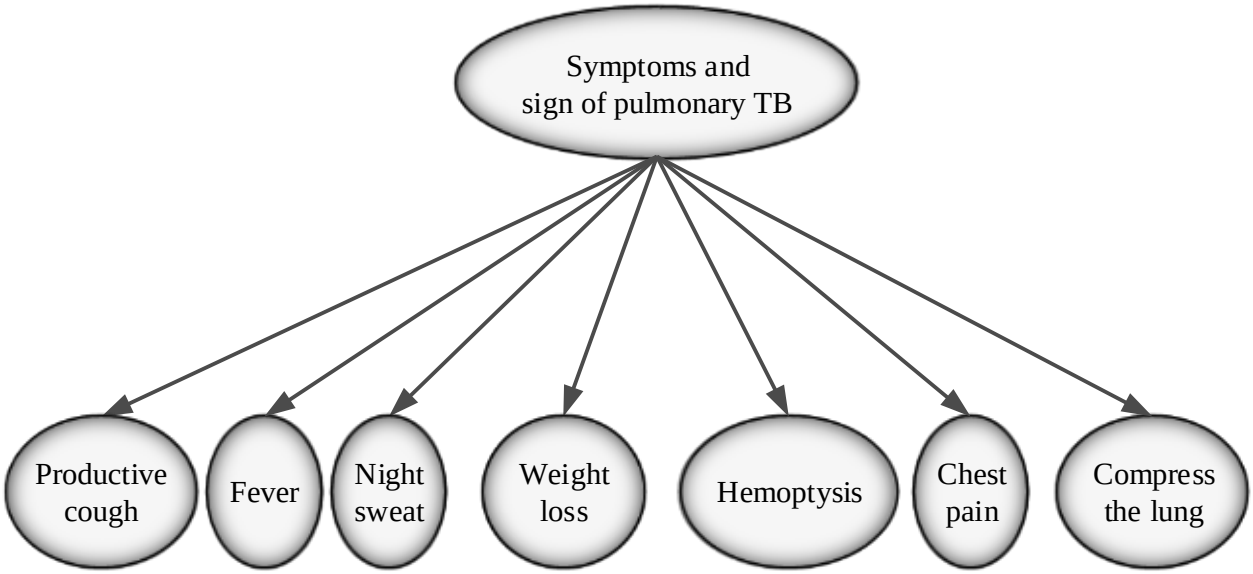

Figure 3-11. Sign and Symptom of Pulmonary TB Patients

As shown in the above figure 3.11, the possible symptoms used to identify whether the patient has a chance to be Pulmonary TB or not is the lab test result that is shown in the figure 3.14 and then domain experts check risk factors is shown in figure 3.16 of a complaint's. Since the symptoms alone are not good enough to conclude.

Concept of Symptoms and Sign of COPD

In the diagnosis of COPD, the domain experts have a concept of symptoms that is used to differentiate the related symptoms of COPD and the non-related symptoms of COPD. For the related symptoms of COPD, the domain experts have a general knowledge about the common symptoms of the two types of COPD (emphysema and chronic bronchitis). In the process of knowledge gathering through interview, the domain experts (physicians) explained that there are symptoms that are used for diagnosing the new patient who came for treatment.

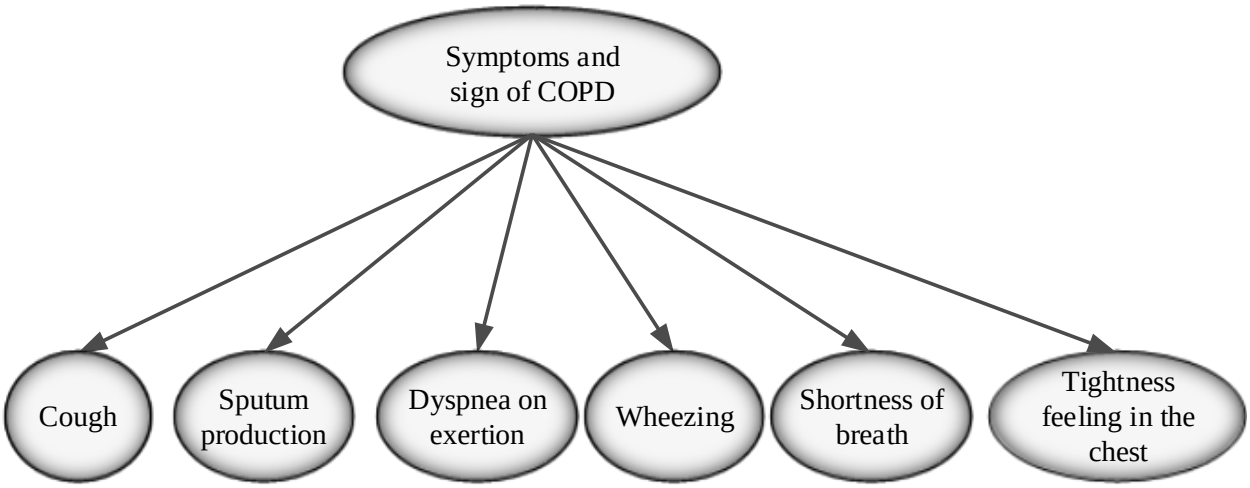

Figure 3-12. Sign and Symptoms of COPD Patients

As shown in the above figure 3.12, depicts the possible common symptoms and sign that can appear with COPD patients then domain experts check risk factors of a complaint's as shown in the figure 3.16.

3.5.1.4 Concepts of Laboratory Test Result

Since considering only the symptoms of chronic diseases are not enough to conclude that whether the patient has chronic diseases or not, laboratory test result is the key part and the most important concept for the decision.

1. **The concepts of fasting blood sugar (FBS) test of diabetic patients**

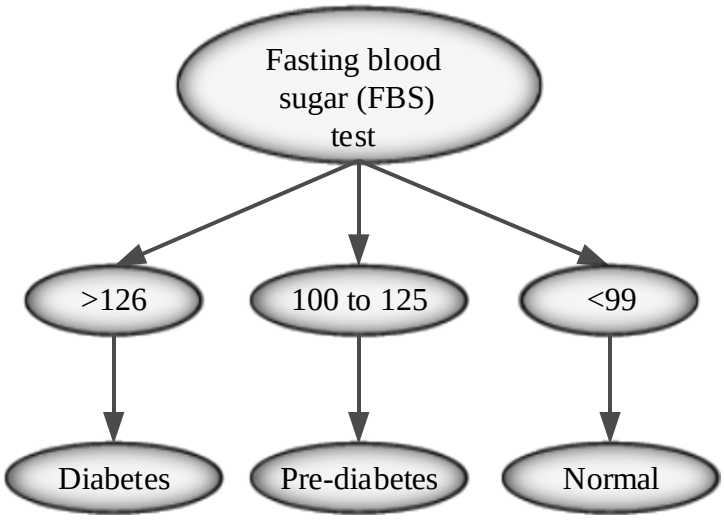

Figure 3-13. Concepts of FBS of a Compliant

As shown in the figures 3.13 the concepts of FBS of a diabetic complaint's helps the domain experts to know in which diabetic type the patient may fall. If FBS > 126 the patient is a diabetes, if FBS is between 100 and 125 the patient is categorized under pre-diabetes and if FBS<99 the patient is free from diabetes.

2. **Concepts of Tuberculin skin test (TST) of pulmonary TB complaint's**

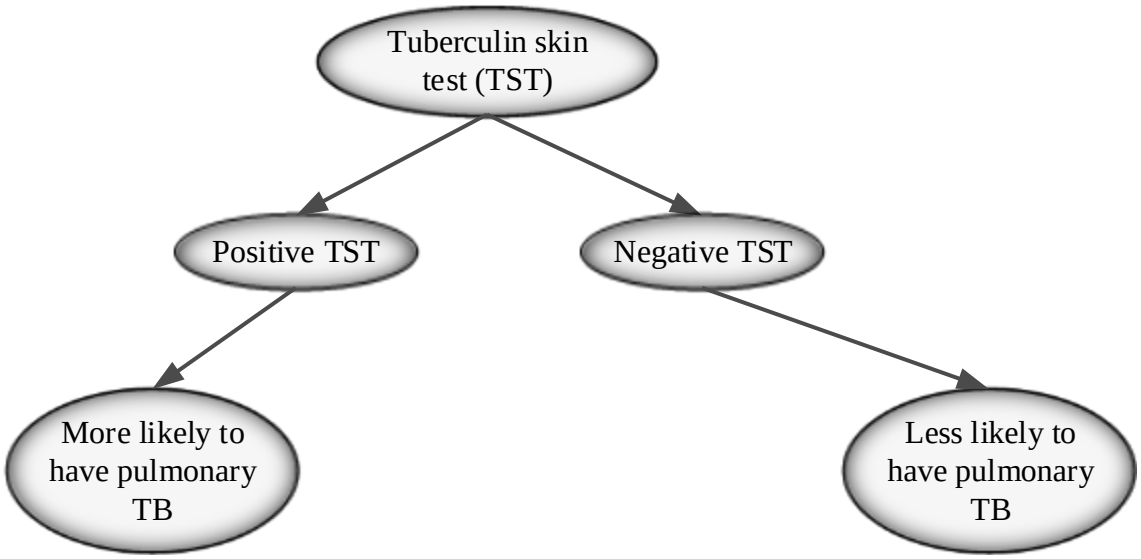

Figure 3.14. Concepts of TST of a Compliant

As shown in the figure 3.14, if a patient has a TST positive means the person‟s body was infected with TB bacteria so the patient is more likely to have pulmonary TB. Additional tests are needed to determine if the person has latent TB infection or TB disease. If a patient has a TST Negative means the person‟s body did not react to the test, and that latent TB infection or TB disease is not likely so the patient is less likely to have pulmonary TB.

3. **Blood Pressure of Chronic Diseases Complaint's**

Diagnosis of hypertension and Diabetes is mainly based on measurement of blood pressure on three separate occasions as show in Figure 3.15 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. If the range of SBP

140-159, DBP 90-99 the level of hypertension is mild, SBP 160-179, DBP 100-109 the level of hypertension is moderate and SBP >=180, DBP>=110 the level of hypertension is severe.

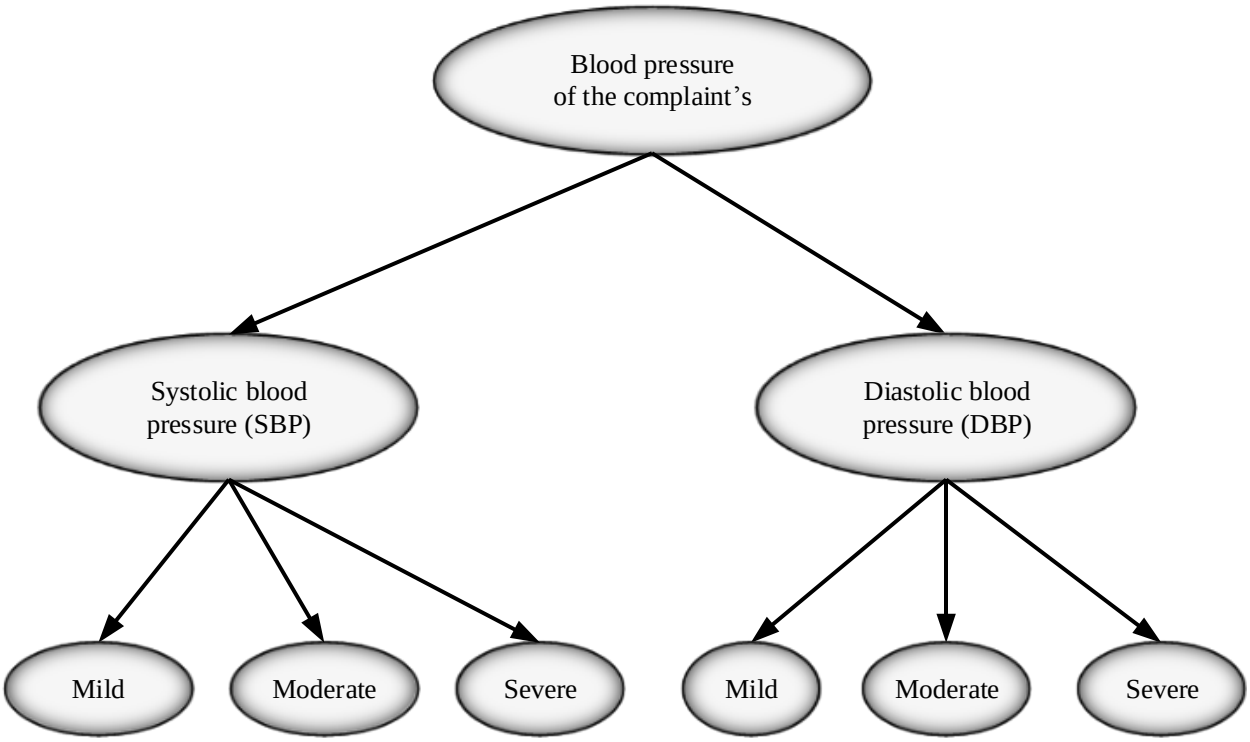

Figure 3-15. Concepts of Blood Pressure of a Compliant

Based on the above discussion the severity of the hypertension and diabetes mellitus will be classified as mild, moderate and severe.

3.5.1.5. Concepts risk Factors for Chronic Diseases complaints

The diagnosis of chronic diseases is not only based on the symptoms and lab results but also it considers different risk factors. These risk factors are modelled in figure 3.16

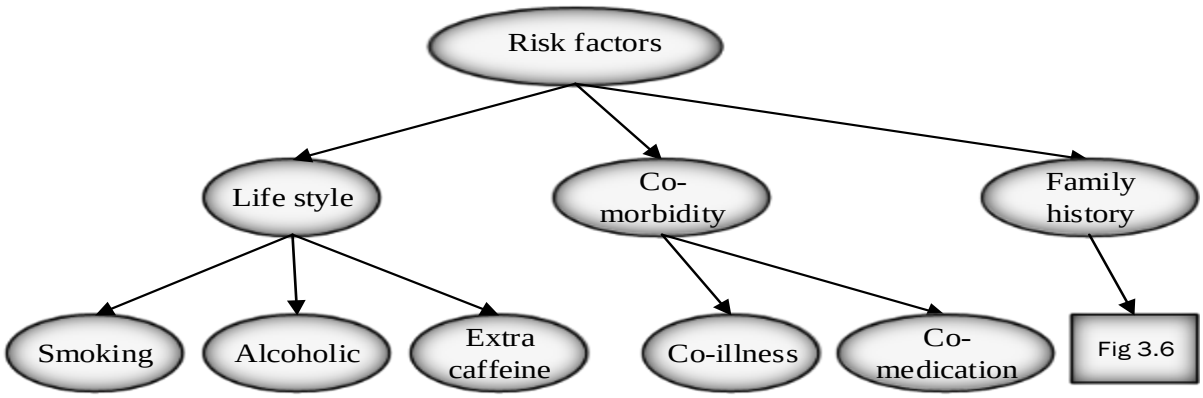

Figure 3-16. Risk Factors of Chronic Diseases

Life style, family history and co-morbidity are the main risk factors that cause chronic diseases. A person who smoke cigarettes or who is alcoholic or who take extra caffeine is highly prone to the disease. A patient who has co-illness will be treated in special way as compared to other compliant. Due to this clinician investigates co-illness and co-medication in chronic diseases diagnosis.

