JIMMA UNIVERSITY

COLLEGE OF NATURAL SCIENCES

DEPARTMENT OF INFORMATION SCIENCE

MSC. IN INFORMATION AND KNOWLEDGE MANAGEMENT

THESIS TITLE: DEVELOPING KNOWLEDGE BASED SYSTEM USING DATA MINING TECHNIQUES FOR DIAGNOSIS AND TREATMENT OF DIABETES

BY: KEDIR EYASU

ADVISORS:

1. **WORKU JIMMA (PHD CANDIDATE)**
2. **TAKELE TADESE (MSC)**

SUBMITTED TO: INFORMATION SCIENCE DEPARTMENT

JIMMA, ETHIOPIA

NOVEMBER, 2018

JIMMA UNIVERSITY

COLLEGE OF NATURAL SCIENCES

DEPARTMENT OF INFORMATION SCIENCE

DEVELOPING KNOWLEDGE BASED SYSTEM USING DATA MINING TECHNIQUES FOR DIAGNOSIS AND TREATMENT OF DIABETES.

BY

KEDIR EYASU

A Thesis Submitted to the Department of Information Science of Jimma University in partial fulfilments for the Degree of Master of Science in Information Science

1. **PRINCIPAL ADVISOR: WORKU JIMMA (PHD CANDIDATE)**
2. **CO-ADVISOR: MR. TAKELE TADESE (MSC.)**

November, 2018

Jimma, Ethiopia

JIMMA UNIVERSITY

COLLEGE OF NATURAL SCIENCES

DEPARTMENT OF INFORMATION SCIENCE

DEVELOPING KNOWLEDGE BASED SYSTEM USING DATA MINING TECHNIQUES FOR DIAGNOSIS AND TREATMENT OF DIABETES.

BY

KEDIR EYASU

Members of the examining board:

Name	Title	Signature	Date
	Chair person		
	Co-Advisor		
	Advisor		
	Internal-Examiner		
	External-Examiner		

DECLARATION

I declare that the thesis is my original work and it has not been presented for a degree in any other university. All the material sources used in this work are appropriately acknowledged.

____

KEDIR EYASU

November, 2018

This thesis has been submitted to the department for examination with our approval as University advisors:

Principal Advisor: Worku Jimma (PhD Candidate) __________

Co-Advisor: Takele Tadese (MSc.) ____________________

November, 2018

DEDICATION

In my heartfelt this work is dedicated to my brothers and sisters those “Qeerroo and Qarree” of Oromoo that were lost their life (from 2013-2017) in autonomy of our people including me and my family.

ACKNOWLEDGEMENT

Frist and Foremost, I would like to express my sincere gratitude to my advisor Worku Jimma (PhD Candidate) for the continuous support of my research, for his patience, motivation, enthusiasm, and immense knowledge. His guidance helped me in all the time of research and writing of this thesis.

I gratefully thank my Co-advisor Takele Tadese for his commitments and patience for reading every section of the thesis, valuable comments, encouragement and guidance from initial to the final level of the thesis that lead me to end my research work.

I would like to say thank you to the staffs of Jimma University Specialized Hospital, Adama Medical College Hospital. Saint Paulos Millennium Medical Hospital and in their assistance by delivering important advice and information on the domain area.

Finally, I would like to thanks to Jimma University Department of Information Science staffs, Mettu University, my families and all the friends and colleagues who provided me with the facilities being required and conductive conditions for this research.

Contents

LIST OF TABLES

LIST O FIGURES

LIST OF ACRONYMS AND ABBREVIATIONS

ADA-------------American Diabetes Association

AI----------------Artificial Intelligent

BMI-----------Body Mass Index

CLIPS-----------C Language Integrated Production System

CMD----------Cardio Vascular Disease

CRISP-DM-----CRoss-Industry Standard Process for Data Mining

DKA---------Diabetic ketoacidosis

DM--------------Diabetes Mellitus

ED---------------Erectile Dysfunction

EM--------------Expectation Measure

FPG-----------Fasting Plasma Glucose

IDDM-----------Insulin Dependent Diabetes Mellitus

IDDM-----------Insulin-dependent diabetes mellitus

IDF--------------International Diabetes Federation

IT----------------Information Technology

JRip-------------Repeated Incremental Pruning

KBS-------------Knowledge Based System

KBSs------------Knowledge Based Systems

KDD-------------Knowledge Discover Database

Km---------------Kilometer

Mg---------------milligram

Ml---------------milliliter

MNT-----------Medical Nutrition Therapy

OGTT--------Oral glucose tolerance test

PART-----------Partial Decision Tree

PSM-------------Problem Solving Method

U.S. -------------United States

ABSTRACT

Diabetes is a disease that affects the body's ability to produce or use insulin. According to International diabetes Federation Ethiopia is one of the 32 countries in African region. 425 million people have diabetes in the world and more than 16 million people in the Africa Region; by 2045 it will be around 41 million. There were 2.567.900 cases of diabetes in Ethiopia in 2015.In the incident of Ethiopia different problems are observed in health care centers. From different perspectives, these problems are the scarcity of domain experts, practitioners, domain experts' skills, health facilities etc. The general objective of this study is to design and develop prototype knowledge based system using data mining techniques for diagnosis and treatment of diabetes. In this study, to develop prototype knowledge base system using data mining techniques for diagnosis and treatment of diabetes is proposed by applying experimental research design. The researcher used domain expert knowledge as supplement of data mining techniques knowledge. To identify the best performance model for extracting the hidden knowledge (i.e. rules), three experiments for three classification algorithms were conducted these are J48, PART and JRip. Finally, the researcher decided to use the results of J48 classification algorithm in the development of the prototype knowledge based system using data mining for diagnosis and treatment of diabetes because it registered better performance than other classifiers. The developed model was tested with test performance of instance classified correctly and accuracy and only which performance score more than 95.1515 % accuracy was used as a knowledge base for the KBS development for a better efficiency and effective. So, Data Mining solves the knowledge acquisition problem of knowledge based perceptive by supplying extracted knowledge to KBS. Weka used for model construction and evaluation, Ultimate Visual basic studio 2013(Vb.net) for using data mining results as store knowledge base and as front side of prototype and common lisp prolog (Clisp) used for obtained knowledge backend coding. So, prototype knowledge based using data mining for diagnosis and treatment of diabetes was developed by integrating Vb.net and Clisp (i.e. clisp.net) tools. Finally, testing of the developed prototype KBS is done to evaluate the work of the system. The first one is testing in terms of using cases collected from hospital with that the system measured an accuracy of 92%. The second one is the user acceptance of the system is evaluated by the potential users' of the system and achieved 91.43% recital. But further exploration has to be done to refine the knowledge base, languages, feature add etc. are need future advances.

Keywords: Knowledge-Based System, Data mining, J48, PART, JRip, Diabetes.

CHAPTER ONE

INTRODUCTION

1.1. BACKGROUND OF THE STUDY

Knowledge based system (KBS) is one of the areas of artificial intelligence. KBS also known as an expert system is a computer program that contains the knowledge and analytical skills of one or more human experts in a specific problem domain. The aim of the design of the expert system is to capture the knowledge of a human expert relative to some specific domain and code this in a computer in such a way that the knowledge of the expert is available to a less experienced user. KBS is a computer program that simulates the judgment and behavior of a human that has expert knowledge and experience in a particular field. It contains a knowledge base containing accumulated experience and a set of rules. Expert system provides high quality experience, domain specific knowledge; apply heuristics, forward or backward reasoning, uncertainty and explanation capability. Rule based expert system contains knowledge base, Inference engine, knowledge acquisition, explanation facility and user interface. For knowledge representation techniques, forward and backward chaining rules are used. Expert systems are designed to emulate an expert in a specialized knowledge domain such as medicine or any other area of knowledge where there is a shortage of expert knowledge (Tripathi, 2011).

During the development of KBS, knowledge about the problem to be solved must be acquired because the most important ingredient in any expert system is the knowledge and the power of expert system resides in the specific, high-quality knowledge it contain about task domains. Knowledge can be acquired from different sources such as making interview with domain experts, document analysis, observation and others. Since tacit knowledge is personal and the knowledge expert may not tell all the knowledge she/he knows during interview, there is hidden knowledge about the problem. Actually the expert knowledge area is "fuzzy" or ambiguous in nature and contains a great deal of procedural knowledge, so the knowledge engineer must be an expert in the process of knowledge collection (Tripathi, 2011). To handle ambiguity and hidden knowledge during knowledge acquisition automatic knowledge discovering is important of which one of the ways is data mining. Data mining, is a more general knowledge discovery techniques, for extracting hidden and previously unknown knowledge from datasets. Data mining methods like

classification, association and clustering, are used to function on large volumes of data to discover hidden patterns and relationships helpful in decision making (Roche & Wang, 2014).
Furthermore, data mining role in problems solve can be grouped in two: prediction and knowledge discovery. For each of these problems it is indicated to use some associated methods. For prediction, classification or regression used, while for knowledge discovery clustering, association rules, database segmentation, sequence analysis or visualization are used.

A classification rule attempts to predict the value of a discrete dependent variable from various known attributes. One of the most frequently used methods is classification based on decision tree. The decision tree can predict a new data instance, by following a path that starts from the root to a leaf node. One of the advantages of decision trees lies in the fact that they can easily translate into a set of 'IF –THEN' rules, easier to understand. Clustering, often referred to as unsupervised learning, involve a process that discovers structures in data without any supervision. As the name clustering implies, unsupervised algorithm is able to discover structures on its own, by exploiting similarities or differences between individual data points on a data set. Association rules mining is an important data mining method that aims to find interesting dependencies in large sets of data items. Interesting associations between data items can lead to information used for decision making (Nega, Adane, 2017).

Diabetes is one of the most chronic and rampant diseases. It's the problem of a metabolic condition of having higher than normal blood sugar levels. This is also known as hyperglycemia. Diabetes occurs when insulin is not being properly produced or responded by the body, which is essentially needed to maintain the proper level of sugar in the human body (Choubey, Paul, & Dhandhenia, 2017).

It is a group of metabolic diseases in which a person has high blood sugar, either because the body does not produce enough insulin, or because cells do not respond to the insulin that is produced a case known as insulin resistance. This high blood sugar produces the classical symptoms of polyuria, polydipsia and polyphagia. The symptoms of high blood sugar include frequent urination, extreme thirst, and increased hunger. If diabetes left untreated, it can cause many complications. Acute complications include diabetic ketoacidosis, non ketotic hyperosmolar coma, or death. Serious long-term complications include heart disease, stroke, chronic kidney failure, foot ulcers, and damage to the eyes.

Pre-diabetes is the pre-cursor of diabetes where the blood glucose levels are higher than normal but not high enough to be considered as diabetes. The prevalence of pre-diabetes is about 25% between population groups older than 45 years. About 70% of pre-diabetic patients are more likely to develop Type II diabetes and cardiovascular disease within 10 years. However, if the condition is tackled at this stage through diet, exercise and other healthy lifestyle changes (weight management program), the risk can be significantly reduced (Pillay, Maunder, & Naidoo, 2009).

The type-I diabetes affects young people below 20 years of age. In type I the pancreatic cells will get affected and fail to function. Because of nil secretion of insulin, the type-I diabetic people suffer throughout their life depend on insulin injection. The type I diabetic patients should regularly follow exercises and healthy diet as suggested by dietitians. Type 2 Diabetes Mellitus (DM) begins with insulin resistance, a condition in which cells fail to respond to insulin properly. As the disease progresses a lack of insulin may also develop. The most common cause is excessive body weight and lack of enough exercise. If the glucose level is not reduced by the above methods then medicines can be prescribed. A 2014 report by National Diabetes Statistics shows 29.1 million people or 9.3% of the U.S. population have diabetes (Roche & Wang, 2014).

Gestational diabetes as one of the main occurs when pregnant women without a previous history of diabetes develop high blood sugar levels. As of 2017 study of diabetes, it is found that around 18% of pregnant women have diabetes. Pregnancy during older age may have a risk of developing the gestational diabetes (Roche & Wang, 2014).

According to International Diabetes Federation (IDF) estimates, about 366 million people affected by type 2 diabetes in 2011 and by 2030 it may be increased to 552 million worldwide. Almost 80% of the diabetic people belong to middle and low income countries. According to American Diabetes Association (ADA), diabetes imposes a significant economic burden on the countries by healthcare expenditures, due to diabetes, account for 11% ($465 billion) of the total healthcare expenses in the world in 2011. By 2030, this number is projected to exceed $595 billion (Roche & Wang, 2014). This chronic disease need taking oral medications (pills), a controlled diet (changing eating habits), and physical exercise programs, but no comprehensive cure is available yet. However, the existing practices for medical treatment need patients to see specialist for diagnosis and treatment. As result, there might be lack of proper cure time, lack of experts, inaccuracy diagnosis and treatment, medical physicians may not have sufficient knowledge etc., in health centers of developing countries like Ethiopia. Thus, using KBS can be an encouraging good solution to reduce human expertise and medical error because as one of the specialized branches of AI, KBS is functioning in a specific domain to offer wise decisions with reasoning. The main benefits provided by such system are intelligent decisions, learning from experience, explanation and/or reasoning, and solving problems (Tripathi, 2011). In this study developing KBS for diagnosis and treatment of diabetes using data mining techniques is proposed.

1.2. STATEMENT OF THE PROBLEM

According to IDF statistics, there are 230 millions of diabetics through the world, at present time of which 80% of them are living in the developing countries. In 2025,the number of diabetics is estimated to reach 380 million (Kulani, 2012).

In 2015, the IDF also estimated that, in the Africa region, 14.2 million adults aged 20–79 years had diabetes, representing a prevalence of 3.2%. The majority (59%) of people with diabetes live in cities, even though the population is predominantly (61%) rural. This region has also the highest proportion of previously undiagnosed diabetes; over two-thirds (67%) of people with diabetes being unaware they have the disease.

Currently, Ethiopia has been challenged by the growing magnitude of non-communicable diseases such as diabetes. Ethiopia is among the top four countries with the highest adult diabetic populations in sub-Saharan Africa. Patient attendance rates and medical admissions related to diabetes in major hospitals have been rising. This requires a shift in healthcare priorities and up-to date data on the prevalence and related complications of diabetes in Ethiopia, to help plan and prioritized health programs. Such information can be an important base for policy on diabetes prevention and treatment (Gizaw et al., 2015).

Many people affected with diabetes is increasing because of population growth, unbalanced diet, aging, urbanization, overweight and lack of physical exercise. Furthermore, several studies indicate shortage numbers of skilled manpower in health care, which contributes to poor diagnosis and treatment of people living with diabetes in Ethiopia (Solomon, 2013).

Moreover, there are no sufficient numbers of experts and medical doctors this cause disproportional numbers of experts and patients. Poor diagnosis and treatment, might be physician inexpert on diabetes. Consequently, to minimize mortality and morbidity from diabetes compilations, intelligent system that effectively and efficiently enables diagnosis and treatment is essential. KBS can be considered as software, which operates on a sophisticated system like a human expert. It explains it’s reasoning or suggested decisions, display intelligent behavior, draw conclusions from complex relationships.

The main conceptual source of KBS is knowledge. KBS can expand to include a knowledge acquisition component the processes data and information into rules. KBS has number of application areas like decision making, prediction, planning, monitoring, process control, forecasting, diagnosis etc. Medical diagnosis is the major application of it. The purpose of medical expert system is to support the diagnosis process of experts. It considers facts and symptoms to provide diagnosis and treatments (Ambilwade Central, Ramesh Manza, 2014).

There are various studies that were conducted locally on knowledge-based systems in order to support reasoning and finding solutions for certain problems like in the area of Health Sciences, some of which are as follows.

A self-learning knowledge based system for diagnosis and treatment of diabetes(Solomon, 2013). Application of knowledge based system for woody plant species identification (Abetwe & A, 2009).Design and development of a prototype knowledge-based system for HIV pre-test counseling (Tagel, 2013).The potential for applying knowledge-based system for diagnosis of acute respiratory tract infections. However, greatest problem for the development of an effective and efficient system is to build an excellent knowledge base. In order to do that, knowledge acquisition is a sensible problem. Developing KBS is based on domain experts knowledge acquiring. Domain experts might not tell all important knowledge and naturally knowledge is ambiguous. As result, reasoning and decision making of the system developed became vague.

Therefore, automatic processing and exploring technology is important in order to solve such problems of KBS. One of modern achievements is data mining. As part of the complex process of knowledge discovery in databases, data mining tries to find useful patterns in large amounts of data, which have no obvious relationship between them. Such a pattern can be the key knowledge used for developing effective and efficient KBS. Due to this fact, the researcher was used automatically generated knowledge of rule based support to diagnosis and treatment in decision making.

Therefore, the main of this study was to design prototype KBS support in decision making for diagnosis and treatment of diabetes by using the knowledge obtained using data mining techniques. Thus, diabetic patients' baseline records were used as resources of data mining techniques in order to obtaining knowledge result.

To this end, the study attempts to answer for the following research questions:

1. What is the suitable system architecture for the proposed KBS?
2. How to identify and acquire appropriate knowledge?
3. What are the main attributes that can properly support for rule based and decision making?
4. Which classification algorithm is suggesting best algorism with more accurate values and less error rate diagnosis and treatment of diabetes?
5. How to develop effective and efficient prototype KBS using data mining techniques for diagnosis and treatment of diabetes?

1.3. **OBJECTIVES OF THE STUDY**

1.3.1. GENERAL OBJECTIVE

The general objective of this study is to design and develop prototype knowledge based system using data mining techniques for diagnosis and treatment of diabetes.

1.3.2. SPECIFIC OBJECTIVES

- ♦ To build the suitable system architecture for the prototype KBS to be developed
- ♦ To identify and acquire appropriate knowledge, both that is explicit and tacit
- ♦ To identify the main attributes that can properly support for rule based and decision making
- ♦ To decide appropriate classification algorithm which is best algorism with more accurate values and less error rate based on diabetic patients' datasets.
- ♦ To develop effective and efficient prototype KBS using data mining techniques for diagnosis and treatment of diabetes

1.4. **SCOPE AND LIMITATIONS OF THE STUDY**

This study is limited to develop prototype of knowledge based system using data mining techniques for diagnosis and treatment of diabetes. The researcher selected diabetes based on the assumption of its complex, fastest grow chronic illness requiring continuous medical care with multifactorial risk-reduction strategies beyond glycemic control. It need more attention when compared to the other chronic diseases. The knowledge for the KBS was acquired from domain experts' interview, documents analysis and diabetic patients' dataset by employing classification data mining techniques. Then a prototype KBS was developed used for diagnosis and treatments of diabetes was developed. However, the medical dataset nature, scarcity and nature of knowledge make the KBS development process difficult and complex which result the study not include dynamic and automatic knowledge discovery from database for develop dynamic rule based and system was not developed with local language.

1.5. SIGNIFICANCE OF STUDY

The intelligent system used for diabetes are important within the medical area because it allows doctors and nurses to quickly gather information and process it in various ways in order to assist with making diagnosis and treatment decisions. These systems could help in diverse areas from the storing and retrieval of medical records, storing and retrieval of key substances in medicines, examination of real-time data gathered from monitors, analysis of patient history for the purposes of diagnosis, analysis of family history, and in many other areas. The KBS applies intelligent reasoning to a domain to solve a problem that require considerable human time, effort and proficiency. KBS have valuable asset to any institutions as a substantial source to support decisions making (Tripathi, 2011).

The developed prototype system supports to reduce the problem of the limited numbers of experts in giving initial diagnosis and treatment of diabetes. In addition the system provide treatment advices effectively and efficiently based on diagnosis result. Furthermore system provides advices for diabetic patients as source of information. Specifically the system help for health care workers and experts working in the diagnosis and treatment of diabetes. Moreover, it can help as a benchmark for researchers who are interested in the area to further related research. The effective and efficient prototype KBS using data mining techniques for diagnosis of diabetes using data mining techniques was developed and used to support for decision making accurately and timely.

1.6. METHODOLOGY

1.6.1. RESEARCH DESIGN

Experimental research is a study that strictly adheres to a scientific research design. The primary goal of an experimental design is to establish a causal connection between the variables. A secondary goal is to extract the maximum amount of information with the minimum expenditure of resources. It is the process of planning a study to meet specified objectives. Planning an experiment properly is very important in order to ensure that the right type of data and a sufficient sample size and influence are available to answer the research questions of interest as clearly and efficiently as possible (Morrison, 2014).

Therefore, in this study the researcher used experimental research design method for model building, investigation, prototype development and testing whereas non-experimental method was used for interviews with experts for domain know how. The researcher used primary and secondary method to collect input to the system namely, knowledge. Data mining was one of the methods, i.e., to extract interesting and previously unknown information or patterns from data sources .Data mining used as the central point of knowledge discovery in databases (KDD) process and it correspond to the modeling step in the knowledge discovery in databases process and CRISP-data mining model as automatic knowledge acquisition method.

Since, manually gathered knowledge is fuzzy in nature, as result data mining techniques and algorisms were used as method of discovering unknown knowledge from preprocessed diabetic patients' medical datasets. Thus, the research design for this study was experimental which used to developed prototype KBS using data mining techniques for diagnosis and treatment of diabetes. The tools used for developing KBS is Clisp.net.

1.6.2. STUDY SITE AND POPULATION

Study sites for this research are Jimma University Specialized Hospital, Saint Paulos Millennium Medical Hospital and Adama Medical College Hospital. Jimma University Specialized Hospital, which is found in Oromia regional state, Jimma town in Jimma University main campus, Saint Paulos Millennium Medical Hospital which is found in Addis Abeba city administration and Adama medical college hospitals which found in Oromia regional state, Adama town. The total population interviewed were twelve. The datasets used for research was also collected from this each hospital.

1.6.3. SAMPLING TECHNIQUES

In this study purposive sampling technique was used for selection of hospitals and domain experts for knowledge acquisition from the above hospitals. The selection criteria of domain experts for the study is based on their profession and their immediate position and those doctors and would be suggested by the medical director of each hospital. Therefore, four experts in internal medicine were selected for interviewed from each hospital. The selected hospitals were the most known and experienced referral and medical hospitals services huge communities in our countries.

1.6.4. SOURCE OF DATA

There are two method of data collections which was used in this study. Those are primary and secondary. The primary method include the data gathered through questionnaire and interview primarily. Whereas secondary method includes referring different documents and journal articles.

For study primary and secondary data are used as source of information. The primary data was gathered by using interviews from domain experts and the secondary data was gathered from different written documents, conference articles and journal publications. Dataset was collected from baseline diabetic patients' medical record using secondary data collection method also known as retrospective method.

1.6.5. DATASETS

The data used for study was collected from the site study namely Jimma University Specialized Hospital, Saint Paulos Millennium Medical Hospital and Adama Medical college hospital. Dataset include diabetic patients medical data consists 2640 record (instances) and twelve attributes, value type, descriptions etc.

1.7. DATA COLLECTION METHOD

For primary data collection interview was used to collect domain knowledge (tacit) from the domain experts and by documents analysis explicit knowledge was extracted from secondary sources of data. For hidden knowledge to be discovered from database by automatic data mining (KDD and Crisp model method) techniques and dataset was collected from baseline diabetic patients' medical record using secondary data collection method also known as retrospective method.

1.7.1.1. ATTRIBUTES SELECTION

Medical data is often very high dimensional. Depending upon the use, some data dimensions might be more relevant than others. In processing medical data, choosing the optimal subset of features is such important, not only to reduce the processing cost but also to improve the usefulness of the model built from the selected data (Llorens et al., 2009). The goal of attribute subset selection was to find a minimum set of attributes such that the resulting possibility distribution of the data classes as close as possible to the original distribution obtained using all attributes. Mining on a reduced set of attributes has an additional benefit of reducing the number of attributes appearing in the discovered patterns, helping to make the patterns easier to understand. In this study, to select the attributes the researcher was discussing with domain experts.

1.8. TECHNIQUES AND ALGORISMS

Data mining process inputs predominantly cleaned, transformed data, searches the data using techniques and algorisms, and outputs patterns and relationships to the interpretation/ evaluation step of the KDD process (Pillay et al., 2009).

1.8.1. CLASSIFICATION TECHNIQUE

Classic data mining technique based on machine learning maps data into predefined groups or classes. Supervised learning is a technique where by the classes are determined before examining the data. Classification algorithms require that the classes be predefined based on data attribute values. It's often describe these classes by looking at the characteristics of data already known to belong to the classes (Ambilwade, 2014). A Rule-based classification extracts a set of rules that show relationships between attributes of the data set and the class label .For this study three algorithms namely J48 pruned, PART, and JRip was used.

1.8.1.1. CLASSIFIER ALGORISM

J48: Decision trees are mainly used in the classification and prediction. It is a simple and a powerful way of representing knowledge. Decision tree is built, many of the branches reflect anomalies in the training data due to noise or outliers. Tree pruning methods use statistical measures to remove the least reliable branches. A classifier system takes input from the cases described by values and attributes and output a classifier that can accurately predict classes of new cases. C 4.5 is a descendant of CLS and IDE, creates classifier and generated decision tree. It can also make classifier in most comprehensive rule-set forms (Abuhay & Tesema, 2015).

PART: rule-based classifier uses a set of IF-THEN rules for classification. An IF-THEN rule is an expression of the form IF condition THEN conclusion. The IF-part (or left-hand side) of a rule is known as the rule antecedent or precondition. The THEN-part (or right-hand side) is the rule consequent. Class for generating a decision list and uses separate-and-conquer. Builds a partial C4.5 decision tree in each iteration and makes the "best" leaf into a rule. A rule induction algorithm which grabs rule from a decision tree (Abuhay & Tesema, 2015).