3.5.1.6. Clinician Decision for Chronic Diseases Diagnosis

The clinician decision on Chronic Diseases treatment is shown in figure 3.17 below.

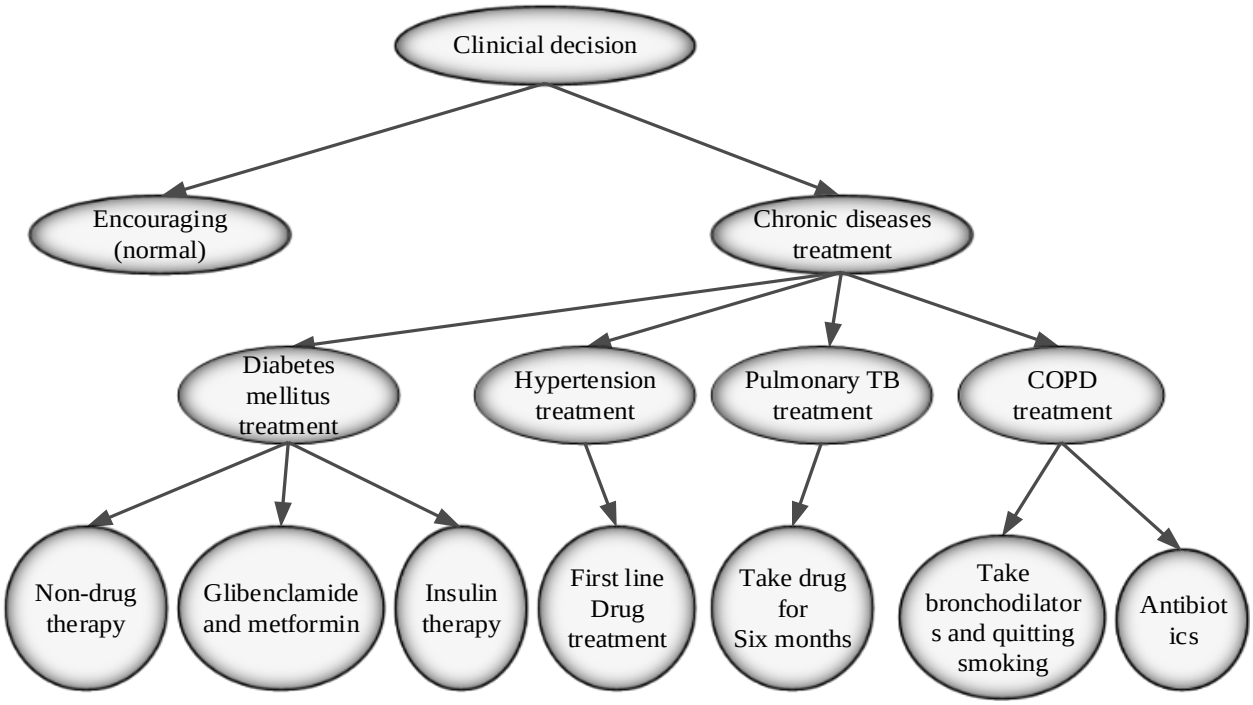

Figure 3.17. Concepts for Clinician Decision of Chronic Disease Cases

When the compliant has chronic diseases then the clinician critically selects how the chronic diseases should be treated. The main decision that the clinician make on chronic diseases case is that weather the disease should be treated with drug therapy, oral hypoglycaemic drug and life modification only by analysing the severity of chronic diseases, 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. As a result of this the clinicians select either the chronic diseases patient should be treated with life modification (advice) only or with drug therapy, bronchodilators and insulin therapy. The drug therapy can be first line drug, mostly Thiazide type diuretic, or urgency of drugs like Captopril, or emergency drug like Hydralazine, if pulmonary TB it takes drug for six, if COPD it takes bronchodilators and antibiotics.

3.5.2. Goal of Chronic Diseases Treatment

The primary focus of treatment is in achieving the FBS between 70 and 120mg/dl , interrupt tuberculosis transmission, cure the patient, relieve symptoms ,treat complications and target systolic blood pressure and it is depicted in the figure 3.18 below. Treatment to lower levels may be useful, particularly to prevent stroke, asthma and 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.

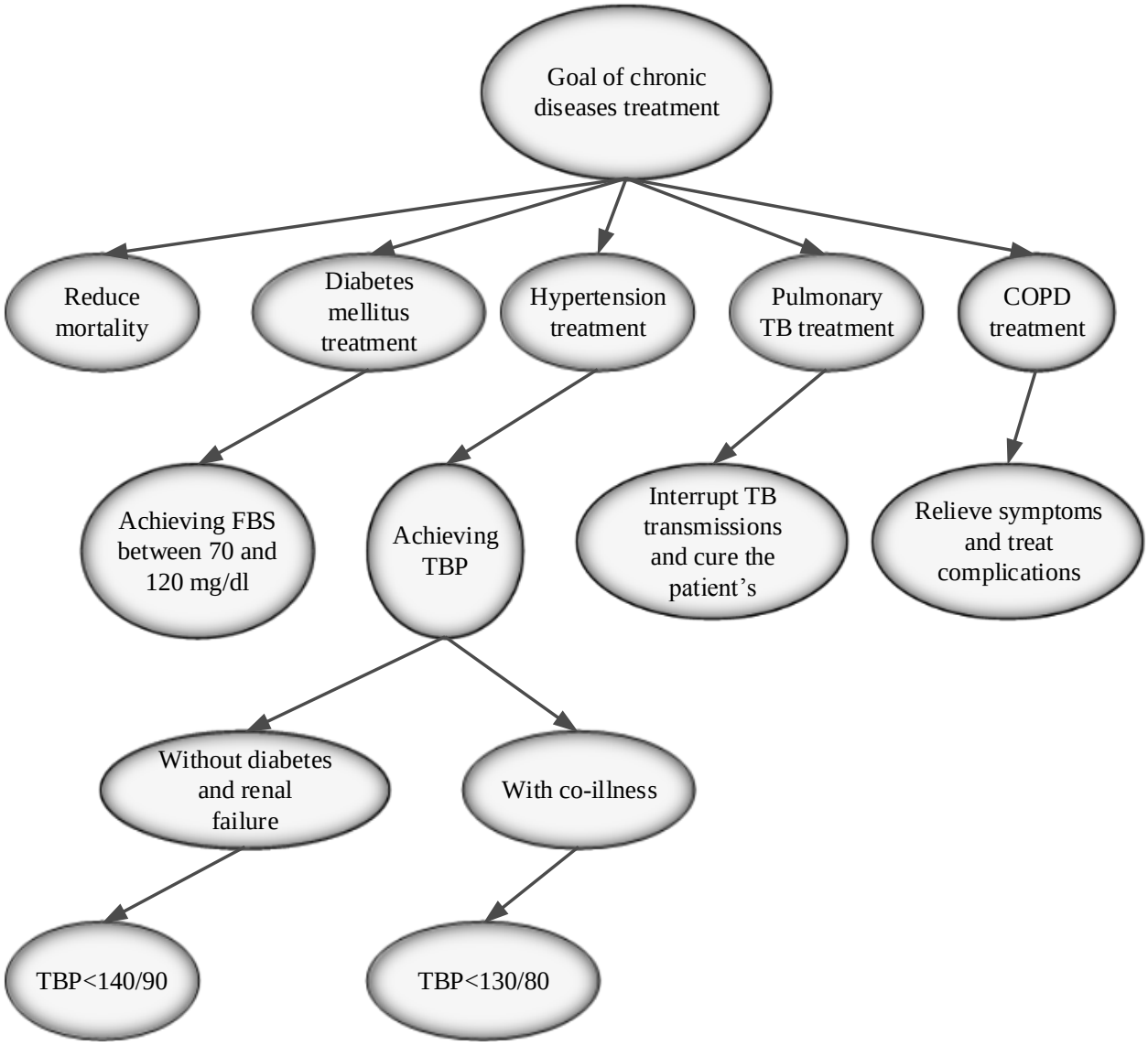

Figure 3-18. Goal of chronic disease treatment

3.6. Diagnosis of Chronic Diseases Case Structure

Diagnosis chronic diseases case structure has two important parts. The first one is the problem descriptions or situation and the second one is the solution.

Problem Description/Situation: It is the part of the case structure that is consisted of attributes which describes the problem to be solved.

Solution: This part of the case structure provides the recommended diagnosis and treatment given to patients based on the problem descriptions. Therefore, the researcher identified the

different description and solution attributes with the help of domain experts and Standard treatment guideline (STG) for Ethiopians health professional. But, there were different challenges during identification and representation of case structure. The first challenge is that the attributes (the symptoms in this case) used to diagnose chronic diseases are too many. Due to the time limitation and scarcity of resources, the researcher selects those symptoms that are most frequently recorded in the patient history. The other challenge is some attributes are difficult for representation. For example the duration of the symptom is difficult for representation because the duration of different symptoms is recorded in different minutes. Since it is so difficult to represent the duration of symptoms, it is not included in the case structure even though it is important in the decision making. Finally, after the case structure is constructed, the cases that build the case base are collected from the chronic patients‟ card history.

Table 3-7: Shows the Case Structure for Diagnosis of Chronic Diseases

Attribute	***Parameter of case***
Age	Is the age of the patient
Sex	Is the term used to refer to a person's self-representation as male or female
Disease Category	Shows the category of a chronic diseases
Present illness1	The symptoms(sign) of a complaint
Present illness2	The symptoms(sign) of a complaint
Premonitory Symptoms	Shows strong sign of a complaint
Tuberculin Skin Test(TST)	Is a test to check if a person has been infected with TB bacteria or not
Fasting Blood Sugar(FBS)	It shows diabetes blood sugar in order to identify the patients' blood sugar level
Ketone	It is used to identify whether a patient is type_1 or other diabetic type
Systolic blood pressure(SBP)	It helps clinician to identify hypertension and diabetes. It is the amount of pressure that blood exerts on vessels while the heart is beating.
Diastolic blood pressure(DBP)	It is the amount of pressure in blood vessels

	between heart beats (when the heart is resting)
Smoking and Drinking Status	Shows whether the complaint is smoker and drinker or not
Co-illness	Show whether the complaint has other kind of disease such as asthma, diabetics, renal failure or CHF.
Co-medication	It is the medication given to the complaint for his/her co-illness
Family History	Genetic background of a complaint
Diagnostic Solution	Solution
Recommended treatment	Solution

According to the domain expert the major decision that the clinician makes for chronic diseases 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 Standard Treatment Guide line for General Hospitals (STGLGH) and domain expert, the decisions that are made for chronic diseases cases are five as listed below:

- To check for other disease if the compliant is free from chronic diseases,
- 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

DESIGNING AND IMPLEMENTATION OF THE PROTOTYPE

4.1. Overview of a Knowledge Base and Expert System

The design and implementation part of this section cover the discussion on developing of case based reasoning system. It involves the actual development of a workable case based reasoning for chronic diseases diagnosis in the case of Adama Referral and Medical Hospital.

4.2. Design of the Case Based Reasoning System

The overall system design is shown in Figure 4.1. The knowledge engineer collected knowledge from various sources and this knowledge is processed and conceptual model was constructed. Then the knowledge engineer represented the knowledge in its appropriate form. The knowledge is represented on the case base as attribute-value pairs and is implemented on the case base.

A system user who seeks advice or recommendation, inputs a new query to the system. The next task is similarity computation. In this phase knowledge about the weights and similarity functions of the features is represented and shown to the user. In the prototype system these weights and similarity functions can also be changed by the user if needed, for instance for testing purposes.

Once all weights and functions are confirmed, the system loads successful cases of previous patients from the case base. Then it computes local similarity of the new case and successful cases feature by using the given similarity functions and weights. The global similarity of the new case, the average of the local similarities is computed with every case in the case base. The system uses the k-nearest algorithm, thus top k nearest cases will be retrieved. A case contains case description attributes and case solution attributes. The result of the voting is ranked list of proposed_solution for new query case. The proposed solution can be derived directly from a retrieved case that matches exactly or partially to the problem of the new case. But, using the proposed solutions directly may have a risk. Therefore, the user of the system should have made an adaptation by altering the differences between the proposed case and the new case.

Information about the difference between the proposed cases and new query are available in the database. In addition to adaptation, case inconsistencies are revised if the retrieved case is not the same as the new case. Finally, the revised solution is retained in the case base for future problem solving.

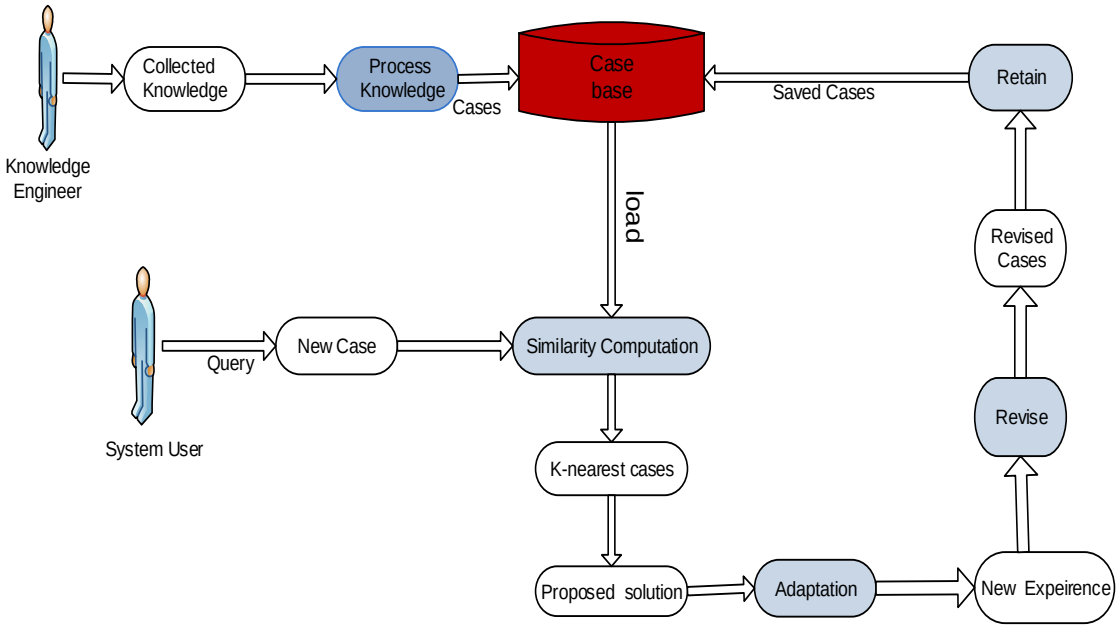

Figure 4-1: Architecture of the System [114]

4.3. Case Based Reasoning for Chronic Disease Diagnosis

In this study, the researcher uses the main feature of jCOLIBRI2 integrated in to Eclipse IDE and by adding some java classes and GUI to deliver the actual prototype of the ACBRCDD. Using JCOLIBRI CBR framework minimizes the effort to develop CBR application from scratch.