JRip: is also a rule-based classifier uses a set of IF-THEN rules for classification and this experiment conducted with default parameters of WEKA and the algorithm generates a model with rules and identify Correctly Classified Instances and Incorrectly Classified Instances (Abuhay & Tesema, 2015).

1.9. **EXPERIMENT TOOL**

1.9.1. WEKA

WEKA has several graphical user interfaces that enable it easy access to primary functionality. Therefore data mining techniques, algorisms and tools was applied on the patients' medical datasets for discovering rules used for develop the prototype KBS. The algorithm can either be applied directly to dataset or called from Java code. It contains tools for data pre-processing, classification, regression, clustering, association rules and visualization. A set of data items, the dataset, is a very basic concept of machine learning (Witten, 2011). A dataset is roughly equivalent to a two-dimensional spreadsheet or database table. In WEKA, it is implemented by the Weka. Core. Instances class. A dataset is a collection of examples, each one of class Weka. Core. Instance. Each Instance consists of a number of attributes, any of which can be nominal (one of a predefined list of values), numeric (a real or integer number) or a string (an arbitrary long list of characters, enclosed in “double quotes”).Weka tool was used as experimental process environment for the study.

1.9.2. KBS DEVELOPING TOOL

KBS shell with the ready-to-wear utilities of self-learning, explanation and inference etc. like Quincy prolog, visual prolog, and Clips rule based, Java Expert System Shell (JESS), GURU, and Vidwan are more specific and can also be useful to develop KBS. KBS can be developed using programming languages like LISP and Prolog (Chala ,Million and T., 2016). Therefore for the study tool the used to develop prototype KBS was Clisp.net.

1.10. TESTING AND EVALUATION

The trend of library-based development approach used in building KBSs out from pre-defined components triggers our proposed approach for testing. The methodology is enriched with an integrated tool set to verify and validate the knowledge base that jointly satisfy the production of high quality knowledge bases (El-korany & Rafea, 1998).

KBSs developers are realizing that in order to compete in the commercial computing community a KBSs must display quality not only in its performance but also in its structure and developing process. Ensuring the quality of KBSs involves two types of activity: a) activities intended to assure that the KBS is structurally correct (verification activities) and b) activities intended to demonstrate KBS ability to reach correct conclusions (validation activities) (El-korany & Rafea, 1998).Testing and Evaluation was accomplished and the implication was involved to trust on its work and other features of the KBS. Therefore, in this testing and evaluation applied on prototyped system using case based (i.e. validity test) and user acceptance.

1.1.1. ETHICAL CONSIDERATION

The study should benefit and cause no harm to the participants and society. Privacy and confidentiality is maintained at all times, the information/data was portrayed in a confidential manner as no personal or identifiable information was recorded or printed in the study. No name was recorded during the interviewing process. The data was not transferred to any third party.

1.1.12. ORGANIZATION OF THESIS

This study comprises six chapters. Chapter one discusses background of the study, the problem statement and research questions, the general and the specific objectives of the study, and methodologies that the researcher used to conduct this study.

Chapter two discusses about conceptual and related works review that are relevant for this study. In this chapter, the researcher discussed about History of diabetes, Type of diabetes, symptoms of diabetes, Cause, Risk factors and complication of diabetes. In addition to this, the researcher discussed about knowledge engineering including problems method, Data Mining and knowledge discovery concepts, DM models and related works which are relevant for this study including Crisp and KDD model. The researcher presented the Knowledge discovery steps such as data set preparation, preprocessing, predictive model creation and experimentation. The researcher also discussed about Artificial Intelligence, KBS and components of KBS. In addition related work which is relevant to the study included in this chapter.

Chapter three discussed about Over view of knowledge based system, knowledge acquisition, knowledge structuring, knowledge about diabetes, common types of diabetes, symptoms of diabetes, risk factor of diabetes, criteria for diagnosis and treatment of Diabetes,Tradional and automatic knowledge acquisition.

Chapter four discusses about Design and Implementation. In this chapter the design and implementation of the prototype are realized by using Data mining results as a knowledge source. The architecture of the new prototype KBS used for diagnosis and treatment of diabetes is developed. The implementation tool used is Clisp.Net prolog integration with visual basic studio 2013 (vb.net).

Chapter five discusses about implementation testing and evaluation of the proposed systems. 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. In addition, discussion was made to show the significance of the proposed approach with previous researches. Finally, the researcher dedicated chapter six for conclusion and recommendation. In this chapter, the researcher discussed the evaluation results and based on the result the researcher presents findings and concludes the study by recommending future works.

CHAPTER TWO

LITERATURE REVIEW

2.1. SYSTEM

System is a regularly interacting or interdependent group of units forming an integrated whole. Each system is delineated by its spatial and temporal boundaries, surrounded and influenced by its environment, described by its structure and purpose and expressed in its functioning. Information systems (IS) involve a variety of in-formation technologies (IT) such as computers, soft-ware, databases, communication systems, the Inter-net, mobile devices and much more, to perform specific tasks, interact with and inform various actors in different organizational or social contexts.

It is also an integrated set of components for collecting, storing, and processing data and for providing information, knowledge, and digital products. Business firms and other organizations rely on information systems to carry out and manage their operations, interact with their customers and suppliers, and compete in the marketplace. Information systems are used to run inter organizational supply chains and electronic markets.

2.2. KNOWLEDGE BASED SYSTEM

A part of AI with availability of advanced computing facilities and other resources, attention is now turning to more and more demanding tasks, which might require intelligence. The society and industry are becoming knowledge oriented and rely on different experts' decision-making ability. KBS can act as an expert on demand without wasting time, anytime and anywhere. KBS can save money by leveraging expert, allowing users to function at higher level and promoting consistency. One may consider the KBS as productive tool, having knowledge of more than one expert for long period of time. In fact, it's a computer based system, which uses and generates knowledge from data, information and knowledge. These systems are capable of understanding the information under process and can take decision based on the residing information/knowledge in the system whereas the traditional computer systems do not know or understand the data/information they process (Sajja & Akerkar, 2010).

Now days, some of the behaviors such as problem solving, learning and understanding are handled by computer programs. Knowledge base system is a computer program that can solve and simplify the problems that is encountered in human expert by using knowledge about the application domain and problem solving techniques. Human experts use their knowledge about the domain and techniques that lead how to use the knowledge to solve problems. Computer knowledge base systems handle problems in the same way as humans can do. The system represents knowledge about a specific application domain and uses one or more techniques that guide to use knowledge to solve problems. Knowledge base system is the general term used for the process of eliciting, structuring and representing knowledge from some knowledge source mostly from human experts and developing a computational problem solving model, specifically a program to be used in some consultative or advisory role (Dagnachew, 2016).

There are many reasons for building an expert system to solve health related problems. Human experts may not always be available or may even be absent from a location. Also, by pooling knowledge of many experts, an expert system may be better than one human expert in its overall performance. An expert system does not get tired and are expected to be more consistent. It can also be used for training and passing on the knowledge derived from the human experts (Tunmibi et al., 2013).

2.2.1. ROLE OF KNOWLEDGE BASED SYSTEM IN HEALTH CARE

Industries and societies are becoming knowledge oriented and dependent of decision making ability of expert. Knowledge base system can act as an expert on response; can save money by leveraging experts; allowing users to function at higher level and promoting availability and consistency (Abbod, 2001). It can increase productivity, document when there is shortage of knowledge for future use, enhances problem solving capability and this leads to increase quality in problem solving process(Dagnachew, 2016).

Knowledge based systems are more useful than the old-fashioned computer based information systems when: There is shortage of experts; expertise is to be multiplied and stored for future use; there is more group of platform than one experts' knowledge and Intelligent assistance is important for decision making (Sajja & Akerkar, 2010).However, the scarcity and nature of knowledge make the knowledge base system development process difficult and complex. Some of the limitations of knowledge base system are due to some reasons (Abbod, 2001).Large volume of Knowledge Acquisition, representation and manipulation; limitations of cognitive science and other scientific methods; abstract nature of the knowledge. Knowledge base systems are systems which are capable of offering solutions to specific problems in a given domain or which are able to give advice, both in a way and at a level comparable to that of experts in the field. The problems in the fields for which knowledge based systems are being developed are those that require considerable human expertise for their solution. Examples of such problem domains are medical diagnosis of disease, financial advice, products design, etc. Most of the present day knowledge based systems are only capable of dealing with restricted problem areas. Nevertheless, even in highly restricted domains, knowledge base systems usually need large amounts of knowledge to arrive at a performance comparable to that of human experts in the field (Dagnachew, 2016).

Medical expert and knowledge-based systems are designed to give expert-level, problem-specific advice in the areas of medical data interpretation, patient monitoring, disease diagnosis, treatment selection, prognosis, and patient management. They capture and make available the knowledge of experts and by applying that knowledge to patient data emulate and assist in the decision-making behavior of medical and administrative personnel. Research in medical expert and knowledge-based systems and the development of such systems is most significant to the broad realm of quality assurance and cost containment in medicine. The growing complexity of the fund of knowledge makes the application of such systems more and more indispensable. Provided that they are used correctly, these systems can reduce much of the repetitive and specialized mental efforts made by the treating physician and enable him or her to devote his or her attention to the personal care of the patient (Adlassnig, 2001).

Medical knowledge is processed by the computer systems on the basis of stored medical knowledge and the current medical and administrative data of a patient, the systems provide a range of alternative suggestions for the course of patient care. The purpose of these decision-oriented suggestions is as follows: To ensure medical quality and to possibly improve patient care; to provide comprehensive quality management with consideration of medical working processes and administrative conditions; to ensure the efficient and cost-oriented utilization of available medical, technical, personnel, and organizational resources. The results of these research activities have impacted a large number of computer applications in medicine:

Clinical patient management: helping monitor patients measured and derived medical data and generate reminders, warnings, and alerts during the automatic processing of medical protocols and guidelines.

Laboratory medicine: providing knowledge-based interpretive reports of laboratory test results and having alerting modules check for notifiable, noteworthy, contradictory, or otherwise remarkable laboratory data.

Anesthesia and intensive care**:** building monitoring systems for disease prevention and early detection of diseases, observing entry criteria for therapies, and building knowledge-based adaptive control systems for medical devices.

Internal medicine: providing knowledge-based filtering, abstraction, and aggregation of medical data considering their context-dependency and temporal course, offering broadly applicable consultation systems for differential diagnosis and therapy to the caring physician in difficult cases, and providing electronic test tools for rare syndromes and rare pathological constellations with knowledge-based searching routines.

Image generating and processing medicine: introducing systems for knowledge-based navigation and monitoring of diagnostic and surgical procedures including routines to avoid undesired events or anatomical regions, displaying differential diagnostic support during image interpretation, and offering clinical data of patients with preceding knowledge-based filtering to assist the image-diagnosing physician in his or her decision.

In building medical expert and knowledge-based systems, special emphasis is put on the utilization of fuzzy set theory and fuzzy logic as methodology underlying the chosen patient data and medical knowledge representation and inference procedures. These methodologies have a number of characteristics that make them highly suitable for modeling uncertain information, which medical concept forming, patient state interpretation, and diagnostic as well as therapeutic decision making is usually based upon. First of all, medical entities such as symptoms, signs, test results, diseases and diagnoses, therapy proposals, and prognostic information items can be defined as fuzzy sets. The inherent vagueness of these entities will thus be conserved. Secondly, fuzzy logic offers reasoning methods capable of drawing strict as well as approximate conclusions. Medicine demands such a broad range of possibilities because the body of medical theory includes definitional, causal, statistical, and heuristic knowledge. Practical medicine even has to accept incomplete medical theories where only vague and uncertain empirical information guides the medical decisions and the diagnostic and therapeutic procedures they are based upon. Finally, fuzzy automata can be used as high-level patient monitoring devices employing real time access to the various medical information systems, such as hospital information systems (HIS), laboratory information systems (LIS), patient data management systems (PDMS), and others (Adlassnig, 2001).

The KBS consists of a Knowledge Base and a search program which mean Inference Engine (IE).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. As an expert's power lies in his explanation and reasoning capabilities, the expert system's creditability also depends on the explanation and reasoning of the decision made/suggested by the system. Also, human beings have an ability to learn new things and forget the unused knowledge from their minds. Simulation of such learning is essential component of KBS. The life of KBS may vary according to the degree of such simulation. KBS may be either manually updated (manual update) or automatically updated by machine (machine learning). Ideally, the basic frame of a KBS rarely needs to be modified. In addition to all these, there should be an appropriate User Interface, which may have the Natural Language Processing facility.