4.3.1. Managing Case Structure of Selected Attributes

- In discussion with clinician the cases in this research consists of 17(seventh) description attributes that served to easily classify the case which is used to make decision by the system and two solutions attributes which holds solution for the recommendations. The acquired cases are saved in MySQL database. Case attributes are Age, Sex, Disease category(DC),Present illness1(PI1),Present illness2(PI2),Premonitory symptoms(PS), Tuberculin skin test(TST),FBS, SBP, DBP, Ketone, Co-illness, Co-medication, Smoking and drinking status, Family history, Diagnostic solution and Recommended Treatment

are the important attributes chosen from all attributes. These attributes have significant impact on chronic diseases patient diagnosis. Most of the case attributes have String data type and a few attributes have Maxstring data type. Age, SBP, DBP and FBS has integer data type. The local similarities of some of the attributes are String data type is equal. Age, SBP, DBP and FBS has interval similarity function. Global similarity of solution attributes is average.

Local similarity functions are used to compare simple attribute values. There are many similarity functions in the jcolibri.method.retrieve.NNretrieval.similarity package. Each local similarity measure can only be applied to some data types. In this research, the following local similarity functions are used.

Equal: If you select equal local similarity for each attribute. Then your input and value of case base must be match. If value matches exactly then it will get result otherwise match failure.

Interval: When you select similarity interval and adjust interval value. Then, jCOLIBRI match value keeping in mind that interval. Exact value match is not compulsory in that type.

Global Similarity is linked with compound attributes and used to get similarity of collected attributes in unique similarity value. Global similarity used in this research is average similarity.

Average: It is a type of global similarity that considers the average of all attribute local similarity values. The algorithm works as follows [70]; [27]; [14].

Step 1: Find the local similarity of step for all attributes of the case which make up the case base.

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

Step 3: Add the value of all attribute results of step 2.

Step 4: Add all weights of attributes that represent the importance value of the attributes and multiply by the number of 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 with the new input case.

Description and Weight of Case Attributes

A single case for this research purpose had fifty description attributes which served to hold the descriptions of the problems and two solutions attributes which could hold the solutions for the problems after having best match cases to the problems at hand. Table 4.1 presents descriptions of case attributes.

Table 4-1: Attribute Description of the Case Structure

Most significance attributes			
Attribute name	***Data types***	***Weight***	***Local similarity***
Premonitory symptoms	String	0.9	Maxstring
SBP	Integer	0.9	Interval
DBP	Integer	0.9	Interval
TST	String	0.9	Equal
FBS	Integer	0.9	Interval
Significant attributes			
Present illness1	String	0.8	Maxstring
Present illness2	String	0.8	Maxstring
Family history	String	0.8	Equal
Ketone	String	0.7	Equal
Co-illness	String	0.6	Equal
Co-medication	String	0.6	Equal
Smoking and drinking status	String	0.6	Equal
Other attributes			
Age	Integer	0.5	Interval
Sex	String	0.5	Equal
Disease category	String	0.4	Maxstring

Solution Attributes			
Diagnostic solution	String	1.0	Average
Recommended treatment	String	1.0	Average

Table 4.1 above shows the description of case attributes regarding attribute name, data types, and weight, local and global similarity. Attributes which are considered to be most significance attributes to the problem domain like premonitory symptoms, SBP, DBP, TST and FBS each have the highest weight value of 0.9 respectively. On the other hand, significant attributes including family history, present illness1, present illness2, ketone, Co-illness, Co-medication and Smoking and drinking status have medium weight values 0.8, 0.8, 0.8, 0.7, 0.6, 0.6 and 0.6 respectively as compared to other attributes like age, sex and disease category of the patient. The assignment of weights to attributes indicates that how often physician use those attributes for diagnosis and recommendations to treat patients with regard to the diagnosis of chronic diseases case patients. In local similarity, four attributes have interval local similarity while the rest of them have equal local similarity. The attributes age, SBP, DBP and FBP which are integer type, the attribute premonitory symptoms, present illness1, present illness2 and disease category which are maxstring types and the rest of them defined as string. The algorithm and the sample code in jCOLIBRI for interval similarity are shown below.

Algorithm for similarity of interval [27]

Step1: Compute the difference of the two given value of attributes and place it in absolute value.

Step2: Compute the difference of the maximum and minimum values of the given attribute from the case base.

Step3: Divide the result of the 1st step by the result of the 2nd step.

Step4: 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.

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

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, TST, Family history, Ketone, Co-illness, Co-medication, Smoking and Drinking Status 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 [27].

Local similarity Algorithm for equal function [27]

If the two value are equal

return 1

Else

return 0

4.3.2. Building the Case Base

To build the case base this study uses a database used to store the cases which is a MySQL database and this database component has two roles in the prototype system. One is the case base which is implemented in a table form, and the second role is it elaborate to the case based reasoning system by providing required facts.

Case base contains the case features of the CBR system, organized in 17 fields Which are Case id, Age, Sex, Disease category(DC),Present illness1(PI1),Present illness2(PI2),Premonitory symptoms(PS), Tuberculin skin test(TST),FBS,SBP,DBP, Ketone, Co-illness, Co-medication, Smoking and drinking status and Family history are case description attributes and Diagnostic solution and Recommended treatment are the case solution in the solution attribute.

4.4. Implementing the System Using Jcolibri2

To develop a new CBR application with jCOLIBRI2, Eclipse IDE was used. A new eclipse project was created and jcolibri2.jar files (and related libraries) were imported into the new project. A CBR application that uses jCOLIBRI must implement or extend the jcolibri.cbrapplications.StandardCBRApplication interface.

This interface divides the CBR application behaviour into 3 steps:

- **Pre-cycle**: Initializes the CBR application, usually loading the case base and precompiling expensive algorithms (really useful when working with texts). It is executed only once.
- **Cycle**: Executes the CBR cycle. It is executed many times.
- **Post-cycle**: Post-execution or maintenance code.

For the prototype system, a class which implements the above interface was created and fully implemented.

```
//Bean that stores the Description of the case
public class RECaseDescription implements jcolibri.cbrcore.CaseComponent {

        String Case Id;
        Integer Age;
        String Sex;
        String Disease_Category;
        String Present_Illness1;
        String Present_Illness2;
        String Premonitory_Symptom;
        Integer Systolic_Blood_Pressure;
        Integer Diastolic_Blood_Pressure;
        Integer Fasting_Blood_Sugar;
        String Tuberculin_Skin_Test;
        String Ketone;
        String Co_illness;
```

String Co_medication;

String Smoking_and_Drinking;

String Family_history;

Figure 4-2: Case Description Bean

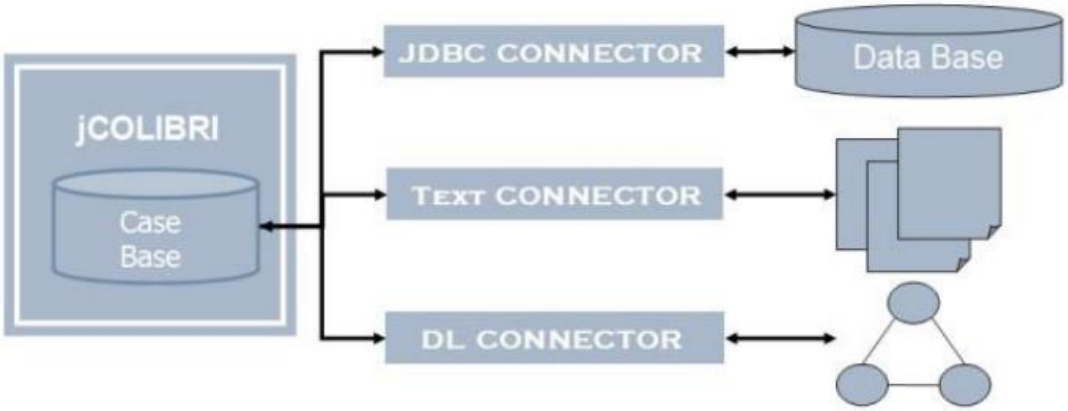

Figure 4-3: Case Base Management in jColibri2

Persistence is built around connectors. A connector represents the first layer of jCOLIBRI on top of the physical storage. Connectors are objects that know how to access and retrieve cases from the medium and return those cases to the CBR system in a uniform way. The second layer of Case Base management is the data structure used to organize the cases once loaded into memory. The organization of the case base (linear, k-d trees, case retrieval nets, etc.) may have a big influence on the CBR processes, so the framework leaves open the election of the data structure that is going to be used. JCOLIBRI2 includes the following Case Bases:

- jcolibri.casebase.LinealCaseBase: Basic Lineal Case Base that stores cases into a List.
- jcolibri.casebase.CachedLinealCaseBase: Cached case base that only persists cases when closing the application.
- jcolibri.casebase.IDIndexedLinealCaseBase: Extension of LinealCaseBase that also keeps an index of cases using their IDs.

In the prototype system the database connector, which is partly implemented with Hibernate (Object-Relational Mapping tool), was used and lineal case base is selected for in memory organization of the cases. To connect and load the case base, the **Configure** and **Pre-cycle** functions of the CBR application were coded.

As it can be seen in the main function of the CBR application below, after configure () and pre-Cycle () the next step is loading a query dialog.

```
public static void main(String[] args) {

    Begin recommender = getInstance();
    recommender.showMainFrame();
    try {
        recommender.configure();
        recommender.preCycle();
        qf = new QueryDialog(main);
        boolean cont = true;
        while (cont) {
            qf.setVisible(true);
            CBRQuery query = qf.getQuery();

            //System.out.println("Query: " + query);
            recommender.cycle(query);
            int ans = javax.swing.JOptionPane.showConfirmDialog(null,
                    "CBR cycle finished, query again?", "Cycle finished",
                    javax.swing.JOptionPane.YES_NO_OPTION);
            cont = (ans == javax.swing.JOptionPane.YES_OPTION);
        }
        recommender.postCycle();
    } catch (Exception e) {
        org.apache.commons.logging.LogFactory.getLog(Begin.class)
                .error(e);
        javax.swing.JOptionPane.showMessageDialog(null,
                "error!! " + e.getMessage());
        e.printStackTrace();

        System.out.println(e.getMessage());
    }
    // System.exit(0);
}
```

Figure 4-4: The Main Function of the System

4.4.1. The Query Dialog

A query dialog with sample input is shown in the figure below.

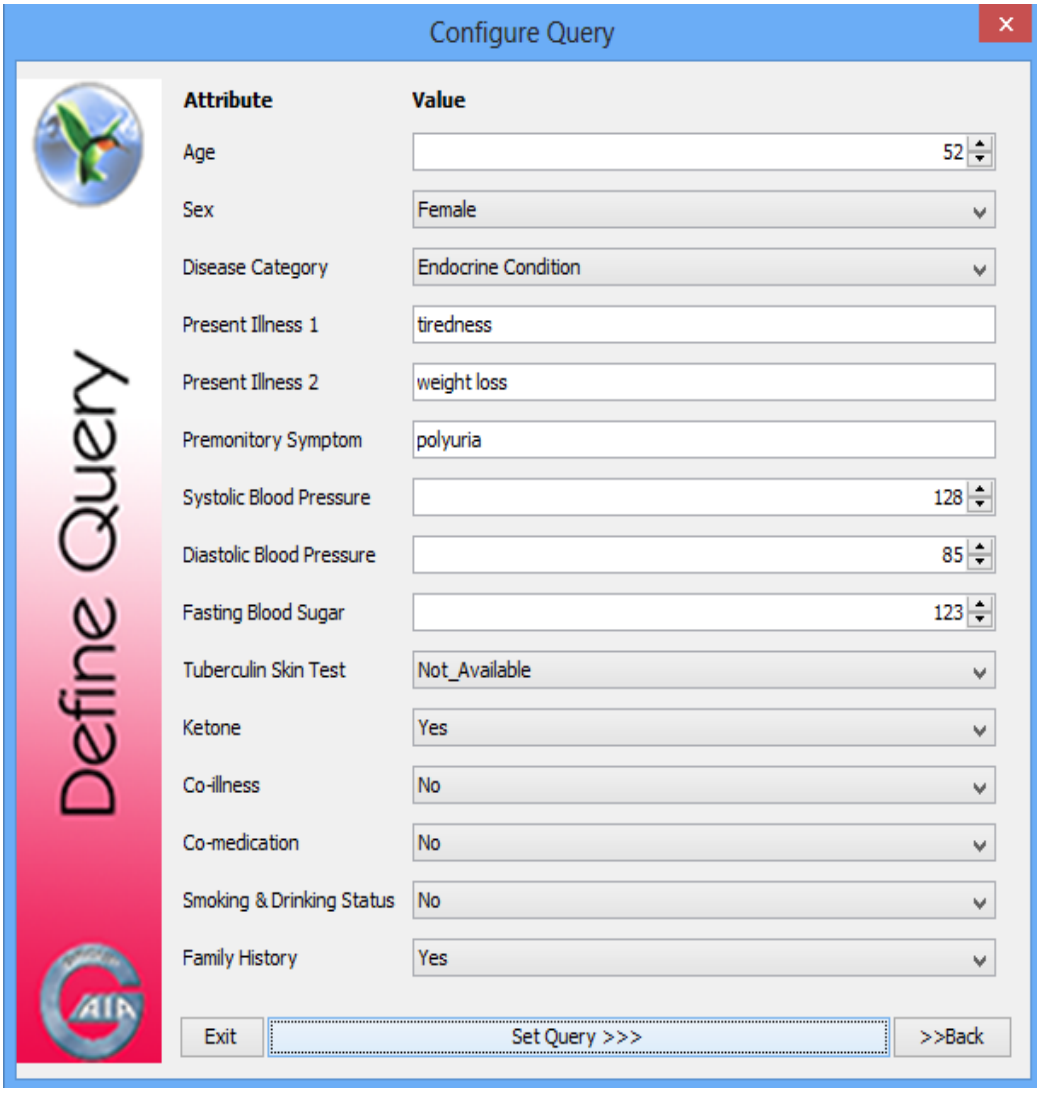

Figure 4-5: Query Dialog

After the query dialog accepts the inputs, the system retrieves similar case from the case base that match the query case and keeps the result in memory for later use (During Adaptation). In the main function of the CBR application after executing configure () and precycle () methods, the cycle () method is executed. The cycle () method performs the 4R's tasks (retrieval, reuse, revise, and retain).

4.4.2. The Similarity Dialog

The default weights and similarity function of the case attributes is set according to the recommended values. The similarity dialog is shown in figure below

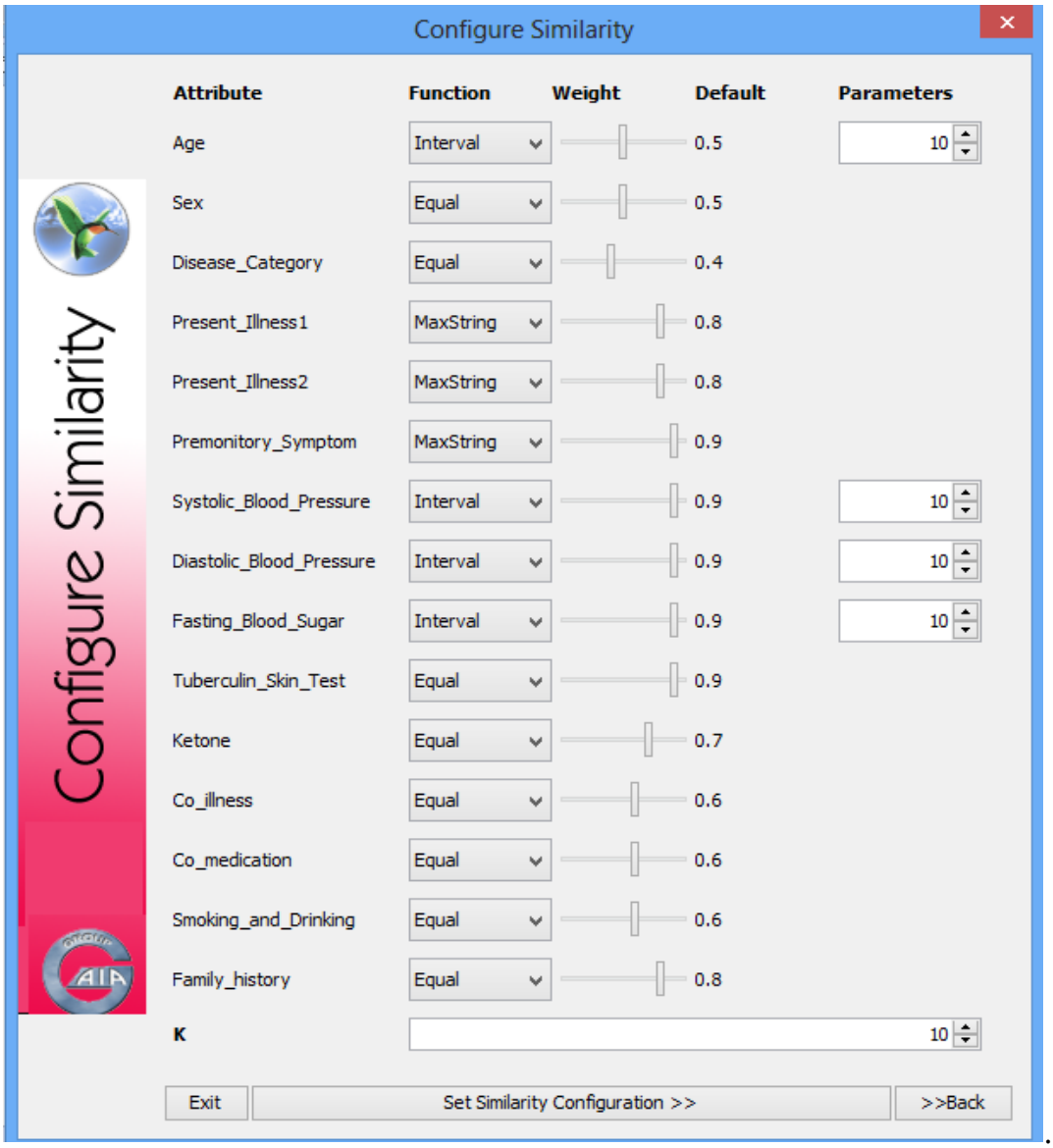

Figure 4-6: Similarity Dialog

The similarity dialog returns the weight, local similarity functions and k value. Therefore using the classes and methods of JColibri the similarity of the query case with cases in the case base is calculated. The next step is loading the result dialog.

4.4.3. Result Dialog

The result dialog shows the top similar cases as presented in the figure below. It also includes the evaluation of cases in a range between 0 and 1.

.

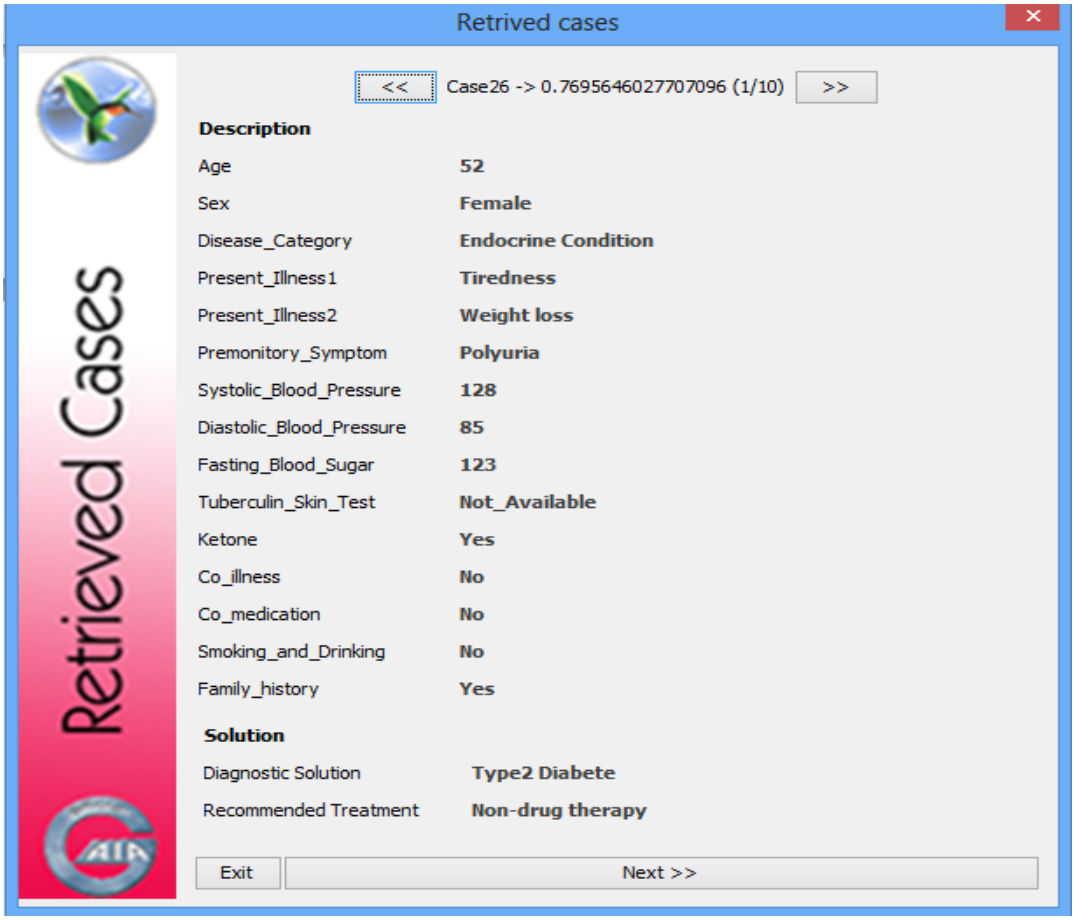

Figure 4-7: Result Dialog

Retrieve tasks used to retrieve case(s) from the stored case base. Retrieve tasks also decomposed in to different subtasks. The subtasks include select working cases task, compute similarity task and select the best case. Select working case task selects cases from case base and stores them into current context. Compute similarity task compute similarity of the stored cases with the case entered by the user using the query window. Select best case shows the best matched of case(s) after computing the similarity of stored cases against the new case. It means that the number of best matched case(s) is shown to the user.

4.4.4. The Revise Dialog

In the revise step the proposed solution is tested for success, e.g. by being applied to the real world environment or evaluated by a domain expert, and repaired if failed. In the prototype system the revise dialog allows a domain expert change the new cases as shown in Figure 4.8 below.

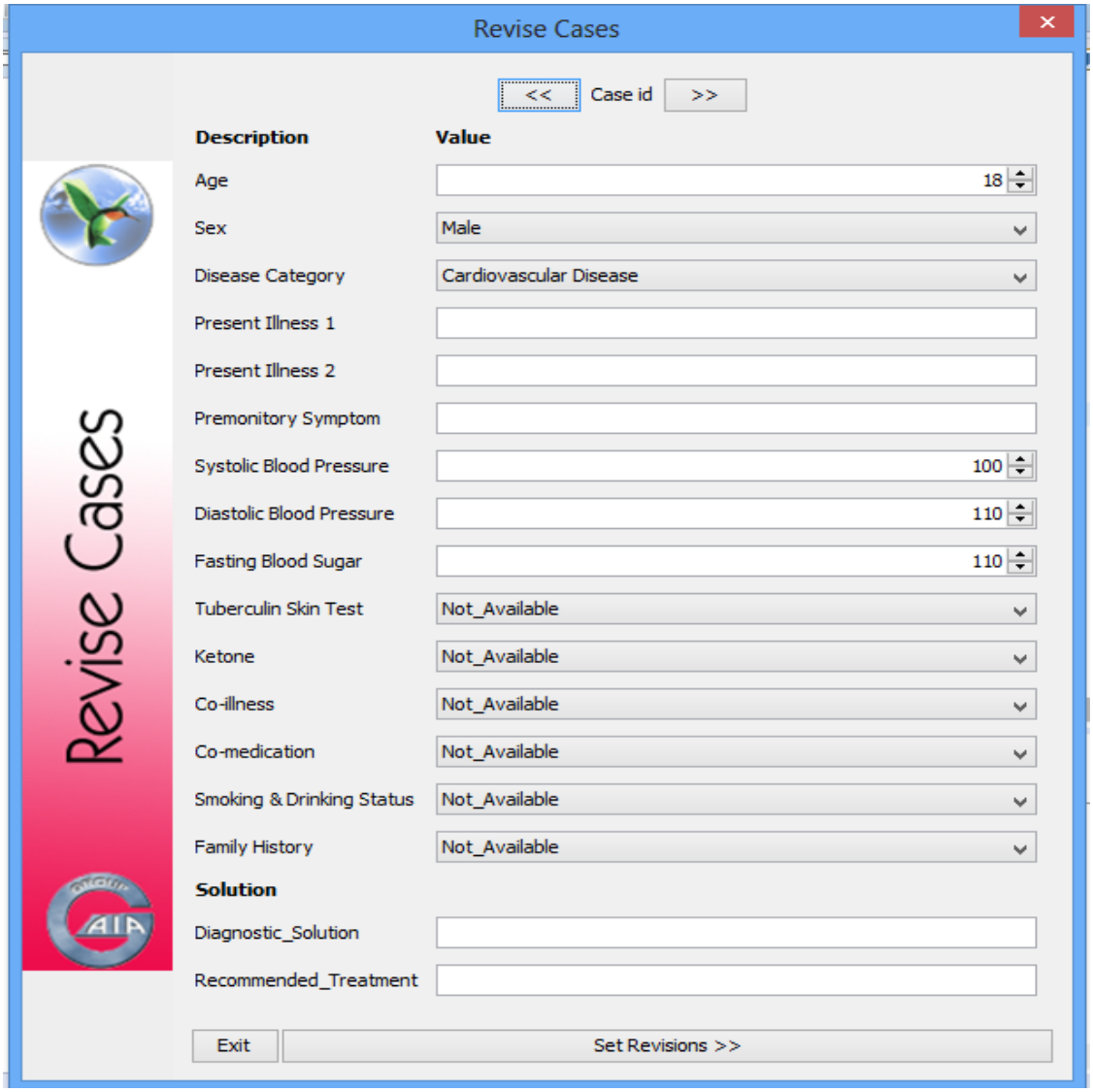

Figure 4-8: The Revise Dialog

4.4.5. Retain Dialog

In the retain step useful new cases are stored in the case base for future reuse. This way the CBR system has learned a new experience. The prototype system allows the domain expert to choose the case he wants to retain and the selected case(s) would be stored permanently to the case base. The retain dialog is shown in Figure 4.9 below.

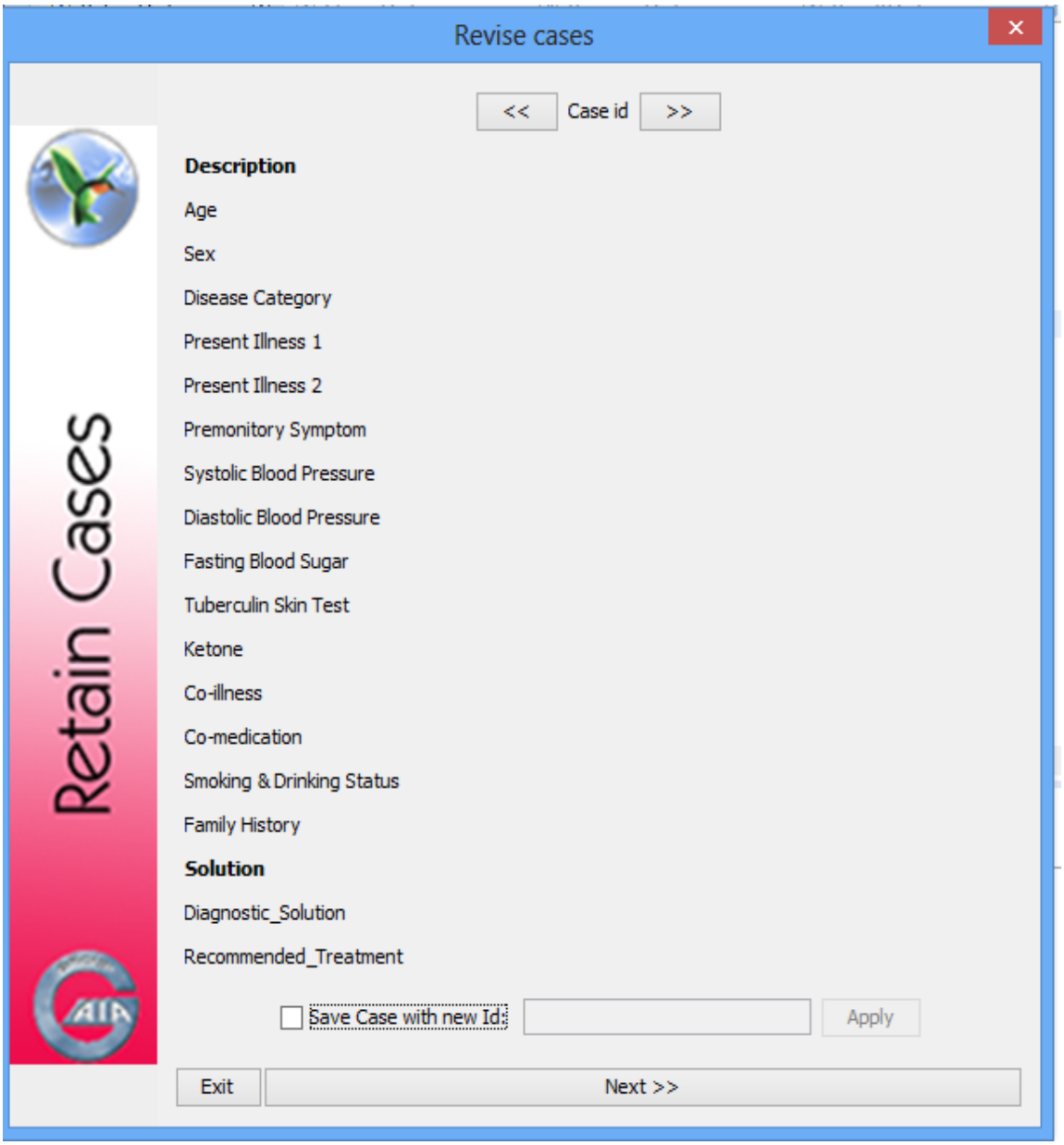

Figure 4-9: Retain Dialog

CHAPTER FIVE

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 match with cases from the case base and evaluation is conducted to investigate the applicability of the prototype.

5.1. Experimental Setting

For this research a total of 164 cases are used to build the case base and to test the prototype with 25 hypertension cases, 53 diabetes mellitus cases, 46 pulmonary TB cases, 40 COPD cases are collected. The evaluation part is basically focused on the performance of the prototype in terms of precision, recall and Accuracy.