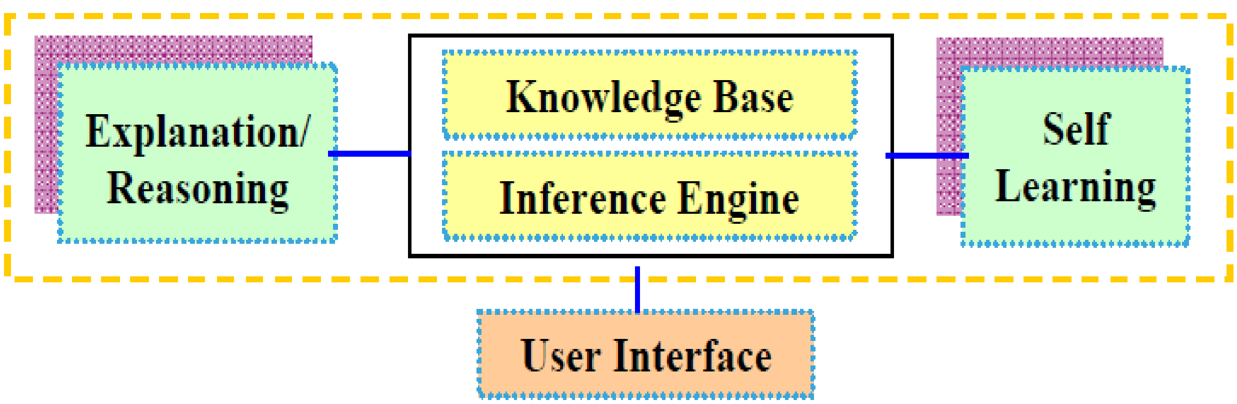

Figure-2.1: Architecture of a Knowledge-Based System (taken from Sajja & Akerkar, 2010).
There are mainly five types of the KBS: i) Expert Systems, (ii) Hypertext Manipulation Systems, (iii) CASE Based Systems, (iv) Database in conjunction with an Intelligent User Interface and (v) Intelligent Tutoring Systems (Sajja & Akerkar, 2010).

2.2.2. KBS DEVELOPMENT

The knowledge of the expert(s) is stored in his mind in a very abstract way. Also every expert may not be familiar with knowledge-based systems terminology and the way to develop an intelligent system. The Knowledge Engineer (KE) is responsible person to acquire, transfer and represent the experts' knowledge in form of computer system. People, Experts, Teachers, Students and Testers are the main users' groups of knowledge based systems.

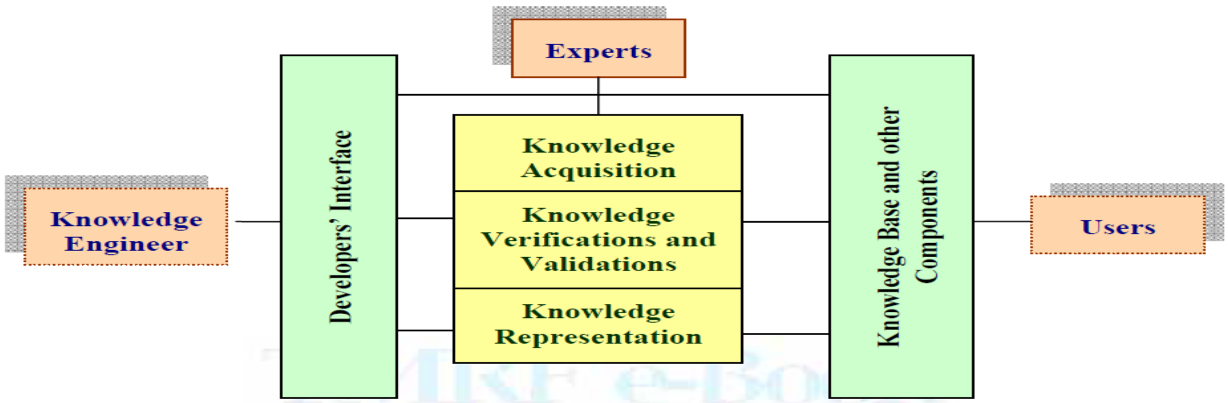

Figure-2.2: Development of a Knowledge-Based System (taken from Sajja & Akerkar, 2010).

The knowledge gaining process incorporates typical fact finding methods like interviews, questionnaires, record reviews and observation to acquire factual and explicit knowledge. However, these methods are not much effective to extract tacit knowledge which is stored in subconscious mind of experts and reflected in the mental models, insights, values, and actions of the experts.

For this, techniques like concept sorting, concept mapping, and protocol analysis are being used. The acquired knowledge should be immediately documented in a knowledge representation scheme. At this initial stage, the selected knowledge representation strategy might not be permanent. However documented knowledge will lead the knowledge engineer/developer to better understanding of the system and provides guidelines to proceed further. Rules, frames, scripts and semantic network are the typical examples of knowledge representation scheme (Sajja & Akerkar, 2010).

It is the responsibility of the knowledge engineer to select appropriate knowledge presentation scheme that is natural, efficient, transparent, and developer friendly. One may think for hybrid knowledge representation strategies like rules within the frames in slots like "on need" and "on request"; semantic network of default frames etc. (Sajja & Akerkar, 2010).

2.2.3. KBS TOOLS

A set of software instructions and utilities taken to be a software package designed to assist the development of knowledge-based systems. Personal computers, typical programming languages like java and framework like .NET can also be used in KBS development. These programming languages are general purpose and also being used to develop other application than AI applications (Choubey et al., 2017).

KBS shell with the ready-to-wear utilities of self-learning, explanation and inference etc. like Java Expert System Shell (JESS), GURU, and Vidwan are more precise and can also be useful to develop KBS. Adapt made KBS can be developed using programming languages like LISP and Prolog. As of 1960 published a remarkable paper showing a handful of simple operators and a notation for functions, one can build a whole programming language. John McCarthy called this language Lisp, for "List Processing," because one of his key ideas was to use a simple data structure called a list for both code and data. There are various versions of Lisp available namely KLISP and C Language Integrated Production System (CLIPS) (Chala ,Million and T., 2016).

2.2.4. KBS FOR INTEGRATED DEVELOPMENT

The Knowledge based systems can be used for interpretation, prediction, diagnosis, design, planning, monitoring, debugging, repair, instruction, control, etc. Such advanced technology should be made available in urban and rural areas to utilize expert knowledge for holistic development. Such systems export knowledge in underdeveloped and remote area where expertise is rare and costly. Hence, knowledge-based systems KBS should be at the primary consideration while designing the development plan for a nation. The share of AI/KBS systems in IT is improved significantly (Sajja & Akerkar, 2010).

In addition, today's KBS are easier to use, less expensive and integrate well with traditional technologies, so it can provide a fundamental technology to the majority of the applications for today's scenario. The four major dimensions of the rural development process namely; Economical, Social, Physical and Health development are considered for the holistic development. Major resources for development are considered as natural resources, human resources, livestock and agricultural resources (Sajja & Akerkar, 2010).

2.3. DATA MINING TECHNIQUES AND MODELS

Data mining uses a combination of explicit knowledge base, sophisticated analytical skills, and domain knowledge to uncover hidden trends and patterns. These trends and patterns form the basis of predictive models that enable analysts to produce new observations from existing data. The process of discovering meaningful new correlations, patterns, and trends by sifting through large amounts of data stored in repositories, and by using pattern recognition technologies, as well as statistical and mathematical techniques. Data mining should be performed on very large or raw datasets using either supervised or unsupervised data mining algorithms (Han and Kamber, 2006).

It has been called exploratory data analysis, among other things. Masses of data generated from cash registers, from scanning, from topic specific databases throughout the company, are explored, analyzed, reduced, and reused. Searches are performed across different models proposed for predicting sales, marketing response, and profit. Classical statistical approaches are fundamental to data mining. Automated AI methods are also used. However, systematic exploration through classical statistical methods is still the basis of data mining. Some of the tools developed by the field of statistical analysis are harnessed through automatic control (with some key human guidance) in dealing with data. A variety of analytic computer models have been used in data mining. The standard model types in data mining include regression (normal regression for prediction, logistic regression for classification), neural networks, and decision trees (Han and Kamber, 2006).

2.3.1. THE ROLE OF DATA MINING IN HEALTH SECTOR

Health organizations today are capable of generating and collecting a large amount of data. This increase in data volume automatically requires the data to be retrieved when needed. With the use of data mining techniques is possible to extract the knowledge and determine interesting and useful patterns or knowledge. The knowledge gained in this way can be used in the proper order to improve work efficiency and enhance the quality of decision making. Above the foregoing is a great need for new generation of theories and computational tools to help people with extracting useful information from the growing volume of digital data (Taranu, 2015).

Information technologies are implemented increasingly often in healthcare organizations to meet the needs of physicians in their daily decision making. Computer systems used in data mining can be very useful to control human limitations such as subjectivity and error due to fatigue and to provide guidance to decision-making processes (Sannita et al., 2011). The essence of data mining is to identify relationships, patterns and models which support predictions and decision-making process for diagnosis and treatment planning. These can be called predictive models, and integrated in hospitals information systems as models of decision making, reduce subjectivity and the necessity for reducing the time for decision making. In addition, the use of information technology in healthcare enables the comprehensive management of medical knowledge and its secure exchange between healthcare providers and beneficiaries (AlSharif, 2011).

Obtaining value added information using computers can help the quality of decision-making and avoiding human error. When there is a large volume of data that must be processed by the people, making decisions is generally of poor quality (Eapen, 2004). Data mining as process of analyzing the raw data using a computer and extracts their meaning. The process is often defined as the discovery of previously unknown and potentially useful information from large volumes of data (unstructured) (Milovic, 2011).Thanks to this technique, it is possible to predict trends and behavior of patients or diseases. This is done by analyzing data from different perspectives and finding connections and relationships between seemingly unrelated information. In the process of data mining previously unknown trends and patterns from a database of information are discovered and transform information into meaningful solutions (Taranu, 2015).

Healthcare abounds various information which causes the necessity of data mining application. It is well known that healthcare is a complex area where new knowledge is being accumulated daily in a growing rate. Big part of this knowledge is in the form of paperwork, resulting from a studies conducted on data and information collected from the patient's healthcare records. There is a big tendency today to make this information available in electronic form, converting information to knowledge, which is not an easy thing to do (Ceusters, 2001). All healthcare institutions need an expert analysis of their medical data, project that is time consuming and expensive (Taranu, 2015). The ability to use a data in databases in order to extract useful information for quality health care is a key of success of healthcare institutions (Eapen, 2004). In medical research, data mining begins with the hypothesis and results are adjusted accordingly, different from standard data mining practice that begins with a set of data without obvious hypothesis (Canlas, 2009). While the traditional data mining is focused on patterns and trends in data sets, data mining in healthcare is more focused on minority that is not in accordance with patterns and trends. The fact that standard data mining is more focused on describing and not explaining the patterns and trends, is the one thing that deepens the difference between standard and healthcare data mining. Healthcare needs these explanations since the small difference can stand between life and death of a patient (Taranu, 2015).

Prevention and diagnosis: Data mining technique made prediction system plays a vital role in strategy preparation for prevention of communicable as well as non-communicable diseases in located area. Lifestyle related diseases like hypertension, diabetes mellitus, cardiovascular diseases; stroke etc. can be easily and accurately classified and possible to locate their etiological area cluster patterns. These techniques are also useful in disease diagnosis. For instance prototype Intelligent Heart Disease Prediction System using three data mining modelling techniques, namely, Decision Trees, Naïve Bayes and Neural Network. Intelligent Heart Disease Prediction System can discover and extract hidden knowledge (patterns and relationships) associated with heart disease from a historical heart disease database. It can answer complex queries for diagnosing heart disease and thus assist healthcare practitioners to make intelligent clinical decisions which traditional decision support systems cannot. By providing effective treatments, it also helps to reduce treatment costs. To enhance visualization and ease of interpretation, it displays the results both in tabular and graphical forms (Tayade, 2014).

Information system (i.e. KBS) simplifies and automates the workflow of health care institution. Integration of data mining in information systems, healthcare institutions reduce subjectivity in decision-making and provide a new useful medical knowledge. Predictive models provide the best knowledge support and experience to healthcare workers. The goal of predictive data mining in medicine is to develop a predictive model that is clear, gives reliable predictions, support doctors to improve their prognosis, diagnosis and treatment planning procedures. A very important application of data mining is for biomedical signal processing expressed by internal regulations and responses to the stimulus conditions, whenever there is a lack of detailed knowledge about the interactions between different subsystems, and when the standard analysis techniques are ineffective, as it is often the case with nonlinear associations (Muller, 2000).

2.3.2. CHALLENGES OF USING DATA MINING IN HEALTH SECTOR

One of the biggest challenges in data mining in medicine is that the raw medical data is voluminous and heterogeneous. These data can be gathered from various sources such as from conversations with patients, laboratory results, review and interpretation of doctors. All these components can have a major impact on diagnosis, prognosis and treatment of the patient, and should not be ignored. Missing, incorrect, inconsistent or nonstandard data such as pieces of information saved in different formats from different data sources create a major obstacle to successful data mining. Also, another obstacle is that almost all diagnoses and treatments in medicine are imprecise and subjected to error rates.