5.2. Testing the Main Cycles and Evaluating the Performance of the System

In this section, the functionality of the prototype system is tested and verified whether it works correctly or not. The retrieval results were also validated by involving domain experts.

5.2.1. Testing Configure and Pre-cycle

As discussed in Chapter 4, in the case based reasoning system, the first two functions executed are configure and pre-cycle.

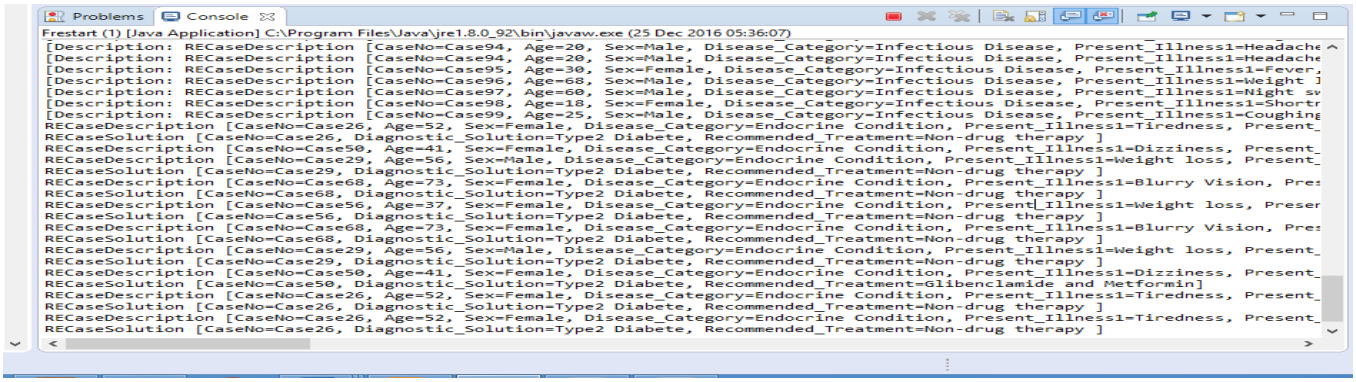

Figure 5-1: Console Output After Configure and Pre-cycle

5.3. User Acceptance Testing

User acceptance testing is performed in a real situation at Adama Referral and Medical Hospital specifically with Hypertension, Diabetes mellitus, Pulmonary TB, COPD case team members. During testing the users‟ acceptance, the applicability of the prototype is evaluated by potential users of the system. The selection is based on the academic qualification in the domain area, work experience, willingness and participation in this research. The chronic diseases case team members are participated consist of four doctors, four nurses and in addition to case team two health officers also participated. Four doctors those who participated in this research work from the beginning to the end by providing the necessary expertise knowledge evaluated the prototype and finally, they give their feedbacks on the questionnaires. The two health officers and four nurses also evaluating the system by appraising the applicability of the system in providing decision support in chronic disease diagnosis.

The system performance evaluation is carry out by using questionnaires. The questionnaire includes both closed and open ended questions. Using these questions system evaluators directly interact with CBR system. Totally fourteen evaluation questions are prepared for system evaluators. The questions are divided into two parts, the first seven questions are close ended questions which helps system evaluators to check on the user interface design aspects, easiness of the system to use, attractiveness, correctness of the decision, adequacy of knowledge content, the problem solving ability and significance of CBR system in chronic diseases diagnosis service. On the other hand, the remaining seven questions are open ended questions which used to collect constructive feedback from the system evaluator’s based on the system’s conclusion.

Based on selected close ended questionnaires, the domain experts put the value of their satisfactory level as Excellent, Very Good, Good, Fair and Poor of evaluation criteria. For their satisfactory level, the researcher puts the value as; 1 for Poor, 2 for fair, 3 for Good, 4 for Very Good and 5 for excellent values. After user acceptance evaluation was done by domain expert, the next step was done by knowledge Engineer about the analysis results that are given by domain experts about visual interaction of the proposed system and finally, give its description about the analysis results of proposed system prototype user acceptance evaluation.

Table 5-1: Depicts User Acceptance Testing [97]

	Evaluation Parameters	***Performance Value***						***Total AV/perf. %***
		1	***2***	***3***	***4***	***5***	***Average***	
1	Adequacy and clarity of decision support			2	4	4	4.2	84
2	Relevancy of the retrieved cases in the decision making			1	4	5	4.4	88
3	Fitness of the final solution to the problem at hand				4	6	4.6	92
4	Ease of use		1	2	3	4	4.0	80
5	Relevance of the attributes in representing the chronic disease case			1	5	4	4.3	86
6	Is the system efficient in time and memory				6	4	4.4	88
7	Is the system has adequate resource		1	3	2	4	3.9	78
8	Is the user interface interactive		1	2	3	4	4.0	80
9	Rate the significance of the system in the domain area				3	7	4.7	94
		Total Average					***4.28***	***85.56***

Note: **AVG** = average, **Perf**. = performance

As depicted in table 5.4, 20% of the respondents‟ rate adequacy and decision support as good, 40 % as very good and the remaining 40% of the respondents rate as excellent. Similarly, relevance of the retrieved cases in the decision making rated good by 10% and 40% as very good of the respondents whereas the remaining 40% of the respondents rate it as excellent. In the case of fitness of the final retrieved solution to the new problem at hand around 40% of the respondents rate the prototype is very good whereas 60% of the respondents rate as excellent. 10% of the respondents‟ rate ease of use of the system is fair, 20% as good, 30% as very good and the remaining 40% of the respondents rate it as excellent.10% of the respondents rate the relevance of attributes in representing chronic diseases case rate as good whereas 50% rate as very good and the remaining 40% of the respondents‟ rate as excellent. The efficiency of the system is evaluated to be very good by 40% of the respondents while the remaining 60% rated it as excellent. Around 10% of the respondents‟ rates the available resources are fair and in the same way 30%, 20% and 40% of the respondents‟ rate the resources available are good, very good and excellent respectively. The user interface of the prototype is also rated as fair, good, very good and excellent by 10%, 20%, 30%, and 40% of the domain experts respectively. To this end 70% of the respondents rate the applicability of the prototype in their domain expert is excellent and the remaining 30% of the respondents very good. Finally, based on the evaluation of all the respondents the average performance of the prototype is 85.56%, which is above very good. This performance result shows the prototype has a promising applicability in diagnosing chronic disease diagnosis.

In general, the significance of CBR in chronic diseases diagnosis has achieved an encouraging result. In addition, the significance of the system in the domain area rated by the respondents with a highest rating value. Domain experts assign a lower value compared to others is the resources available in the case base. Since the prototype learns over time, the resources will grow quickly.

One interesting feedback provided by the domain experts is adding explanation facilities are in the system is very important. According to the domain experts, explanation facility is important in the adaptation of the retrieved solution. If users of the system get a more explanation about the retrieved solution and the problems itself, they can easily decide whether to use or not use the

retrieved solution. According to the experts if the recommended treatment is a pharmacotherapy, the user should get more explanation about the medication recommended such as adverse effect and instruction of the dose.

5.4. System performance Using Test Cases

As the goal of retrieve process in this research is to recommend correctly for chronic diseases cases, i.e. to solve the problem correctly, the performance of the retrieve process is measured by using accuracy. Accuracy is one of the useful measurements in case based reasoning [115]. This measurement had been used by [27] on their research. Accuracy is defined as the percentage of the number of correctly recommended cases [116]. System performance evaluation was done using sample case testing mechanism in order to check the accuracy and correctness of the system prototype performance. System performance testing is the process of determining whether the prototype system is correct, that is whether it meets the level of accuracy as required. Since the research uses top 5 neighbour's votes to validate testing proportion, using sample tests experiments are conducted to evaluate the performance of retrieve process of the case based reasoning system prototype for chronic diseases diagnosis.

The setting of a parameter K significantly impacts on the recommendation of accuracy. Since K is the number of voting neighbours and the KNN method uses a majority vote scheme, value of a samples top 5 and top 10 were studied. The similarity method have to be identified for this purpose the k-nearest algorithm is used so that similar cases with top k example if K is 5 the top 5 Ks and if k is 10 the top 10 Ks in the case base are retrieved by the CBR system, a solution for a query, instead of propose just the solution of nearest neighbour or the k-nearest neighbours, a method in which the k-nearest neighbours vote was implemented. Although it is often assumed that better accuracy values can be obtained if the number of neighbours is increased, this is not the case. As shown in the sample tests the maximum accuracy was achieved when the value of K was equal to 5 rather than 10. Thus, when K was more than 5, the accuracy decreased as K increased. Therefore, the value of K was set at 5 to test the performance of the system.

To confirm whether the right prototype system has been built. The correct and incorrect outcomes are identified by comparing decisions made by domain experts on the sample chronic disorder case and with the conclusions of the system prototype. This evaluation strategy is used

to validate the achievement level of the system and to measure the accuracy of the system. To address the issue of validation, test cases are selected that have similar parameter with the system prototype they are selected purposely from Adama Referral and Medical Hospital. These test cases are categorized in to four types, based on knowledge structured for the chronic diseases categories of the proposed system. The types of diseases that have been tested are Diabetes mellitus, Hypertension, Pulmonary TB, COPD cases. The system evaluator selects cases from any categories of tested cases. After the retrieve task of the system is tested for each case, summery of the result is shown in the table 5.2 below:

Table 5-2: ACBRCDD Accuracy Testing

No.	***Chronic Diseases Name***	***Total Number of Cases***	***Correctly Diagnosed Disease***	***Incorrectly Diagnosed Disease***	***Accuracy of the Prototypes (100%)***
1	Diabetes Mellitus	53	41	12	86.8%
2	Hypertension	25	20	5	88%
3	Pulmonary TB	46	35	11	87%
4	COPD	40	32	8	90%
Total		***164***	***128***	***36***	***87.8%***

From the data shown in the above table, of the total of 164 patients, 128 were diagnosed correctly as chronic diseases cases, and the other 36 were diagnosed incorrectly as chronic diseases cases. During the testing procedure, the expert evaluators identify correctly and incorrectly diagnosed cases by comparing the decisions made by the system with that of the experts' decision on those cases. This testing method is applied to evaluate the performance of the prototype system using the parameters precision, recall and F-measure. These three parameters are used in order to measure the accuracy of the prototype system. The following table shows confusion matrix output for correct diagnosed and incorrect diagnosed chronic diseases. 146 Positive cases are correctly diagnosed and 18 false negative cases are incorrectly diagnosed.

5.4.1 Confusion Matrix

The confusion matrix is a simple square matrix that compares the relative performance of the human and the new system – the match and mismatch. The columns of the matrix correspond to the number of instances by human and the rows correspond to the number of instances by the system as a particular value (Brule, 1986). One benefit of a confusion matrix is that it is easy to see if the system is confusing two classes (i.e. commonly mislabeling one as another). For example, the following table shows the confusion matrix for a two class classifier. The entries in the confusion matrix have the following meaning in the context of our study: *a* is the number of correct predictions that an instance is negative, *b* is the number of incorrect predictions that an instance is positive, *c* is the number of incorrect of predictions that an instance negative, and *d* is the number of correct predictions that an instance is positive.

Table 5.3. Confusion Matrix

		Predicted	
		Positive	Negative
Actual	Positive	a (True Positive)	b (False Positive)
	Negative	c (False Negative)	d (True Negative)

Several standard terms have been defined for the 2 class matrix:

- The accuracy (AC) is the proportion of the total number of predictions that were correct. It is determined using the equation:

  AC=a+d/a+b+c+d

- The recall or true positive rate (TP) is the proportion of positive cases that were correctly identified, as calculated using the equation:

  TP=a/a+b

- The false positive rate (FP) is the proportion of negatives cases that were incorrectly classified as positive, as calculated using the equation:

  FP=b/a+b

- The true negative rate (TN) is defined as the proportion of negatives cases that were classified correctly, as calculated using the equation:

  TN=d/c+d

- The false negative rate (FN) is the proportion of positives cases that were incorrectly classified as negative, as calculated using the equation:

  FN=c/c+d

The classification performance of the model used in this research was evaluated using three statistical measures; classification accuracy, sensitivity and specificity. These measures are defined using the values of true positive (TP), true negative (TN), false positive (FP) and false negative (FN) as depicted in table 5.3.

- A true positive decision occurs when the positive prediction of the system coincided with a positive prediction of the medical expert or when an input is detected as a patient with ARTI diagnosed by the expert clinicians.
- A true negative decision occurs when both the system and the medical expert suggested the absence of a positive prediction or in other word, when an input is detected as normal that is labeled as a healthy person by the expert clinicians.
- False positive occurs when an input is detected as a patient that is labeled as healthy by the expert clinicians.
- Finally, false negative occurs when the system labels a positive case as negative or when an input is detected as normal with ARTI case diagnosed by the expert clinicians.
- Classification accuracy is defined as the ratio of the number of correctly classified cases and is equal to the sum of TP and TN divided by the total number of cases N.

  Accuracy=TP+TN/N

- Sensitivity refers to the rate of correctly classified positive and is equal to TP divided by the sum of TP and FN. Sensitivity may be referred as a *True Positive Rate*

  Sensitivity=TP/TP+FN

- Specificity refers to the rate of correctly classified negative and is equal to the ratio of TN to the sum of TN and FP. *False Positive Rate* equals (100-specificity).

  Specificity=TN/TN+FP

Table 5-4: Confusion Matrix for ACBRCDD

Chronic Diseases			
ACBRCDD		Positive	Negative
	Positive	TP=128	FP=18
	Negative	FN=2	TN=16

1. Sensitivity (True positive rate) = TP/TP+FN

   =128/128+2

   =128/130 which is equivalent to 100%

   =0.977(97.7%)

2. Specificity (False Positive Rate) = TN/FP+TN

   =16/18+16

   =16/34

   =0.4705(47.05%)

3. Accuracy= TP+TN/N

   =128+16/164

   =144/164

   =0.878(87.8%)

Sensitivity (true positive rate) of 97.7% indicates that the positive prediction of the system coincided with a positive prediction of the medical expert. This means that the system is almost fully functioning to identify a patient with any of the four chronic diseases and to give the necessary diagnosis.

A specificity of 47.05% indicates that, at initial diagnosis, ACBRCDD correctly identifies 47.05% of those who do not have chronic diseases. It is also to mean that out of 100%, the system recognizes around (47.05%) of healthy people as healthy. This implicates that further investigations shall be undertaken to ensure high chronic diseases confirming rate and to recognize all health people as healthy.

The result shows the system has an accuracy of 87.8%. The remaining 12.2% is the variation between the system and the human experts. Out of the 12.2% the output of the human expert is found to be 'chronic bronchitis' whereas the output using the system contains outputs, 'pulmonary TB'. The variation happened because of two main reasons. The first one is related with the symptom similarity between chronic bronchitis and pulmonary TB. Most of the time one of a severe complication of chronic bronchitis is pulmonary TB. If a patient has pulmonary TB previously it will be the cause for chronic bronchitis. If there is coughing or hack on the patient with pulmonary TB that may not go away for a long time, after one or two week, the disease became severe and it gave rise for chronic bronchitis diseases. The other one is the patient those who have hypertension may have diabetes mellitus and those who have diabetes mellitus may have hypertension the system is not take this case into consideration that is way the variation will happened between human experts.

After the retrieved task of the system is tested for the 164 cases, summary of the result is shown in the table 5.6 below

Table 5.5 Accuracy Value of the Retrieve Process

Total number of tested cases	***Total number of correctly recommended cases***	***Total number of incorrectly recommended cases***	***Accuracy of the prototypes (100%)***
164	128	36	87.8%

5.4.2 Comparison of the Prototype Performance with Previous Case Based Reasoning System

The performance of the diagnosis of chronic diseases 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 system is not evaluated. Thus, the retrieval performance of the systems are compared and shown in the table 5.6 below.

Table 5.6 Comparison of the Prototype with the Previous CBR System

Domain area	***Programming tools Used***	***Accuracy***
AIDS	jCOLIBRI	Not evaluated
Hypertension	Python	88.89%
Anxiety disorder diagnosis	jCOLIBRI	Not evaluated

The above table 5.7 shows that the result of the accuracy value of the case based reasoning system of chronic diseases is nearly the same with the accuracy value of system developed for hypertension by [14]. The reduction in the accuracy value compared with [14] could be caused by the composition of the attribute for the case structure of chronic diseases case. As it was indicated in chapter three of this thesis, the patient record for the chronic diseases patient does not include attributes like height, weight and race which affect the diagnosis of chronic diseases. 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 accuracy value of the diagnosis of chronic diseases could be the tools that are used in the research. In study we use JCOLIBRI that retrieve nearest cases for a given query but the tool that Henok use is python that retrieve nearest cases for a given query and also python use voting automatically for the proposed solution from the retrieved alternatives relevant cases. In addition to this, as the number of relevant cases for a query 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 chronic diseases cases. This deals with the accuracy. And more than an average accuracy is registered which is a promising result. This task is not evaluated in the [96] and [97] thesis research.

5.5. Results and 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). Additionally, the prototype provides a high percentage of the relevant cases for query which enable the user to diagnose and treat the new chronic diseases case. Therefore, retrieval performance of the case based reasoning system for chronic diseases diagnosis registers 87.8% accuracy. This is an encouraging result as the prototype system has given a high percentage of appropriate recommendations for the new chronic diseases cases. This shows that a promising result is achieved in the study. The reasons that the prototype couldn‘t achieve 100% retrieval 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 chronic diseases diagnosis 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 Neighbour 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 chronic diseases 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 chronic diseases diagnosis. This could affect the result of the retrieval performance of the prototype. The performance of the retrieval 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 chronic diseases cases to the case base can also improve the performance of the prototype, as it was suggested for the case similarity testing above. The overall user evaluation for the case based

reasoning chronic diseases diagnosis of a 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 chronic diseases diagnosis system shows an encouraging result for retrieving relevant cases and proposing solution so as to diagnosis and treat new chronic diseases cases. It also attain promising user acceptance as it is evaluated by the domain experts.

CHAPTER SIX

CONCLUSION AND FUTURE WORK

Nowadays, the application of AI in medical domain attracts many researchers especially applying the sub field of CBR for medical diagnosis. The main focus of this research is diagnosing chronic disorder using CBR approach. Therefore, this chapter concludes the overall findings of this research and based on the findings recommends the remaining problems to be investigated by other researchers.

6.1. Conclusion

Case based reasoning is a 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.

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

For the purpose of the study the required knowledge is acquired from domain experts, previous solved chronic diseases cases and other relevant documents through semi-structured interview and document analysis. The domain experts are selected from Adama Referral and Medical Hospital.

Hierarchal tree modelling technique is employed in order to model the acquired knowledge. Then case structure, which includes attributes that have direct impact on the chronic diseases diagnosis, is formulated to build the case base. The case base is constructed with 164 chronic diseases cases from Adama Referral and Medical Hospital.

The prototype is developed by using JCOLIBRI2 tool which is most appropriate and suitable tool to develop CBR applications. The retrieval and reuse task of the prototype uses K Nearest

Neighbour retrieval algorithm and voting method respectively. Feedback of domain expert and incremental learning are used to design the improvement 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 chronic diseases case). The evaluation is conducted through user evaluation.

The system testing and evaluation is done by visual interaction of user acceptance testing and using case testing method. The user acceptance testing done by prepared questionnaires which evaluated by the domain experts evaluator by using open and close ended questions, After each questions was evaluated by the respondents, the analysis was done by collecting each respondents evaluation points. Finally, calculating the average performance of the user acceptance testing for ACBRCDD prototype was done. The system performance testing was done by using test cases and its accuracy measured by confusion matrix.

Based on the result obtained from respondent' evaluation points, the proposed prototype for ACBRCDD the total average performance of the user acceptance evaluation is about 85.56% of the proposed system prototype satisfied by domain experts. Similarly, the total system performance testing using test cases were obtained about 87.8% of proposed system prototype was accuracy tested. This results shows how the domain experts satisfied with proposed ACBRCDD and they hope the developed prototype helps to solve domain problems.

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 CBR system for chronic diseases diagnosis.

6.2. Future Work

The main goal of this research is to develop a prototype CBR application for chronic diseases diagnosis. At the beginning of this research, the researcher set up different specific objectives in harmony with the overall general objective of this study. To this end, almost all objectives are achieved successfully with some challenges and constraints. Therefore, there are a number of problems to be investigated by future researchers in applying CBR in chronic diseases diagnosis.

- The retrieval algorithm used for retrieval of cases for ACBRCDD application is nearest neighbour retrieval algorithm. Since the case base of the system increases through incremental learning, the retrieval time increases linearly. Therefore, the retrieval performance will decrease from time to time. To overcome this problem investigating case maintenance techniques essential.
- In this research, the importance values of the attributes are assigned manually with the help of domain experts. The performance of the retrieval process reuse process (adaption) 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 chronic diseases diagnosis. 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 chronic diseases patient was not included. Further investigation can be conducted in order to improve the applicability of the prototype on chronic diseases diagnosis. 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 more chronic diseases so as to solve the shortage of domain experts.

References

[1] "Preventing chronic disease. A vital investment," in *WHO*, Geneva, 2005.

[2] "International Statistical Classification of Diseases and Related Health Problem," in *WHO*, Geneva, 2011.

[3] Global Status Report on Non-Communicable Disease, 2010.

[4] "Report of a joint WHO/FAO expert consultation," in *WHO/FAO*, Geneva, 2003.

[5] A. Sebhatu, The implementation of Ethiopia's Health Extension Program:Overview, Addis Ababa, Ethiopia, 2008.

[6] "World Report on Knowledge for Better Health: Strengthening Health Systems," in *WHO*, 2003/2004.

[7] Global TB report, WHO, 2011.

[8] "Health Status Indicator Fact Sheet," in *FDROEMH*, 2014.

[9] Health and health related indicators, Federal ministry of health, 2009/2010.

[10] F. J. C. C. Saydah SH, "Poor control of risk factors for vascular disease among adults with previously diagnosed diabetes," vol. 291, JAMA., 2004, pp. 335-342.

[11] B. G. B. H. e. a. Chobanian AV., "The Seventh Report of the Joint National Committee on Prevention, Detection, Evaluation,and Treatment of High Blood Pressure," JAMA., the JNC 7 report, 2003, pp. 2560-2572.

[12] O. B. D. D. Schneider KM., "Prevalence of multiple chronic conditions in the United States' Medicare," Health Quality Life Outcomes, 2009. [Online]. [Accessed 7 January 2013].

[13] S. A. L. T. G. T. M. W. W. K. e. a. Vogeli C., "Multiple chronic conditions: prevalence,health consequences, and implications for quality, care management, and costs.," in *J Gen Intern Med*, vol. 3, 2007, pp. 5-391.

[14] B. Henok, "A Case Based Reasoning Knowledge Based System for hypertion Management," Addis Ababa, Ethiopia, 2011.

[15] Health Sector Development Programme, Addis Ababa, Ethiopia: The Federal Democratic Republic of Ethiopia:Ministry of Finance and Economic Development, 2010.

[16] P. prakash, "Decision Support system in Heart Disease Diagnosis by Case Based Recommendation," vol. 4, Kerala,india, International Journal of Scientific & Technology Research, 2015.

[17] C. M. a. I. B. Stefania Montani, "CBR in Health Sciences.," international journal of scientific and technology research, Lyon,France, 2012.

[18] "The prevalence of multimorbidity in primary care and its effect on health care utilization and Cost," in *WHO-AIMS*, Gevena, 2009.

[19] B. P. a. W. S. Tesfaye F., Population Based Prevalence of High Blood Pressure among Adults in Addis Ababa:Uncovering a Silent Epidemic.BMC Cardiovascular Disorder, vol. 9, 2009.

[20] T. H. D. L. H. H. S. C. S. L. M. e. a. R. Lehnert, "health care utilization and costs of elderly persons with multiple chronic conditions," vol. 4, no. 68, p. 387–420, 2011.

[21] "World Health Statistics," in *WHO*, Geneva, 2009.

[22] "Managing multiple chronic condition," in *WFMH*, Geneva, 2010.

[23] R. K. J. a. P. E. Bergmann, "Representation in Case-Based Reasoning," in *The Knowledge Engineering Review*, vol. 20, 2005, pp. 1-4.

[24] B. Martin, "Instance-Based Learning: Nearest Neighbour with Generalisation.Msc Thesis , University of Waikato," 1995. [Online]. [Accessed 27 march 2011].

[25] R. Jagannathan, "Feature Selection for Cancer Classification:A Signal-to-noise Ratio Approach," International Journal of Scientific & Engineering Research, July 2013.

[26] H. a. S. J. Fang, "Case-Based Reasoning for Logistics Outsourcing Risk Assessment Model," Proceedings of International Conference on Enterprise and Management Innovation, 2007, pp. 1133-1138.

[27] M. R. M. &. H. A. Salem, "A Case Base Experts System for Diagnosis of Heart Disease. International Journal on Artificial Intelligence and Machine Learning," vol. 5, 2005, pp. 33-39.

[28] M. B.-S. B. B. H. &. W. S. Lenz, Case-based Reasoning Technology: From Foundations to Applications, LNAI: State of the Art. Springer, 1998.

[29] R. P. A. Sajja, ""Knowledge-Based Systems for Development," in Advanced Knowledge Based Systems: Model, Applications & Research.," vol. 1, 2010, pp. 1-11.

[30] M. R. M. M. A. a. D. K. Rabee, ""A Knowledge based system for comfort analysis of internal environment of hotels" in Proceedings of Creating Environment for Tourism," Nevada,USA, 1994, pp. 394-405.

[31] R. Burke, ""Knowledge-based recommender systems," in Encyclopedia of Library and Information Science," vol. 69, 2000, pp. 180-200.

[32] Randall Davis and Jonathan J. King, ""The Origin of Rule-Based Systems in AI," in Rule-Based Expert Systems," in *The MYCIN Experiments of the Stanford Heuristic Programming Project*, 1984, pp. 21-52.

[33] N. Deshpande, ""Expert System," in Artificial Intellegence," Pune, India, Technical Publications,

2008.

[34] J. &. H. I. Prentzas, "Categorizing approaches combining rule based and case-based reasoning," in *Expert Systems*, 2007, pp. 97-122.

[35] S. Chen, "Knowledge Representation and Reasoning Methodology based on CBR Algorithm for Modular Fixture Design," Journal of the Chinese Society of Mechanical Engineers, 2007, pp. 593-604.

[36] A. Salem, "Case Based Reasoning Technology for Medical Diagnosis," 2007.

[37] A. P. Aamodt, "Case-Based Reasoning: Foundational Issues,Methodological Variations, and System Approaches," vol. 1, no. 7, pp. 39-59, 1994.

[38] J. Kolodner, "An Introduction to Case-Based Reasoning. Artificial Intelligence Review," vol. 6, no. 1, pp. 3-34, 1992.

[39] M. Robert.J, Knowledge Based System: An Introduction to Expert System, New York:Macmillan, 1992.

[40] F. a. M. corduck, the fifth generation reading ed., A. wesley, Ed., 1983.

[41] T. Byrd, "Expert system Implementation," interview with knowledge engineers: Industrial Management and data system, vol. 95, 1995.

[42] K. a. H. McGraw, Knowledge acquisition principle and Guidelines, New Jersey:Prentice Hall, 1989.

[43] J. Ranjan, Knongwledge Acquisition, representation, and Reason, 2006.

[44] E. Cordingley, Knowledge Elicitation techniques for knowledge elicitation:principle,techniques and applications, Chichester,England: Ellis Horwood ltd., 1989.

[45] M. Gau, Knowledge Acquisition for a Diagnosis based task, NATO ASI Series, 1990.

[46] K. a. S. Ericsson, Protocol Analysis: Verbal Reports as Data, Cambridge,MA.: MIT press, 1984.

[47] C. Leodes, "Knowledge based system," technique and application, 2000.

[48] F. Raza, "Arteficial intelligence Technique in software Engineering", Hong Kong: in processing of the international multicoference of Engineering and computer science, 2009.

[49] M. K. B. a. P. Abdullah, "Knowledge based system," re-evaluation, Journal of knowledge management, 2006, pp. 22-142.

[50] J. K. a. E. P. Ralph Bergmann, ""Representation in case-based reasoning,"The Knowledge Engineering Review," 2005, pp. 1-4.

[51] D. K. D. a. A. M. Aha, "Instance-based learning algorithms," in Machine Learning, vol. 6, Boston: Kluwer Acadamic Publishers, 1991, pp. 37-66.

[52] R. Bergmann, "Formalizing Structural Case Representations," in Experience Management Foundations, Development Methodology, and Internet-Based Applications, Berlin: Springer-Verlag, 2002, pp. 63-69.

[53] N. Juristo and J.L. Morant, ""Common Framework for the Evaluation Process of KBS and Conventional Software," Knowledge Based Systems," vol. 11, pp. 145-159, October 1998.

[54] Y. B. K. &. S. O. Singh, "A Review of Studies in Machine Learning Techniques," vol. 1, no. 1, pp. 70-84, 2007.

[55] I. &. M. F. Watson, "Case-based reasoning: A Review.," vol. 4, no. 9, pp. 327-354, 1994.

[56] L. &. D. L. Kyung-Ho, "An Approach to Case-Based System for Conceptual Ship Design Assistant.," vol. 16, Expert Systems with Applications, 1999, pp. 97-104.

[57] B. Azeb, "Integrated Case Based and Rule Based Reasoning for Decision Support," Norwegian University of Science and Technology, 2009.

[58] B. D.-A. a. P. G.-C. Juan A. Recio-García, "GAIA – Group of Artificial Intelligence Applications, University Complutense of Madrid.," 2008. [Online]. [Accessed 14 May 2014].

[59] J. D. U. a. J. W. Hector Garcia-Molina, "The Worlds of Database Systems," in Database Systems, Upper Saddle River New Jersey 07458: Prentice Hall: The Complete Book, 2009, pp. 2-21.

[60] M. R. a. R. Weber, Case-Based Reasoning, Springer-Verlag Berlin Heidelberg, 2013.

[61] P. R. a. H. S. Christopher D. Manning, ""Boolean retrieval," An Introduction to Information Retrieval," Cambridge University Press, 2009, pp. 1-2.

[62] R. E. Pino, "Network Science and Cybersecurity," Springer Science and Business Media, 2013, pp. 155-157.

[63] I. &. A. H., "Evaluation of jCOLIBRI: Master‟s Thesis,Maradalen University," Sweden, 2008.