Here the analysis of specificity and sensitivity are being used for the measurement of these errors. Within the issue of knowledge integrity assessment, two biggest challenges are: (1) How to develop efficient algorithms for comparing content of two knowledge versions (before and after). This challenge demands development of efficient algorithms and data structures for evaluation of knowledge integrity in the data set; and (2) How to develop algorithms for evaluating the influence of particular data modifications on statistical importance of individual patterns that are collected with the help of common classes of data mining algorithm. Algorithms that measure the influence that modifications of data values have on discovered statistical importance of patterns are being developed, although it would be impossible to develop a universal measure for all data mining algorithms (Taranu, 2015).

2.3.3. KNOWLEDGE DISCOVERY IN DATABASES

Knowledge Discovery in Databases (KDD) is an automatic, exploratory analysis and modeling of large data repositories. KDD is the organized process of identifying valid, novel, useful, and understandable patterns from large and complex data sets. Data Mining is the core of the KDD process, involving the inferring of algorithms that explore the data, develop the model and discover previously unknown patterns. The model is used for understanding phenomena from the data, analysis and prediction. The accessibility and abundance of data today makes knowledge discovery and Data Mining a matter of considerable importance and necessity (Han and Kamber, 2006).

2.3.3.1. THE KDD PROCESS MODEL

This process includes several stages, consisting of data selection, data treatment, data pre-processing, data mining and interpretation of the results. This process is interactive, since there are many decisions that must be taken by the decision-maker during the process. The steps of data mining (Han and Kamber, 2006) are presented in figure 2.1 and described as follows.

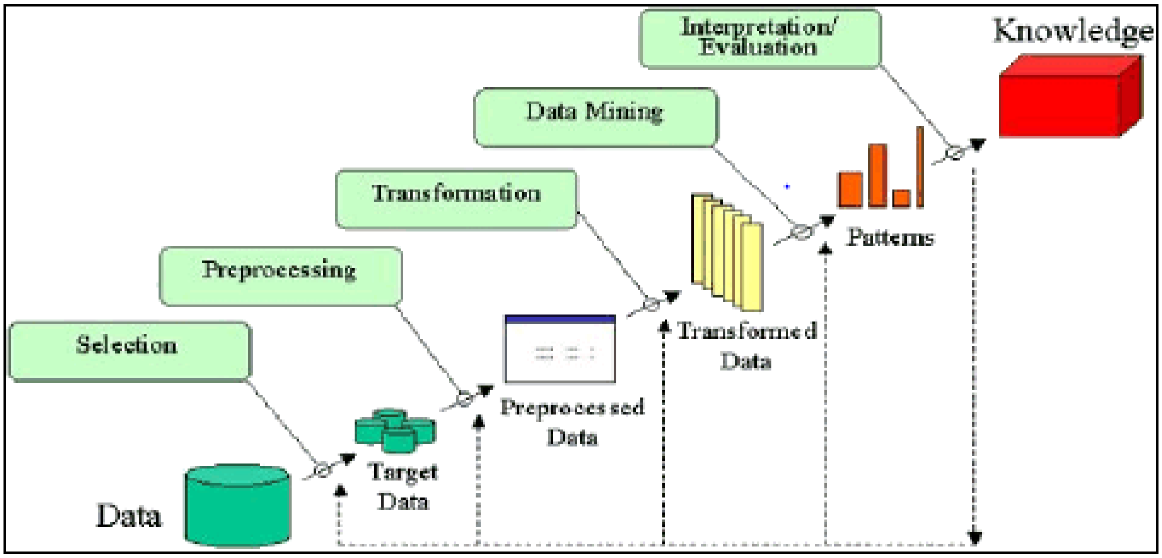

Figure-2.3: the process of knowledge discovery (Han and Kamber, 2006).

Data selection: This stage includes the study of the application domain, and the selection of the data. The domains study intends to contextualize the project in the company's operations, by understanding the business language and defining the goals of the project. In this stage, it is necessary to evaluate the minimum subset of data to be selected, the relevant attributes and the appropriate period of time to consider.

Data pre-processing: This stage includes basic operations, such as: removing noise or outliers, collecting the necessary information to model or account for noise, deciding on strategies for handling missing data attributes, and accounting for time sequence information and known changes. This stage also includes issues regarding the database management system, such as data types, schema, and mapping of missing and unknown values.

Data transformation: This stage consists of processing the data, in order to convert the data in the appropriate formats for applying data mining algorithms. The most common transformations are: data normalization, data aggregation and data discretization. To normalize the data, each value is subtracted the mean and divided by the standard deviation. Some algorithms only deal with quantitative or qualitative data. Therefore, it may be necessary to discredit the data, i.e. map qualitative data to quantitative data, or map quantitative data to qualitative data.

Data mining stage: This stage consists of discovering patterns in a dataset previously prepared. Several algorithms are evaluated in order to identify the most appropriate for a specific task. The selected one is then applied to the pertinent data, in order to find indirect relationships or other interesting patterns.

Interpretation/Evaluation: This stage consists of interpreting the discovered patterns and evaluating their utility and importance with respect to the application domain. In this stage it can be concluded that some relevant attributes were ignored in the analysis, thus suggesting the need to replicate the process with an updated set of attributes.

2.3.4. CRISP-DM PROCESS MODEL

The CRISP-DM (CRoss-Industry Standard Process for Data Mining) contains the corresponding phases of a project, their respective tasks, and relationships between these tasks. At this description level, it is not possible to identify all relationships. Essentially, there possibly exist relationships between all data mining tasks depending on the goals, the background and interest of the user, and most importantly on the data. The steps of Crisp data mining (Ncr et al., 1999).are presented in figure 2.1 and described as follows.

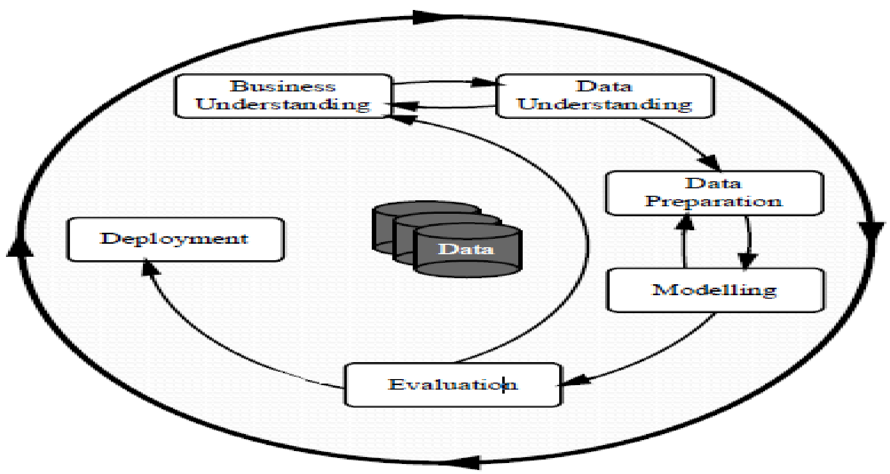

Figure-2.4: the Crisp-DM KDD Process Model (taken from Ncr et al., 1999).

Business Understanding

This initial phase focuses on understanding the project objectives and requirements from a business perspective, and then converting this knowledge into a data mining problem definition, and a preliminary plan designed to achieve the objectives.

Data Understanding

The data understanding phase starts with an initial data collection and proceeds with activities in order to get familiar with the data, to identify data quality problems, to discover first insights into the data, or to detect interesting subsets to form hypotheses for hidden information.`

Data Preparation

The data preparation phase covers all activities to construct the final dataset (data that will be fed into the modeling tool(s)) from the initial raw data. Data preparation tasks are likely to be performed multiple times, and not in any prescribed order. Tasks include table, record, and attribute selection as well as transformation and cleaning of data for modeling tools.

Modeling

In this phase, various modeling techniques are selected and applied, and their parameters are calibrated to optimal values. Typically, there are several techniques for the same data mining problem type. Some techniques have specific requirements on the form of data. Therefore, stepping back to the data preparation phase is often needed.

Evaluation

At this stage built a model (or models) that appears to have high quality, from a data analysis perspective. Before proceeding to final deployment of the model, it is important to more thoroughly evaluate the model, and review the steps executed to construct the model, to be certain it properly achieves the business objectives. A key objective is to determine if there is some important business issue that has not been sufficiently considered. At the end of this phase, a decision on the use of the data mining results should be reached.

Deployment

Creation of the model is generally not the end of the development. Even if the purpose of the model is to increase knowledge of the data, the knowledge gained will need to be organized and presented in a way that the customer can use it. Depending on the requirements, the deployment phase can be as simple as generating a report or as complex as implementing a repeatable data mining process. In many cases it will be the customer, not the data analyst, who will carry out the deployment steps.

2.3.5. DATA MINING TECHNIQUES

Data mining techniques are used to operate on large volumes of data to discover hidden patterns and relationships helpful in decision making. Various algorithms and techniques like Classification, Clustering, Association Rules, Regression, Artificial Intelligence, Neural Networks, Decision Trees, Genetic Algorithm, Nearest Neighbor method etc., are used for knowledge discovery from databases. These techniques and methods in data mining need brief mention to have better understanding (Han and Kamber, 2006).

2.3.5.1. CLASSIFICATION

The most commonly applied data mining technique, which services a set of pre-classified examples to develop a model that can classify the population of records at large. This approach frequently employs decision tree or neural network-based classification algorithms. The data classification process involves learning and classification. In Learning the training data are analyzed by classification algorithm. In classification test data are used to estimate the accuracy of the classification rules. If the accuracy is acceptable the rules can be applied to the new data tuples. The classifier-training algorithm uses these pre-classified examples to determine the set of parameters required for proper discrimination. The algorithm then encodes these parameters into a model called a classifier (Ambilwade Central, Ramesh Manza, 2014).

Neural networks: A technique that classifies each record in a dataset based on a combination of the classes of the k record(s) most similar to it in a historical dataset (where k is greater than or equal to 1).

Decision Trees: Decision tree is tree-shaped structures that represent sets of decisions. These decisions generate rules for the classification of a dataset (Ambilwade Central, Ramesh Manza, 2014).

2.3.5.2. CLUSTERING

Clustering is finding groups of objects such that the objects in one group will be similar to one another and different from the objects in another group. In educational data mining, clustering has been used to group students according to their behavior (Ambilwade Central, Ramesh Manza, 2014).

2.3.5.3. ASSOCIATION RULE

Association analysis is the discovery of association rules showing attribute-value conditions that occur frequently together in a given set of data (Ambilwade Central, Ramesh Manza, 2014).Association and correlation is usually to find frequent item set findings among large data sets. This type of finding helps businesses to make certain decisions, such as catalogue design, behaviors analysis, cross marketing and customer shopping behavior analysis. Association Rule algorithms need to be able to generate rules with confidence values less than one. However the number of possible Association Rules for a given dataset is generally very large and a high proportion of the rules are usually of little (if any) value (Ambilwade Central, Ramesh Manza, 2014).

2.4. DIABETES

Diabetes Mellitus (DM) is a set of related diseases in which the body cannot regulate the amount of sugar in the blood. In a healthy person, the blood glucose level is regulated by several hormones, including insulin. Insulin is produced by the pancreas, a small organ between the stomach and liver. The pancreas secretes other important enzymes that help to digest food. Insulin allows glucose to move from the blood into liver, muscle, and fat cells, where it is used for fuel (Lingaraj et al., 2015).

Diabetes is a long-lasting disease in which the human body's cells either do not respond properly to insulin or insulin production becomes insufficient. In order to understand the concept of diabetes, it is important to know how our body cells use glucose to produce energy. As a matter of biological process, the food that someone eats is transformed into glucose and energy is generated from glucose. This energy makes the body cells to accomplish their tasks. During the process of energy production, pancreas (a gland organ in the digestive system) produces a hormone called insulin. Insulin enables the body cells to absorb glucose. Nevertheless, diabetic individuals either cannot produce this hormone properly or their body cannot appropriately use it for usual body functions. Consequently, this creates the accumulation of glucose in the blood and when the accumulated glucose is not used by the body cells, then signs of diabetes start to come out (Kulani,2012). There are three main types of diabetes mellitus (DM) or simply diabetes.

2.4.1. TYPES OF DIABETES

2.4.1.1. PRE DIABETES

Pre diabetes is a milder form of diabetes that is sometimes called impaired glucose tolerance. It can be diagnosed with a simple blood test (Choubey, Paul, & Dhandhenia, 2017).Prediabetes raises short-term absolute risk of type 2 diabetes by 3-to 10-fold, with some populations exhibiting greater risk than others. People with diabetes are vulnerable to multiple and complex medical complications. These complications involve both cardiovascular disease (CVD) (heart disease, stroke, and peripheral vascular disease) and microvascular disease (i.e., retinopathy, neuropathy, and micro albuminuria). Most patients with diabetes die of CVD (Garber et al., 2008).

2.4.1.2. TYPE I DIABETES

Type 1 diabetes mostly affects children and young adults but, can affect at any age. In general, 5-10% of diabetes patients suffer from type 1 diabetes. In this case, body is unable to produce insulin. Insulin helps the body to use sugar from food as a source of energy. People with type-1 diabetes need insulin therapy. There is a destruction of insulin secreting cells (β-cells) of pancreas in our body. The cause is immune-mediated or idiopathic. It usually requires external insulin therapy hence known as Insulin Dependent Diabetes Mellitus (IDDM) (Choubey, Paul, & Dhandhenia, 2017).

2.4.1.3. TYPE II DIABETES

This type of diabetes is the most common type of diabetes with 90-95% of all the diabetes patients suffering from it. Type II diabetes symptoms often develop slowly. This type mostly occurs in the people who are more than forty years old but can also be found in younger age group. Type II diabetes symptoms are usually the same for men and women however certain symptoms are gender based like urological problems such as Erectile Dysfunction (ED) etc. which is seen only in males. This is asymptomatic for many years. Body loses the ability to produce adequate insulin (Insulin deficiency) or the body cells do not respond to insulin (Insulin resistance) or both. It can be controlled but, lifestyle modification, taking oral medications (pills) and if required insulin therapy but no complete cure for this diabetes type is available (Choubey, Paul, & Dhandhenia, 2017).

2.4.1.4. GESTATIONAL DIABETES

Gestational diabetes occurs during pregnancy. It raises mother's risk of getting diabetes for the rest of her life. It also raises the child's risk of being overweight and getting diabetes. It displays a high blood sugar level during pregnancy, usually occurs at around 28 weeks or later, and affects about 4% of all pregnant women. This type of diabetes usually goes away after pregnancy and causes are yet unknown (Choubey, Paul, & Dhandhenia, 2017).Gestational diabetes, during pregnancy usually around the 24^{th} week many women develop gestational diabetes(Ahmed, Mahmoud, Aref, & Salem, 2012).A diabetic condition that appears during pregnancy and usually goes away after the birth of the baby. Gestational diabetes is best controlled by dietary adjustment. Gestational diabetes can cause birth complications. One complication is microsomal, in which the baby is considerably larger than normal due to large deposits of fat; such a baby can grow too large to be delivered through the vagina. Gestational diabetes also increases the risk of low blood sugar, low serum calcium and low serum magnesium in the baby immediately after delivery. The key to prevention is careful control of the mother's blood sugar levels. If the mother maintains normal blood sugar levels, it is less likely that the fetus will develop microsomal, hypoglycemia, or other chemical abnormalities.

2.4.2. CAUSES OF DIABETES

Diabetes is a disease that causes blood glucose level rise in human body. The normal blood glucose level lies between (70-100) mg/100 ml during fasting and approximately 140 mg/100 ml otherwise. For a diabetic person, the blood glucose is around 126 mg/100 ml during fasting and 200 mg/100 ml otherwise. The most common symptoms observed in diabetic patients are: polyuria, weight loss, excessive thirst, continuous hunger, blurring and changes in vision, and fatigue (Kulani, 2012).

2.4.3. RISK FACTORS OF DIABETES

The cause of diabetes type I is generally unknown but research has shown that genetic factors are important risk factors. Coupled with an abnormal immune response. Medical experts believe that Type II diabetes is caused by a combination of lifestyle and genetic factors. However, in some cases it can be caused by the following factors as much sedentary lifestyle, stress, obesity, advanced age, unhealthy diet, family history of diabetes, improper functioning of the pancreas, previously diagnosed with gestational diabetes (Llorens et al., 2009).

2.4.4. COMPLICATIONS OF DIABETES

Diabetes can cause severe complications in both short and long term. The short term complications are: hyperglycemia, hypoglycemia and diabetic ketoacidosis. Hyperglycemia is an effect of high blood sugar level. This condition occurs when a large amount of food is consumed and the body cannot use all the sugar. Frequent urination, excessive thirst and nausea are the most common warning signs of hyperglycemia. Hypoglycemia is a condition which results from low blood sugar levels. It can occur suddenly in people using insulin if too little food is eaten, if a meal is delayed or in the case of extreme exercise. Symptoms include feeling cold, nervous, weak or hungry, and some people usually have headaches. Diabetic ketoacidosis is a condition which develops from a severe shortage of insulin. This condition develops when insulin and blood sugar levels are out of balance such that ketones accumulate in the blood. Symptoms include high blood sugar or ketones in the urine, dry mouth, and loss of appetite, fruity-smelling breath and possible vomiting. Long term complications include: Blindness, Kidney failure, fungal diseases, cardiovascular diseases and Lower limb amputations (Kulani, 2012).

2.5. RELATED WORK

Intelligence deals with many cognitive skills, comprising the capability to find solution of problems, learn, and understand language. However, developments up to the present time in AI have been made in the area of problem solving concepts and methods for constructing program that reason about problems instead of finding the solution (Tagel, 2013).Knowledge-Based System Model to Support Diabetes Research and Clinical Process Among different applications of knowledge-based systems, it is possible to find applications such as diagnosis, planning, designing, monitoring and control of processes, or training and consultation services. Therefore expert systems that focus onto support organizational making decisions are equipped with components denominated knowledge-based management subsystems (Hernández-medrano, 2011).

The expert system can be used effectively in all areas of medical sciences. In particular, in terms of vast number of diabetics throughout the world, the expert system can be highly helpful for the patients. These patients in many cases are not aware of their disease and how to control it. In addition, some of these patients do not access to the physicians during necessary times. Therefore, such a system can provide necessary information about the indications, diagnosis and primary treatment advices to the diabetics. Since this expert system gathers its knowledge from several medical specialists, the system has a broader scope and can be more helpful to the patients in comparison to just one physician (Zeki et al., 2012).

The functionality of the Tele-healthcare solution is to interact with the patient via the internet through the user interface at the patient kiosk. The Tele-expert system will conduct a virtual consultation session with the user, through the user interface, to determine his/her current health profile. The virtual consultation session is a graphic user interface based dialogue between the user and he Tele-expert system application, during which the expert system forwards a number of questions to the user either to collect information/verify available information and/or derived conclusions. The user is expected to provide valid responses to the questions presented by the expert system, in a touch screen mode or through the customized keyboard (Beulah Devamalar,Thulasi Bai, 2008).

According to Alemu (2009) Application of Knowledge Based System for Woody Plant Species Identification. This study attempts to design prototype KBS for woody plant species identification. The knowledge based system uses rule based approach for the proposed system. The system is modeled in decision laddering; domain knowledge is represented using production rules in prolog to construct the knowledge base. The prolog built in backward inferring mechanism is used for the identification of the species. Finally, the system is tested and evaluated by the users. The result shows that, the system identifies the woody plant species correctly and can be applicable in woody plant species identification. As compared to existing way of identification we come up with new knowledge/rules with minimum features that registers comparable performance.

According to Solomon (2013) the developed a prototype of a self-learning knowledge-based system using KBS approach process and Swi-prolog tool editor that can provide advice for physicians and patients to facilitate the diagnosis and treatment of patients living with diabetes. The prototype system achieves a good performance and meets the objectives of the study. In order to make the system applicable in the domain area for diagnosis and treatment of diabetes, some adjustments like automatically updating the rules in the knowledge base of the system, incorporating a well-designed user interface and a mechanism of natural language facilities are future work.

According to Dagnachew (2016) the developed knowledge based system using design science research methodology and Mat-lab tool was used for premedical triage using fuzzy inference system, and the user acceptance of the system is evaluated by the domain experts. The system gives a significance contribution for the domain area because the domain experts rated the user acceptance test of the system as 74.28%. This is because the since the symptoms of the patients are subjective, it is difficult for fuzzing. In addition to that there is no universal agreement on fuzzing the symptoms. The developed system can be able to identify the diseases based on the query given by the user. The user in this case the triage nurse gives each fuzzified values of the symptoms, and then the system recommended the possible result in ranking by using percentage.

Knowledge based system for diagnosis and treatment of malaria was developed using Knowledge Engineering research design to develop prototype system. Purposive sampling technique was used to select domain experts for knowledge acquisition. The knowledge was acquired using both structured and unstructured interviews from domain experts and represented by production rule. Developing the system in local languages, improving the user interface and applying other techniques are the future works of the study. The system can be used for diagnosis and treatment of malaria where there is shortage of health professionals and/or lack of laboratory equipment and can learn new facts about signs and symptoms and treatments. During KBS development the challenges faced were acquiring knowledge from the domain experts, therefore further knowledge for diagnosis and treatment of malaria is needed to improve the knowledge base (Chala ,Million and T., 2016).

Hybrid intelligent system is a combination of artificial intelligence (AI) techniques that can be applied in healthcare to solve complex medical problems. Case-based reasoning (CBR) and rule based reasoning (RBR) are the two more popular AI techniques which can be easily combined. Both techniques deal with medical data and domain knowledge in diagnosing patient conditions. They proposes a hybrid intelligent system that uses data mining technique as a tool for knowledge acquisition process. Data Mining solves the knowledge acquisition problem of rule based reasoning by supplying extracted knowledge to rule based reasoning system. They use WEKA for model construction and evaluation, Java Net-Beans for integrating data mining results with rule based reasoning and Prolog for knowledge representation. To select the best model for Tuberculosis diagnosis. To select the best model for disease diagnosis, four experiments were carried out using J48, BFTree, JRIP and PART. The PART classification algorithm is selected as best classification algorithm and the rules generated from the PART The use of data mining techniques to build the knowledge base of the hybrid system can be taken as strong features of the system. However, the system lacks to update rules in the knowledge base of the hybrid system and the user interface need to be enhanced with a better graphical user interface that allows users to choose their language preferences. So, further study is needed to improve user interface of hybrid intelligent system and to design a system that can update rules of the knowledge base (Nega and Adane, 2017).

Generally, several studies have been developed using rule-based representation technique to reason out the solution of a particular problem. But, the developed KBSs were not used automatic knowledge extraction from datasets using data mining techniques and the system was not with smart user interface fundamentally .Thus, in this study an attempt is made to design a prototype KBS using data mining techniques for diagnosis and treatment of diabetes that is effective and efficient in decision making.

CHAPTER THREE

KNOWLEDGE ACQUISITION, MODELING AND KNOWLEDGE REPRESENTATION

3.1. OVER VIEW OF KNOWLEDGE BASED SYSTEM

KBS uses computational intelligence and machine learning techniques to acquire information from data, whereas a knowledge-based system has well defined information organization and manipulation schemes. A knowledge based system consists of three components: a modeling subsystem, processing or inference subsystem and a knowledge base. The central component of the knowledge-based system is the knowledge base (KB). The knowledge base is domain specific. The quality, consistency, integrity and efficient usage of the KB profoundly affect the performance of these systems. If health information technology is going to transform healthcare, a deeper understanding of the complex dynamics underlying the system implementation and application. Since poor clinical decisions may lead to adverse effects, there is a need for efficient and effective decision support systems (Christopher, Nehemiah, & Arputharaj, 2016).

3.2. KNOWLEDGE ACQUISITION

Knowledge acquisition (KA) is the process of acquiring relevant knowledge from domain experts and other sources of information such as books, databases, guidelines, manuals, journal articles, computer files, etc. KA is the process of eliciting, structuring and representing domain knowledge acquired from different sources. The knowledge acquirement component allows the expert to enter their knowledge or expertise into the expert system, and to refine it later as and when required. Historically, the knowledge engineer played a major role in this process, but automated systems that allow the expert to interact directly with the system are becoming increasingly common. The acquired knowledge can be specific to the problem domain, it can be general or it is meta-knowledge (knowledge about knowledge). Knowledge acquisition is the first step and time consuming task in the development of knowledge based system. There are certain important steps that the knowledge engineer needs to carry out during knowledge acquisition process (Dagnachew, 2016).

Knowledge acquisition is one of the major bottlenecks in the stage of knowledge based system development. Two primary approaches to knowledge acquisition are elicitation of knowledge from experts (traditional knowledge acquisition) and Data mining (Nega and Adane, 2017). The development of an efficient KBS based on the development of an efficient knowledge base that has to be complete, coherent and non-redundant in order to make knowledge extraction as much as correct as possible (i.e. in order to keep the correctness of the knowledge as it is kept at the source) different techniques could be applied. Among these techniques, data mining techniques and, more general, knowledge discovery techniques became the most used in the recent years. The researcher acquires knowledge using two types of knowledge acquisition methods which are manual and automatic knowledge extraction.

It involves extracting knowledge from human experts, and/or written documents to build a knowledge-based system. In this study, the knowledge required to build a knowledge-based system was elicited from both tacit and explicit sources of knowledge. Tacit knowledge is collected from four experts in each domain area from Jimma University Specialized Hospital, Saint Paulos Millennium Medical Hospital and Adama Medical College Hospital by using structured and unstructured interviews (the interview questions used are found in Appendix I). Domain experts are chosen purposefully for wide-ranging discussion using structured and unstructured interviews to understand the domain knowledge. Furthermore to gain tacit knowledge using data mining techniques base line medical datasets were collected from each listed hospital. Explicit source of knowledge has been collected from the Internet, manuals, research papers and journal articles, guidelines etc.