[64] K. D. Ashley, Modeling legal arguments: Reasoning with cases and hypotheticals,Bradford Books, 1991.

[65] E. L. A. K. D. &. B. L. K. Rissland, "Case-based reasoning and law. The Knowledge Engineering Review," no. 20, pp. 293-298, 2005.

[66] A. &. B.-C. T. Wyner, Argument schemes for legal case-based reasoning.Proceedings of the 2007 conference on Legal Knowledge and Information.The Twentieth Annual Conference, IOS Press, 2007.

[67] B. P. Allen, "Case-based reasoning: Business applications. Communications of the ACM," vol. 37, pp. 40-42, 1994.

[68] M. Lenz, "Experiences from deploying cbr applications in electronic commerce," 1999.

[69] S. V. S. P. &. a. A. Shokouhi, "An overview of case-based reasoning applications in drilling

engineering. Artificial Intelligence Review," pp. 1-13, 2011.

[70] I. Watson, "Case-based reasoning: A review. Knowledge Engineering Review," vol. 9, 1994.

[71] J. &. K. R. Kolodner, "Using Experience in Clinical Problem Solving:Introduction and Framework," vol. 3, no. 17, pp. 420-431, 1987.

[72] P. Koton, "Reasoning about Evidence in Causal Explanations," Morgan Kaufmann, San Mateo, CA, In the Proceedings of the American Association for Artificial Intelligence, 1988, pp. 256-261.

[73] R. Turner, "Organizing and using schematic knowledge for medical diagnosis," J. Kolodner, Ed., Morgan Kaufmann, San Mateo, In the Proceedings of the First Workshop on CBR, 1988, pp. 435-446.

[74] I. K. E. &. S. K. Bichindaritz, "Case-Based Reasoning in CARE PARTNER: Gathering Evidence for Evidence-Based Medical Practice," pp. 334-345, 1998.

[75] A. B. I. S. R. &. P. P. Holt, "Medical Applications in case based reasoning," vol. 3, no. 20, pp. 289-292, 2006.

[76] R. &. G. L. Schmidt, "Temporal abstractions and Case-based Reasoning for Medical Course Data Two Prognostic Applications," in *Lecture Notes in Computer Science Springer-Verlag Heidelberg*, P. Perner, Ed., In the Proceedings of the Second International Workshop on Machine Learning Pattern Recognition and Data Mining in Pattern Recognition, 2001, pp. 23-34.

[77] B. &. P. E. Lopez, "Case-Base Planning for Medical Diagnosis," pp. 96-105, 1993.

[78] C. &. Z. J. Bradburn, "The Application of Case-based Reasoning to the Tasks of Healthcare Planning," pp. 365-378, 1993.

[79] I. Bichindaritz, "Case-Based Reasoning Adaptive to Several Cognitive Tasks," pp. 391-400, 1995.

[80] M. A. P. &. P. G. Haddad, "Feasibility Analysis of a Case-based Reasoning System for Automated Detection of Coronary Heart Disease from Myocardial Scintigrams," vol. 1, no. 9, pp. 61-78, 1997.

[81] C. &. W. P. Marling, "Case-Based Reasoning in the Care of Alzheimer‟s Disease Patients," pp.702-715, 2001.

[82] E. &. P. B. Bareiss, "Protos: An Exemplar-Based Learning Apprentice," pp. 12-23, 1987.

[83] R. M. K. T. V. B. E. T. J. &. J. K. Macura, "Computerized Case-based Instructional System for Computed Tomography and Magnetic Resonance Imaging of Brain Tumors," vol. 4, no. 29, pp. 497-506, 1994.

[84] M. &. A. A. Grimnes, "A Two Layer Case-Based Reasoning Architecture for Medical Image Understanding," pp. 164-178, 1996.

[85] I. &. P. S. Bichindaritz, "PhylSyst:un système d'intelligence artificielle pour l'analyse cladistique, Biosystema," no. 12, pp. 17-55, 1994.

[86] I. &. P. S. &. S. S. Bichindaritz, "Knowledge Based Phylogenetic Classification Mining," pp. 163-172, 2004.

[87] P. Perner, "An Architecture for a CBR Image Segmentation System," vol. 6, no. 12, pp. 749-759, 1999.

[88] I. &. S. B. Bichindaritz, "Concept Learning in a Real World Domain Technique et Sciences Informatiques," vol. 4, no. 11, pp. 69-98, 1992.

[89] J. Berger, "Roentgen: Radiation Therapy and Case-based Reasoning," pp. 171-177, 1994.

[90] G. M. C. &. S. L. Petot, "An Artificial Intelligence System for Computer Assisted Menu Planning," vol. 9, no. 98, pp. 1009-1014, 1999.

[91] S. B. R. P. L. &. S. M. Montani, "A Multi-Modal Reasoning Methodology for Managing IDDM Patients.Informatics, International Journal of Medical," vol. 58, no. 59, pp. 243-256, 2000.

[92] L. Gierl, "ICONS: Cognitive basic functions in a case-based consultation system for intensive care.," pp. 230-236, 1993.

[93] K. Fenstermacher, "An Application of Case-based Instruction in Medical Domains," pp. 50-54, 1996.

[94] R. M. S. B. R. P. L. &. G. L. Schmidt, "Case-Based Reasoning for Medical Knowledge-based Systems," vol. 3, no. 64, pp. 355-367, 2001.

[95] G. Biazen, "Application of case based recommender system to advise students in field of study selection at higher education in Ehiopia," Addis Ababa University, Ethiopia, 2013.

[96] J. Alemu, "A Case-Based Approach for Designing Knowledge Base System for Addis Resource Center (ARC):The Case of Warmline Clincian Consultation Service.," Addis Ababa University, Ethiopia, 2010.

[97] W. Getachew, "Application of Case-Based Reasoning for Anxiety Disorder Diognosis," Addis Ababa University, Ethiopia, 2012.

[98] M. A. A. Ahmar, ""A Prototye Student Advising Expert System Supported with an Object-Oriented Database,"" International Journal of Advanced Computer Science and Applications,Special Issue on Artificial Intelligence, 2011, pp. 6.

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

[100] T. A. C. K. L. a. Z. R. W. Byrd, "A Synthesis of research on requirements analysis and knowledge acquisition techniques," vol. 1, no. 16, pp. 117-138, 1992.

[101] T. Sagheb, "A Conceptual Model of Knowledge Elicitation," College of Business,technology and Communication, 2009.

[102] A. Miller, "Problem Detection for Situation Assessment In Case-Based Reasoning For Diabetes Management," Msc thesis,The Russ College of Engineering and technology,Ohio University, 2009.

[103] "Chronic Disease," 15 May 2008. [Online]. [Accessed 15 May 2008].

[104] A. S. A. J. e. a. McGlynn EA., "The quality of health care delivered to adults in the United States," vol. 348, N Engl J Med., 2003, pp. 2635-2645.

[105] T. Singh, ""Prototype Expert System for Diagnosis and Treatment of Diabetes"," Graduate Project, USA, 2001.

[106] "American Diabetes Association, "Insulin Basics"," [Online]. [Accessed 10 November 2012].

[107] "International Diabetes Federation, "What is Diabetes"," [Online]. [Accessed 20 October 2012].

[108] "National Diabetes Information Clearinghouse, "Diagnosis of Diabetes and Prediabetes"," [Online]. [Accessed 19 October 2012].

[109] Standard Treatment Guide line for General Hospitals(STGLGH), Ethiopia: Drug Administration Control Agency, 2010.

[110] s. Sharm, ""Restrictive Lung Disease"," 19 04 2008. [Online]. [Accessed 19 December 2008].

[111] P. Makhfi, "Introduction to knowledge modeling and neural network," 2011. [Online]. [Accessed 27 March 2011].

[112] K. P. G. M. A. L.-C. K. P. G. A. B. F. R. J. R. I. Lundgren-Cayrol, "Explor@ Advisory Agent:Tracing the Student's Trail," Proceeding of WebNet'2001, World Conference of the WWW and Internet, 2001, pp. 802–808.

[113] H. Yemisrach, "Application of Case-Based Reasoning System in legal Knowledge-Based Systems:A Prototype in Children Criminal Cases in Ethiopia," Addis Ababa University , Ethiopia, 2010.

[114] G. Tinsae, "Application of an Enhanced Case based Recommender System to assist Students' in Field of Study Selection at Higher Education in Ethiopia," Adama Science and Technology University, Ethiopia, January 2015.

[115] D. McSherry, "Precision and Recall in Interactive Case-Based Reasoning. In Case Based Reasoning Research and Development (ICCBR), Lecture notes in Artificial Intelligence," 2001, pp. 392-306.

[116] M. H. R. a. D. A. Junker, "On the Evaluation of Document Analysis Components by Recall, Precision, and Accuracy.Proceedings of the Fifth International Conference on Document Analysis and Recognition," 1999, pp. 713-716.

Appendixes

Appendix I

Interview Questions

The main objective of this study is to design and develop a CBR (Case-Based Reasoning) system for diagnosing patients with chronic diseases in order to provide advice to domain experts and enhance services given to patients.

1. How health problems are defined from the aspect of their categories in chronic disease?
2. What are the method and techniques general practitioner used to identify the causes of patient diseases in each chronic diseases category?
3. What are the common types of diseases that can identify using their common pillar symptom?
4. How general practitioners identify the unique characteristics of each diseases?
5. What are common symptoms each disease and their association with one another in their respective categories?
6. What are the main inherits and environmental risk factors of diseases during patient diagnosis?
7. Does a disease have common behaviors and their complication due to different factors?
8. What are the decisions of general practitioner in providing good treatment?
9. What are the most important areas focused by general practitioner to diagnose the patient?
10. What are recommended techniques used to manage specific diseases?

Appendix II

ADAMA SCIENCE AND TECHNOLOGY UNIVERSITY

School of Electrical and Computing

Department of Computing

Dear Participant,

I' am currently writing my research thesis on the area of Application of Case Based Reasoning for Chronic Diseases Diagnosis to finalize this research study that I've been attending at Adama Science and Technology University, I need to have some survey question to be answer by domain experts.

My research title is Application of Case Based Reasoning for Chronic Diseases Diagnosis; this survey questionnaire helps to evaluate the developed system prototype its performance testing as well as user acceptance testing by the domain experts. So that you have to give your opinion and feedback and the assessment of Application of Case Based Reasoning for Chronic Diseases Diagnosis would be developed as a solution. You have every right to with draw from this survey at any point, if you are not comfortable. Thank you very much for spending time to read and answer these short questionnaires!

Thank you very much again for your support!

Sultan, Ermias

Adama Science and Technology University,

Adama.

January, 2017

Appendix III

Questionnaires for ACBRDD (Application of Case Based Reasoning for Chronic Disease Diagnosis) in the Case of Adama Referral and Medical College

Prototype Evaluation form for the Domain Expert

This is an evaluation form to be filled by chronic disease diagnosis experts(team) in order to evaluate the applicability of the case-based reasoning system in chronic disease diagnosis such as (Hypertension, Diabetes mellitus, Pulmonary TB, COPD). I thank you in advance for your willingness and valuable time.

Please, read and answer all questions according to the instruction given below.

I. Please circle the options for each of the following questions given below.

1. How do you rate the adequacy and clarity of the system in decision support?

   1. Poor 2. Fair 3. Good 4. Very good 5. Excellent

2. How do you rate the relevancy of the attributes in representing chronic disease diagnosis?

   1. Poor 2. Fair 3. Good 4. Very good 5. Excellent

3. Is the user interface of prototype interactive?

   1. Poor 2. Fair 3. Good 4. Very good 5. Excellent

4. Is the prototype easy to use?

   1. Poor 2. Fair 3. Good 4. Very good 5. Excellent

5. How do you rate relevancy of the retrieved cases in the decision making process?

   1. Poor 2. Fair 3. Good 4. Very good 5. Excellent

6. Is the system efficient in time and memory?

   1. Poor 2. Fair 3. Good 4. Very good 5. Excellent

7. How do you rate the significance of the system in the domain area?

    1. Poor 2. Fair 3. Good 4. Very good 5. Excellent

II. <u>Please state your opinions and suggestions regarding the performance and applicability of the ACBRCDD in the domain area.</u>

8. How is ACBRCDD different from diagnosing the chronic diseases from the human experts?

    _____

9. What issues are covered by the ACBCRDD Advisory system?

    _____

10. Please state your suggestion if you think that there are uncovered issues by prototype ACBCRDD?

    _____

11. Do you believe that, ACBRCDD prototype handle chronic diseases diagnosing tasks?

    _____

12. Do you think that, ACBRCDD contribute in diagnosing and treatment of chronic diseases can support human experts particularly, Primary healthcare workers and General Practitioners in the domain area? How?

    _____

13. What are the limitations of ACBRCDD as you see its performance when you are interacting? With it?

    _____

14. What are the strengths of the ACBRCDD in solving problems in the domain areas?

Appendix IV.

SAMPLE CASES

case_id	age	sex	disease_category	present_illness_1	present_illness_2	premonitory_symptom	systolic_blood_pressure	diastolic_blood_pressure	fasting_blood_sugar
Case1	35	Female	Cardiovascular Disease	Nausea	Headache	Fainting	135	80	0
Case10	23	Male	Cardiovascular Disease	Shortness of breath	Headache	Fainting	128	85	0
Case100	17	Female	Infectious Disease	Weight loss	Night sweats	High Fever	0	0	0
Case101	27	Female	Infectious Disease	Headache	Blood present in urina	Productive Cough	0	0	0
Case102	69	Female	Infectious Disease	Fever	Night sweats	Productive Cough	0	0	0
Case103	23	Male	Infectious Disease	Weight loss	Night sweats	High Fever	0	0	0
Case104	30	Male	Infectious Disease	Night sweats	Weight loss	Productive Cough	0	0	0
Case105	40	Female	Infectious Disease	Hemoptysis	Fever	Productive Cough	0	0	0
Case106	30	Male	Infectious Disease	Night sweats	Headache	Productive Cough	0	0	0
Case107	55	Female	Infectious Disease	Fever	Weight loss	Productive Cough	0	0	0
Case108	28	Female	Infectious Disease	Night sweats	Chest pain	Amount of fluid increases dramat	0	0	0
Case109	20	Male	Infectious Disease	Weakness	Fever	Productive Cough	0	0	0
Case11	56	Male	Cardiovascular Disease	Nausea	Shortness of breath	Dizziness	173	107	0
Case110	22	Male	Infectious Disease	Weight loss	Headache	Productive Cough	0	0	0
Case111	45	Female	Infectious Disease	Fever	Night sweats	Productive Cough	0	0	0
Case112	23	Female	Infectious Disease	Weight loss	Night sweats	High Fever	0	0	0
Case113	26	Male	Infectious Disease	Chest pain	Weight loss	Productive cough	0	0	0
Case114	32	Female	Infectious Disease	Fever	Headache	Productive cough	0	0	0
Case115	36	Male	Infectious Disease	Night sweats	Chest pain	Amount of fluid increases dramat	0	0	0
Case116	47	Male	Infectious Disease	Fever	Night sweats	Amount of fluid increases dramat	0	0	0
Case117	24	Female	Infectious Disease	Weight loss	Night sweats	High Fever	0	0	0
Case118	58	Male	Infectious Disease	Hemoptysis	Fever	Productive Cough	0	0	0
Case119	63	Male	Infectious Disease	Night sweats	Headache	Productive Cough	0	0	0
Case12	69	Male	Cardiovascular Disease	Nausea	Shortness of breath	Heart palpitation	164	105	0
Case120	29	Female	Infectious Disease	Coughing up blood	Weight loss	Productive Cough	0	0	0