3.2.1. KNOWLEDGE ABOUT DIABETES

The term diabetes mellitus describes a metabolic disorder of multiple etiology characterized by chronic hyperglycemia with disturbances of carbohydrate, fat and protein metabolism resulting from defects in insulin secretion, insulin action, or both. The effects of diabetes mellitus include long–term damage, dysfunction and failure of various organs. Diabetes mellitus may present with characteristic symptoms such as thirst, polyuria, blurring of vision, and weight loss. In its most severe forms, ketoacidosis or a non–ketotic hyperosmolar state may develop and lead to stupor, coma and, in absence of effective treatment, death (Cefalu, 2017).

Frequently symptoms are not severe, or may be absent, and consequently hyperglycemia sufficient to cause pathological and functional changes may be present for a long time before the diagnosis is made. The long–term effects of diabetes mellitus include progressive development of the specific complications of retinopathy with potential blindness, nephropathy that may lead to renal failure, and/or neuropathy with risk of foot ulcers, amputation, Charcot joints, and features of autonomic dysfunction, including sexual dysfunction. People with diabetes are at increased risk of cardiovascular, peripheral vascular and cerebrovascular disease. Several pathogenic processes are involved in the development of diabetes. These include processes which destroy the beta cells of the pancreas with consequent insulin deficiency, and others that result in resistance to insulin action. The abnormalities of carbohydrate, fat and protein metabolism are due to deficient action of insulin on target tissues resulting from insensitivity or lack of insulin (Cefalu, 2017).

Table 3.2.1: Types of diabetes summary

Type of Diabetes	**Description of Diabetes Mellitus**
Prediabetes	A condition characterized by slightly elevated blood glucose levels, regarded as indicative that a person is at risk of progressing to Type 2 diabetes. That has high blood pressure and was recently diagnosed with prediabetes.
Type I Diabetes	Immune mediated and idiopathic forms of cell dysfunction, which lead to absolute insulin deficiency. This is an autoimmune mediated disease process which gives rise to absolute deficiency of insulin and therefore total dependency upon insulin for survival.
Type II Diabetes	Disease of adult onset, which may originate from insulin resistance and relative insulin deficiency or from a secretory defect. This is a disease, which appears to have a very strong genetic predisposition and is caused by a combination of inadequate insulin secretion and an insensitivity of the body tissues to insulin so leaving patients with this condition relatively deficient in insulin.
Gestational Diabetes	The third main form and occurs when pregnant women without a previous history of diabetes develop a high blood glucose level.

(Cefalu, 2017).

3.2.2. COMMON SYMPTOMS OF DIABETES TYPES

Table 3.2.2: Common Symptoms Diabetes

Common Symptoms of diabetes	
✓ Polydipsia or excessive thirst ✓ Polyuria or frequent urination ✓ Polyphagia or excessive hunger ✓ Unusual weight loss ✓ Blurred vision ✓ Slow-healing wounds ✓ Recurring infections ✓ Headache ✓ Excessive Urination	✓ Feeling Pursiness ✓ Loss of Consciousness ✓ Nausea & Vomiting ✓ Tiredness ✓ Fraction& Skin Tag ✓ Late Ameliorative Wounds ✓ Frequent Infections ✓ Loss of Sensation in Hands and Feet ✓ Bad Breath ✓ Falling of Blood Sugar in ✓ Midnight with Cold Sweat ✓ Bellyache

3.2.3. SYMPTOMS OF DIFFERENT DIABETES TYPES

Table 3.2.3: Symptoms of different diabetes types

Diagnosis/Symptoms	**Type I Diabetes**	**Type II Diabetes**	**Pre Diabetes**	**Gestation Diabetes**
Increased thirst	Yes	Yes	Yes	No
Increased urge to urinate	Yes	Yes	Yes	No
Increased appetite	Yes	Yes	Yes	No
Weight reduction	Yes	No	No	No
Over weight	No	Yes	Yes	Yes
Weight variation	No	Yes	No	No
Impaired vision	No	Yes	Yes	No
Tiredness	Yes	Yes	No	No
Impatience	No	Yes	Yes	No
Infection	No	Yes	Yes	No
Itchy skin	No	No	Yes	No
Family history	No	Yes	Yes	No
Depression and stress	Yes	Yes	Yes	No
Tingling sensation	No	No	Yes	No

Fruity breath odour	Yes	No	No	No
Bed wetting	Yes	No	No	No
Slow-healing wounds	No	Yes	No	No
Family history of diabetes during pregnancy	No	No	No	Yes
Previous pregnancy	No	No	No	Yes
Baby over 9 pounds during previous pregnancy	No	No	No	Yes
Sleeplessness	Yes	Yes	No	No
Trembling	Yes	Yes	No	No
Sweating	Yes	Yes	No	No
Anxiety	Yes	Yes	No	No
Confusion	Yes	Yes	No	No
Weakness	Yes	No	No	No
Mood swings	Yes	No	No	No
Nausea	Yes	No	No	No
Vomiting	Yes	No	No	No
Dry skin	No	Yes	No	No
Aches and pains	No	Yes	No	No
Recurrent fungal infection	No	Yes	No	No
Nightmares	Yes	Yes	No	No
Seizures	Yes	Yes	No	No
Unconsciousness	Yes	Yes	No	No
Numbness	Yes	Yes	No	No
Vaginal meiotic infection	Yes	Yes	No	No
Rapid heart beat	No	Yes	No	No
Recurring gum infections	No	Yes	Yes	No
Impotency	No	No	Yes	No
High blood pressure	No	No	No	Yes
Sleep walking	Yes	Yes	No	No
Making unusual noises	Yes	Yes	No	No
leg cramps	Yes	Yes	No	No
Slurred speech	Yes	Yes	No	No
flushed face	Yes	Yes	No	No
Pale skin	Yes	No	No	No
Loss of menstruation	Yes	Yes	No	No
Stomach pain	Yes	Yes	No	No
Deep breathing	Yes	Yes	No	No
Areas of darkened skin	No	Yes	No	No

Difficult Concentrating	Yes	Yes	No	No
Dehydration	Yes	Yes	No	No
Lack of coordination	Yes	Yes	No	No
History of heart disease	No	No	Yes	No
Polycystic ovary syndrome	No	No	Yes	No

(Zeki et al., 2012).

3.2.4. RISK FACTORS OF DIABETES

Table 3.2.4: Risk factors of diabetes

Diagnosis/Risk Factors	**Healthy**	**At Risk**
Overweight	No	Yes
Age	<25	>=25
Blood Pressure	<140/90 Mm Hg	>=140/90 Mm Hg
Diabetic Parents or siblings	No	Yes
Hidden Diabetes	No	Yes
Rate of Triglycerides	<200	>=200
Abortion	No	Yes
Gestational Diabetes or Having Baby with over 4 Kg weight	No	Yes
Low Physical Activity (less than 3 times per Week)	No	Yes
Disorder in Glucose Tolerance in Previous Tests	No	Yes
Diabetics in Relatives	No	Yes
Rate of Fasting Blood Sugar between 110 and 125	No	Yes
History of Vascular Disease	No	Yes
Ovary Syndrome or Numerous Cysts	No	Yes

(Taken from Zeki et al., 2012).

3.2.5. CRITERIA FOR DIAGNOSIS AND TREATMENT OF DIABETES

Type 1 diabetes and type 2 diabetes are heterogeneous diseases in which clinical presentation and disease progression may vary considerably. Classification is important for determining therapy, but some individuals cannot be clearly classified as having type 1 or type 2 diabetes at the time of diagnosis. The traditional paradigms of type 2 diabetes occurring only in adults and type 1 diabetes only in children are no longer accurate, as both diseases occur in both cohorts. Occasionally, patients with type 2 diabetes may present with diabetic ketoacidosis (DKA), particularly ethnic minorities. Children with type 1 diabetes typically present with the hallmark symptoms of polyuria/polydipsia, and approximately one-third present with DKA (3). The onset of type 1 diabetes may be more variable in adults, and they may not present with the classic symptoms seen in children. Although difficulties in distinguishing diabetes type may occur in all age-groups at onset, the true diagnosis becomes more obvious over time (Cefalu, 2017).

3.2.6. DIAGNOSTIC CRITEIA

If a diagnosis of diabetes is made, the clinician must feel confident that the diagnosis is fully established since the consequences for the individual are considerable and lifelong. The requirements for diagnostic confirmation for a person presenting with severe symptoms and gross hyperglycemia differ from those for the asymptomatic person with blood glucose values found to be just above the diagnostic cut–off value. Severe hyperglycemia detected under conditions of acute infective, traumatic, circulatory or other stress may be transitory and should not in itself be regarded as diagnostic of diabetes (Cefalu, 2017).

The diagnosis of diabetes in an asymptomatic subject should never be made on the basis of a single abnormal blood glucose value. For the asymptomatic person, at least one additional plasma/blood glucose test result with a value in the diabetic range is essential, either fasting, from a random (casual) sample, or from the oral glucose tolerance test (OGTT). If such samples fail to confirm the diagnosis of diabetes mellitus, it will usually be advisable to maintain surveillance with periodic re–testing until the diagnostic situation becomes clear (Cefalu, 2017).

In these circumstances, the clinician should take into consideration such additional factors as ethnicity, family history, age, adiposity, and concomitant disorders, before deciding on a diagnostic or therapeutic course of action. An alternative to blood glucose estimation or the OGTT has long been sought to simplify the diagnosis of diabetes. Glaciated hemoglobin, reflecting average glycaemia over a period of weeks, was thought to provide such a test. Although in certain cases it gives equal or almost equal sensitivity and specificity to glucose measurement, it is not available in many parts of the world and is not well enough standardized for its use to be recommended at this time. The clinical diagnosis of diabetes is often prompted by symptoms such as increased thirst and urine volume, recurrent infections, unexplained weight loss and, in severe cases, drowsiness and coma; high levels of glycosuria are usually present (Cefalu, 2017). The research was used different international guideline help for diagnosis and treatment of diabetes like ADA, IDF, and WHO etc.

3.2.6.1. FASTING BLOOD SUGAR

A test to determine how much glucose (sugar) is in a blood sample after an overnight fast. The fasting blood glucose test is commonly used to detect diabetes mellitus. A blood sample is taken in a lab, physician's office, or hospital. The test is done in the morning, before the person has eaten. The normal range for blood glucose is 70 to 100 mg/dl. Levels between 100 and 126 mg/dl are referred to as impaired fasting glucose or pre-diabetes. Diabetes is typically diagnosed when fasting blood glucose levels are 126 mg/dl or higher (Cefalu, 2017).

Table 3.2.6.1: Diabetes Test Condition

Testing diabetes	**Normal**	**Diabetes**
Fasting blood Sugar(FBS)	80-99mg/dl	≥126mg/dl or (7.0 mmol/L)
Fasting blood Sugar(FBS)	100-125mg/dl	Pre-diabetes (impaired fasting glucose)
Random Blood Sugar	80-139mg/dl	≥200mg/dl or (11.1 mmol/L)
2hour glucose tolerance test	80-139mg/dl	≥200mg/dl or (11.1 mmol/L)
2hour glucose tolerance test	140to199mg/dl	Pre-diabetes (impaired glucose tolerance)
Glaciated hemoglobin (HbA1c)	≤6%	Glaciated hemoglobin ≥6.5%
mg = milligram, dL = deciliter ,mmol=millimol		

(Taken from Cefalu, 2017).

3.2.6.2. ORAL GLUCOSE TOLERANCE TEST

In the test, a person fasts overnight (at least 8 but not more than 16 hours). Then first, the fasting plasma glucose is tested. After this test, the person receives 75 grams of glucose (100 grams for pregnant women). Usually, the glucose is in a sweet-tasting liquid that the person drinks. Blood samples are taken up to four times to measure the blood glucose.

3.2.6.3. *HEMOGLOBIN BLOOD TEST*

The term Hemoglobin Blood test (HbA1c) refers to glaciated hemoglobin. It develops when hemoglobin, a protein within red blood cells that carries oxygen throughout the body, joins with glucose in the blood, becoming glaciated. By measuring glaciated hemoglobin (HbA1C), clinicians are able to get an overall picture of what our average blood sugar levels have been over a period of weeks/months. For people with diabetes this is important as the higher the HbA1c, the greater the risk of developing diabetes-related complications. HbA1c is also referred to as hemoglobin A1C or simply A1C.

The A1C test should be performed using a method that is certified by the NGSP (www.ngsp.org) and standardized or traceable to the Diabetes Control and Complications Trial (DCCT) reference assay. Although point-of-care A1C assays may be NGSP certified, proficiency testing is not mandated for performing the test, so use of point-of-care assays for diagnostic purposes is not recommended but may be considered in the future if proficiency testing is performed and documented. The A1C has several advantages compared with the FPG and OGTT, including greater convenience (fasting not required), greater pre analytical stability, and less day-to-day perturbations during stress and illness.