tuberculin_skin_test	ketone	co_illness	co_medication	smoking_drinking	family_history	diagnostic_solution	recommended_treatment
Not_Available	Not_Available	No	No	Yes	Yes	Hypertension	Start first line drug by consic
Not_Available	Not_Available	Yes	No	No	Yes	Hypertension	Start first line drug by consic
Postive TST	Not_Available	No	No	Not_Available	Yes	Pulmonary TB	Take drugs(tablets) for six n
Postive TST	Not_Available	No	No	Not_Available	No	Pulmonary TB	Take drugs(tablets) for six n
Postive TST	Not_Available	No	No	Not_Available	No	Pulmonary TB	Take drugs(tablets) for six n
Postive TST	Not_Available	No	No	Not_Available	No	Pulmonary TB	Take drugs(tablets) for six n
Postive TST	Not_Available	No	No	Not_Available	No	Pulmonary TB	Take drugs(tablets) for six n
Postive TST	Not_Available	No	No	Not_Available	No	Pulmonary TB	Take drugs(tablets) for six n
Postive TST	Not_Available	No	No	Not_Available	Yes	Pulmonary TB	Take drugs(tablets) for six n
Postive TST	Not_Available	Yes	Yes	Not_Available	No	Pulmonary TB	Take drugs(tablets) for six n
Postive TST	Not_Available	No	No	Not_Available	Yes	Pulmonary TB	Take drugs(tablets) for six n
Postive TST	Not_Available	No	No	Not_Available	No	Pulmonary TB	Take drugs(tablets) for six n
Not_Available	Not_Available	No	No	No	No	Hypertension	Start first line drug by consic
Postive TST	Not_Available	No	No	Not_Available	No	Pulmonary TB	Take drugs(tablets) for six n
Postive TST	Not_Available	Yes	Yes	Not_Available	No	Pulmonary TB	Take drugs(tablets) for six n
Postive TST	Not_Available	No	No	Not_Available	Yes	Pulmonary TB	Take drugs(tablets) for six n
Postive TST	Not_Available	No	No	Not_Available	Yes	Pulmonary TB	Take drugs(tablets) for six n
Postive TST	Not_Available	No	No	Not_Available	Yes	Pulmonary TB	Take drugs(tablets) for six n
Postive TST	Not_Available	No	No	Not_Available	Yes	Pulmonary TB	Take drugs(tablets) for six n
Postive TST	Not_Available	No	No	Not_Available	Yes	Pulmonary TB	Take drugs(tablets) for six n
Postive TST	Not_Available	Yes	Yes	Not_Available	No	Pulmonary TB	Take drugs(tablets) for six n
Postive TST	Not_Available	Yes	Yes	Not_Available	Yes	Pulmonary TB	Take drugs(tablets) for six n
Postive TST	Not_Available	No	No	Not_Available	No	Pulmonary TB	Take drugs(tablets) for six n
Not_Available	Not_Available	No	No	Yes	Yes	Hypertension	Start first line drug by consic
Postive TST	Not_Available	No	No	Not_Available	No	Pulmonary TB	Take drugs(tablets) for six n

Appendix V.

SOURCE CODE OF THE PROTOTYE SYSTEM

The Main Function

```
package sulti1;

import java.awt.Dimension;

public class Frestart implements StandardCBRApplication {
    static QueryDialog qf;

    private static Frestart _instance = null;

    public static Frestart getInstance() {
        if (_instance == null)
            _instance = new Frestart();
        return _instance;
    }
    private Frestart() {
    }
    /** Connector object */
    Connector _connector;
    /** CaseBase object */
    CBRCaseBase _caseBase;
    SimilarityDialog similarityDialog;
    public ResultDialog resultDialog;
    public void configure() throws ExecutionException {
        try {
                // Create a data base connector
            _connector = new DataBaseConnector();
            // Init the ddbb connector with the config file
            _connector.initFromXMLfile(jcolibri.util.FileIO
                    .findFile("sulti1/databaseconfig.xml"));
            // Create a Lineal case base for in-memory organization
            _caseBase = new LinealCaseBase();

            resultDialog = new ResultDialog(main);

        } catch (Exception e) {
```

```
                        e.printStackTrace();
                throw new ExecutionException(e);
        }
    }
    public CBRCaseBase preCycle() throws ExecutionException {
        // Load cases from connector into the case base
        _caseBase.init(_connector);
        System.out.println("connector is" + _connector);
        // Print the cases
        java.util.Collection<CBRCase> cases = _caseBase.getCases();
        System.out.println(cases.size() + " found");
        for (CBRCase c : cases)
            System.out.println(c);
        return _caseBase;
    }
    public void cycle(CBRQuery query) throws ExecutionException {
        // Obtain configuration for KNN
        // similarity configuration is based on the query
        similarityDialog = new SimilarityDialog(main);
        similarityDialog.setVisible(true);
        NNConfig simConfig = similarityDialog.getSimilarityConfig();
        simConfig.setDescriptionSimFunction(new Average());
        // Execute NN
        Collection<RetrievalResult> eval = NNScoringMethod.evaluateSimilarity(
                _caseBase.getCases(), query, simConfig);
        // Select k cases
        Collection<CBRCase> selectedcases = SelectCases.selectTopK(eval,
                similarityDialog.getK());
        // Show result
        resultDialog.showCases(eval, selectedcases);
        resultDialog.setVisible(true);



    }
```

```
    public void postCycle() throws ExecutionException {
        MySql.close();
        _connector.close();
    }

    public static JFrame main;

    void showMainFrame() {
        main = new JFrame("Application of  CBR");
        main.setResizable(true);
        main.setUndecorated(true);
        JLabel label = new JLabel(new ImageIcon(
                jcolibri.util.FileIO
                        .findFile("sulti1/gui/jcolibri2.jpg")));
        main.getContentPane().add(label);
        main.pack();
        Dimension screenSize = java.awt.Toolkit.getDefaultToolkit()
                .getScreenSize();
        main.setBounds((screenSize.width - main.getWidth()) / 2,
                (screenSize.height - main.getHeight()) / 2, main.getWidth(),
                main.getHeight());
        main.setVisible(true);
    }

    public static void main(String[] args) {

        Frestart recommender = getInstance();
        recommender.showMainFrame();
        try {
            recommender.configure();
            recommender.preCycle();
            qf = new QueryDialog(main);
            boolean cont = true;
            while (cont) {
                qf.setVisible(true);
```

```
            CBRQuery query = qf.getQuery();

            //System.out.println("Query: " + query);
            recommender.cycle(query);
            int ans = javax.swing.JOptionPane.showConfirmDialog(null,
                    "CBR cycle finished, query again?", "Cycle finished",
                    javax.swing.JOptionPane.YES_NO_OPTION);
            cont = (ans == javax.swing.JOptionPane.YES_OPTION);
        }
        recommender.postCycle();
    } catch (Exception e) {
        org.apache.commons.logging.LogFactory.getLog(Frestart.class)
                .error(e);
        javax.swing.JOptionPane.showMessageDialog(null,
                "error!! " + e.getMessage());
        e.printStackTrace();

        System.out.println(e.getMessage());
    }
    // System.exit(0);
}
```

Appendix VI

When the system starts the following screen is displayed.

Appendix VII

Sample Cases of the Experiment system

Initial situation of case1: A compliant comes with the age of 52 to Adama Referral and Medical Hospital. He doesn't have co-illness and he didn't take any medication. He has a symptom of tiredness, weight loss, and polyuria. His systolic blood pressure and systolic blood pressure shows the same reading in different check-up. The systolic blood pressure is 170, diastolic blood pressure is 100 and the fasting blood sugar is 123. He does not smoke cigarette and drink alcohol. He Have family history of diabetes mellitus. Question. Is the compliant a chronic diseases patient? If he is a chronic diseases patient, which types of chronic diseases he has. What kind of chronic diseases treatment should be applied? From the above case description the following facts are extracted and fed as the chronic diseases case query to the prototype. The query form is found in the appendix VII.

- The complaint's age is 52 and sex female
- He didn't show a sign of unconsciousness
- The systolic blood pressure is 170, diastolic blood pressure is 100 and fasting blood sugar 123
- The compliant has no co-illness and didn't take any medication
- He is not smoker and alcoholic
- He has a symptom of tiredness
- He has a symptom of weight loss
- He has a symptom of polyuria
- Disease categories of a compliant is Endocrine condition

Recommendation by the system: After the above facts of the new case are fed for the prototype, the prototype computes the similarity of the chronic disease 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 type2 diabetes patient and the patient has to take glibenclamide and metformin. The recommendation proposed by the system is correct as compared to the physician recommendation for the

compliant at Adama Referral and Medical Hospital. The results of the retrieval process are presented in the appendix VIII, appendix IV and appendix X respectively.

Initial situation of case2: This case describe a 32 years old patient complain of sex male come to Adama Referral and Medical Hospital. The person is smoke cigarettes so the person is facing in increasing breathlessness when exercising or moving around, a persistent cough with phlegm that never seems to go away, frequent chest infections, particularly in winter, and wheezing. Question. Is the compliant a chronic diseases patient? If he is a chronic diseases patient, which types of chronic diseases he has. What kind of chronic diseases treatment should be applied? From the above case description the following facts are extracted and fed as the chronic diseases case query to the prototype.

- The complaint‘s age is 50 and sex male
- He has a symptom of sputum production
- They have a symptom of frequent respiratory infection
- He didn’t show a sign of wheezing
- They have a symptom of shortness of breath
- He is smoker and alcoholic
- Disease categories of a complaint is respiratory disease

Recommendation by the system: After the above facts of the new case are fed for the prototype, the prototype computes the similarity of the chronic disease 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 chronic bronchitis patient and the patient has to take bronchodilator to treat chronic bronchitis. The recommendation proposed by the system is correct as compared to the physician recommendation for the compliant at Adama Referral and Medical Hospital.

Initial situation of case3: A compliant comes with the age of 50 to Adama Referral and Medical Hospital. 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 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 Adama Referral and Medical Hospital.

Appendix VIII:

Query form for new chronic diseases case

CASE BASED REASONING FOR CHRONIC DISEASES DIAGNOSIS

Enter age of a complaint: **52**

Enter sex of a complaint: **Female**

Enter disease categories of a complaint: **Endocrine Condition**

Enter present illness1 of a complaint: **Tiredness**

Enter present illness2 of a complaint: **Weight loss**

Enter premonitory symptoms of a complaint: **Polyuria**

Enter systolic blood pressure of a complaint: **128**

Enter diastolic blood pressure of a complaint: **85**

Enter fasting blood pressure of a compliant: **123**

Has the complaint have a ketone (yes/no): **Yes**

Is the complaint has a co-illness (yes/no): **No**

Is the complaint take co-medication (yes/no): **No**

Is the compliant a smoker and alcoholic (yes/no): **No**

Enter family history of a complaint (yes/no): **Yes**

Appendix IX:

Retrieved relevant cases from the case base for the sample query by the prototype

RELEVANT CASES

With Case 51

Has similarity of 0.862581231758

A person with age of 41, Sex Female, with DC of endocrinal condition, who was PI1 of dizziness, who was PI2 of tiredness, who has PS of polyuria, SBP 124, DBP 84, FBS 125, who was a ketone, who was no co-illness, who haven't took co-medication, who was not smoker and drinker, who was family history of diabetes

The physician decision was

The type of disease is type2 diabetes

The recommendation given by physician is take medication of glibenclamide and metformin and increase physical exercise to control the condition.

With Case 29

Has similarity of 0.847583266117

A person with age of 56, Sex male, with DC of endocrinal condition, who was PI1 of weight loss, who was PI2 of dizziness, who has PS of polyphagia, SBP 125, DBP 85, FBS 181, who was a ketone, who was no co-illness, who haven't took co-medication, who was not smoker and drinker, who was family history of diabetes.

The physician decision was

The type of disease is type2 diabetes

The recommendation given by physician is take medication of glibenclamide and metformin and increase physical exercise to control the condition.

Appendix X:

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

With Case 68

Has similarity of 0.830084295203

A person with age of 73, Sex Female, with DC of endocrinal condition, who was PI1 of blurry vision, who was PI2 of dizziness, who has PS of polyuria, SBP 125, DBP 84, FBS 111, who was a ketone, who was no co-illness, who haven't took co-medication, who was not smoker and drinker, who was family history of diabetes

The physician decision was

The type of disease is type2 diabetes

The recommendation given by physician is take medication of glibenclamide and metformin and increase physical exercise to control the condition.

With Case 56

Has similarity of 0.815411117737

A person with age of 37, Sex Female, with DC of endocrinal condition, who was PI1 of weight loss, who was PI2 of tiredness, who has PS of polydipsia, SBP 125, DBP 85, FBS 180, who was a ketone, who was no co-illness, who haven't took co-medication, who was not smoker and drinker, who was family history of diabetes

The physician decision was

The type of disease is type_2 diabetes

The recommendation given by physician is take medication of glibenclamide and metformin and increase physical exercise to control the condition.

Appendix XI:

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

With Case 75

Has similarity of 0.815146094216

A person with age of 60, Sex Female, with DC of endocrinal condition, who was PI1 of blurry vision, who was PI2 of dizziness, who has PS of polydipsia, SBP 120, DBP 82, FBS 170, who was a ketone, who was no co-illness, who haven't took co-medication, who was not smoker and drinker, who was family history of diabetes

The physician decision was

The type of disease is type_2 diabetes

The recommendation given by physician is take medication of glibenclamide and metformin and increase physical exercise to control the condition.

With Case 62

Has similarity of 0.815146094216

A person with age of 40, Sex Female, with DC of endocrinal condition, who was PI1 of blurry vision, who was PI2 of dizziness, who has PS of polyuria, SBP 123, DBP 80, FBS 105, who was a ketone, who was no co-illness, who haven't took co-medication, who was not smoker and drinker, who was not family history of diabetes

The physician decision was:

The complaint is at normal stage it is free from diabetes

However; the complaint is at higher risk for diabetes

With Case 27

Has similarity of 0.8051460942136083

A person with age of 50, Sex Female, with DC of endocrinal condition, who was PI1 of dizziness, who was PI2 of blurry vision, who has PS of polydipsia, SBP 130, DBP 82, FBS 93, who was not have a ketone, who was no co-illness, who haven't took co-medication, who was not smoker and drinker, who was not family history of diabetes

The physician decision was

The complaint is at normal stage it is free from diabetes

However; the complaint is at higher risk for diabetes

With Case 28

Has similarity of 0.800510940242

A person with age of 62, Sex Male, with DC of endocrinal condition, who was PI1 of blurry vision, who was PI2 of tiredness, who has PS of polyphagia, SBP 138, DBP 92, FBS 170, who was a ketone, who was no co-illness, who haven't took co-medication, who was not smoker and drinker, who was family history of diabetes

The physician decision was

The type of disease is type_2 diabetes

The recommendation given by physician is take medication of glibenclamide and metformin and increase physical exercise to control the condition.

Therefore, the type of disease and recommended treatment for all cases by the system is:

Type_2 diabetes and take glibenclamide and metformin