Table 3.2.6.3.1: Criteria for the diagnosis of diabetes

Criteria for the Diagnosis of Diabetes
Fasting Blood Sugar at cut point of FPG ≥126 mg/dL (7.0 mmol/L) of FPG ≥126 mg/dL (7.0 mmol/L). Fasting is defined as no caloric intake for at least 8 hour.
Or
2-h PG ≥200 mg/dL (11.1 mmol/L) during an OGTT. The test should be performed as described by the WHO, using a glucose load containing the equivalent of 75 g anhydrous glucose dissolved in water.
Or
A1C≥6.5% (48 mmol/mol). The test should be performed in a laboratory using a method that is NGSP certified and standardized to the DCCT assay.
Or
In a patient with classic symptoms of hyperglycemia or hyperglycemic crisis, a random plasma glucose ≥200 mg/dL (11.1 mmol/L).
In the absence of unequivocal hyperglycemia, results should be confirmed by repeat testing.

Table 3.2.6.3.2: Criteria for testing for diabetes or prediabetes in asymptomatic adults

Criteria for testing for diabetes or prediabetes in asymptomatic adults
1. Testing should be considered in overweight or obese (BMI ≥ 25 kg/m2 or23 kg/m2 in Asian Americans) adults who have one or more of the following risk factors:A1C ≥ 5.7% (39 mmol/mol), **IGT**, or **IFG** on previous testing first-degree relative with diabetes high-risk race/ethnicity (e.g., African American, Latino, Native American, Asian American, Pacific Islander) women who were diagnosed with **GDM** history of **CVD** hypertension (≥140/90 mmHg or on therapy for hypertension) HDL cholesterol level, ≤35 mg/dL (0.90 mmol/L) and/or a triglyceride level ≥250 mg/dL (2.82 mmol/L) women with polycystic ovary syndrome physical inactivity other clinical conditions associated with insulin resistance (e.g., severe obesity, a canthosis nigricans).
2. For all patients, testing should begin at age 45 years.
3. If results are normal, testing should be repeated at a minimum of 3-year intervals

(Taken from Cefalu, 2017)

Table 3.2.6.3.3: Criteria for testing for type 2 diabetes or prediabetes in asymptomatic children

Testing for type 2 diabetes or prediabetes in asymptomatic children Criteria
✓ Overweight (BMI.85th percentile for age and sex, weight for height.85th percentile, or Weight .120% of ideal for height).Plus any two of the following risk factors: ✓ Family history of type 2 diabetes in first- or second-degree relative ✓ Race/ethnicity (Native American, African American, Latino, Asian American, Pacific Islander) ✓ Signs of insulin resistance or conditions associated with insulin resistance (a canthosis nigricans, hypertension, dyslipidemia, polycystic ovary syndrome, or small-for gestational- age birth weight) ✓ Maternal history of diabetes or GDM during the child’s gestation ✓ Age of initiation: age 10 years or at onset of puberty, if puberty occurs at a younger age ✓ Frequency: every 3 years Persons aged ≤18 years.

3.2.7. TREATMENT AND ADVICES OF DIABETES

Nowadays, there are operational treatments for diabetes although not possible to make diabetic patient healthy again. If the patient is able to get the proper medication, quality of care and well medical advice, then he/she can have a healthy life and lessen the possibility of growing complications.

3.2.7.1. PREDIABETES TREATMENT

Prediabetes progresses, drug therapies directed towards hyperglycemia and the individual coronary heart disease risk factors may be required. Strict control of all known risk factors for CVD and microvascular complications in patients with type 2 diabetes by aggressive management of hypertension, dyslipidemia, and glycaemia and use of aspirin (as well as smoking cessation) has proved to be highly beneficial. Treatment goals for blood pressure and lipid control matching those for diabetes, given the strong evidence of increased cardiovascular risk for persons with prediabetes. These interventions, after or with lifestyle changes, may reduce CVD risks independently of treatments focused on the issue of glucose control and the prevention of microvascular risk (Garber et al., 2008).

If pre-diabetes is detected during investigation for diabetes, the treatment involves the same lifestyle changes that are recommended for people diagnosed with diabetes. For most, this will include regular physical activity, healthy eating and if necessary, losing weight.

Healthy eating: A healthy eating plan for losing weight and reducing the risk of type 2

Diabetes should include a reduction in total energy (kilojoule) and fat intake, particularly saturated fat foods such as butter, full fat dairy products, fatty meats, takeaway foods, biscuits, cakes and pastries. Instead choose a wide range of high fiber, low GI carbohydrate foods such as wholegrain breads and cereals, legumes and fruit.

Regular physical activity: Regular physical activity helps the body to use insulin better and to feel fit and healthy. Aim to do at least 30 minutes of 'moderate intensity' physical activity (such as brisk walking or swimming) on most, if not all, days of the week or three 20-minute sessions of 'vigorous intensity' exercise per week (such as jogging, aerobics class, strenuous gardening). Try to include some resistance training twice a week to improve the way your muscles work, such as body weight exercises or lifting weights such as cans of food (Alphafarm, 2009).

3.2.7.2. TYPE I DIABETES TREATMENT

Insulin adjustments in response to planned variations in eating or exercise

Diet: Calculate the carbohydrate content of the meal, and adjust the insulin dose based on the carbohydrate ratio that was prescribed (e.g., 1 unit for each 15 g of carbohydrate). The actual ratio of insulin units to grams of carbohydrate may vary in individuals from 1 unit/5 g of carbohydrate to 1 unit/20 g of carbohydrate.

Exercise: Insulin requirements may change by up to 50% during periods of exercise. Patients should monitor their glucose level before, during, and after exercise to determine the effects on their glucose levels. If the effects of the exercise are predictable, insulin doses can be adjusted.

Stress: Whether due to physical injury, infection or illness, iatrogenic use of steroids, or psychological factors, stress causes an increase in hormones that antagonize insulin (and thus increases glucose unless adjustments are made). Although stress usually causes glucose to rise, some people become more agitated and active during stress, leading to a drop in glucose (Kaiser, 2017).

Insulin therapy: In case of type 1 DM the only available therapy is insulin however in type 2 DM the patients who are not able to achieve glycaemia targets by oral agents, the insulin therapy should be instituted. Sulphonyureas should preferably be omitted from the treatment and patient should be subjected to insulin therapy. All insulin injections should preferably be administered in the abdomen, although they can also be given in the thigh, hip, or buttock regions (Srivastava, 2015).

3.2.7.3. TYPE II DIABETES TREATMENT

Medical Nutrition Therapy (MNT)

Nutrition therapy is recommended for all people with type 1 and type 2 diabetes as an effective component of the overall treatment plan by ADA. All individuals who have diabetes should receive individualized MNT as needed to achieve treatment goals. For overweight or obese adults with type 2 diabetes, reducing energy intake while maintaining a healthful eating pattern is recommended to promote weight loss which translates in clinical benefit in form of improved glycaemia, blood pressure, and/ or lipids.

The amount of carbohydrates and available insulin are the most important. Carbohydrate intake should be monitored either by carbohydrate counting or experience-based estimation. Substituting low–glycaemia load foods for higher–glycaemia load foods may modestly improve glycaemia control. The patients with no evidence of diabetic kidney disease, the goals should be individualized as there is no recommendation for ideal protein intake, however in patients with kidney disease reducing the amount of dietary protein below usual intake is not recommended because it does not alter glycaemia measures, cardiovascular risk measures, or the course of Glomerular Filtration Rate (GFR) decline. The quality of fat is more important than quantity so trans-fat should be reduced in diet. The goal of MNT in the individual with type 1 DM is to coordinate and match the caloric intake. MNT for type 2 DM should emphasize modest caloric reduction (low-carbohydrate or low-fat), reduced fat intake, and increased physical activity

Exercise

For individuals with type 1 or type 2 DM, exercise is also useful for lowering plasma glucose (during and following exercise) and increasing insulin sensitivity. In patients with diabetes, moderate aerobic physical activity of 150 min/week (distributed over at least 3 days) is recommended by ADA.

Pharmacologic Therapy

Pharmacological therapy is aimed at maintaining the glycaemia and reducing the long term complications of Diabetes. Drug classes used for the treatment of type 2 diabetes such as:

(1) Insulin sensitizers: (a) Biguanides ; (b)Thiazolidinedione’s (TZDs); (2) Insulin secretagogues: (a) Sulfonylureas; (b) Meglitinide derivatives; (3) Alphaglucosidase inhibitors; (4) Glucagonlike peptide–1 (GLP-1) agonists; (5) Dipeptidy-l peptidase IV (DPP-4) inhibitors; (6) Selective sodium-glucose transporter-2 (SGLT-2) inhibitors (7) Insulin; (8)Amylinomimetics(Srivastava, 2015).

3.2.7.4. GESTATIONAL DIABETES TREATMENT

Gestational diabetes mellitus (GDM) is a common metabolic disorder that occurs during pregnancy. GDM can cause significant problems, including maternal complications, perinatal complications, and metabolic disorders in off spring of mothers with GDM. The primary management method for women with GDM is nutritional therapy. Some women with GDM require diet therapy alone, while some women require both diet therapy and insulin therapy. Currently, there is no universal management method for GDM because there are no universal diagnostic criteria and genomic backgrounds differ according to ethnicity (Sugiyama, 2011).

Table 3.2.6.2: Guideline for blood sugar level monitoring and advice of treatment

A1C (%)	**Advice**	**FPG (mg/dL)**
> 13.0	Patient need to consult his/her experts, if the result is above 325 mg/dL.	>325
≥8.0 and ≤13.0	If the glucose level is ≥183 and ≤325 mg/dL, then it indicates that the glucose level control was not effective. If the glucose level is always in this range, then patient likely to be affected with the complications of diabetes disease.	≥183 and ≤325
≥7.0 and <8.0	If the glucose level is ≥155 and <183 mg/dL, patient control of the glucose level is not as much. Re-examination of the glucose level is required to correct the necessary things.	≥155 and <183
≥6.0 and <7.0	If the glucose level is ≥126 and <155 mg/dL, it indicates that patient previously exercising well, had a healthy diet and take insulin properly. So, patient should continue on doing it.	≥126 and <155
≥4.0 and <6.0	If a person is non-diabetic, then FPG test result is ≥68 and <126 mg/dL. But if the person is diabetic and the number is in this range, this indicates repeated low glucose level.	

3.3. KNOWLEDGE MODE

It involves using concepts discovered during the knowledge elicitation session to build a model or representation of the domain experts’ knowledge and automatic knowledge acquisition. It is a process where knowledge engineer uses concepts discovered during the knowledge acquisition phase to build a model of the domain. The knowledge used for building of the knowledge-based system in this study focused on knowledge regarding the diagnosis and treatment of diabetes.

3.4. KNOWLEDGE REPRESENTATION

Data mining refers to the application of algorithms for extracting patterns from data. Data mining is the step in the process of knowledge discovery in databases, that inputs predominantly cleaned, transformed data, searches the data using algorithms, and outputs patterns and relationships to the interpretation/ evaluation step of the KDD process (Inderpal, 2013). Classification maps data into predefined groups or classes. It is often referred to as supervised learning because the classes are determined before examining the data. Classification algorithms require that the classes be defined based on data attribute values. They often describe these classes by looking at the characteristics of data already known to belong to the classes (Bhambri, 2011). Classification is form of data analysis that can be used to extract models describing important data classes or to predict future data trends and classification predicts categorical (discrete, unordered) labels (Asghar, 2009). The objective of this step is to apply three classification technique algorithms on Patients' diabetic data set which have been collected from different hospital and develop a model that can produce the automatic knowledge so that to use the model for KBS development.

3.4.1. DOMAIN EXPERTS KNOWLEDGE ACQUISITION

In this work, the tradition methods (such as interview and document analysis) are used primarily for understanding the basic concepts related to diagnosis, treatment and prognosis of diabetes disease. More specifically, interviews and document analysis are used to access the general and domain-specific knowledge and to obtain comprehensive example sets. On the other hand, data mining approach is particularly fruitful in automating the knowledge acquisition task of rule based system. However, it is a mistake to believe that one can do data mining process without a domain expert. Because at the very least the researchers need an expert to select the training examples and to explain the domain terminology as well as to identify the features of the examples which are likely to be relevant. Therefore, the researcher used the traditional methods to supplement the automatic knowledge's acquisition of the KBS development.

3.4.2. KNOWLEDGE ACQUISITION AUTOMATICALLY

The development of an efficient and effective smart knowledge based system involves the development of an efficient knowledge base that has to be complete, coherent and non-redundant. Knowledge acquisition is most difficult and error-prone task in the development of KBS due to that knowledge acquisition involves communications between people with completely different backgrounds, human experts and knowledge engineers, who must formulate the concepts, relations and control mechanisms needed for the knowledge based system (Ambilwade Central, Ramesh Manza, 2014).

Moreover, tacit knowledge is difficult to transfer and understand from interviews .In this case, the knowledge acquisition problem can be addressed by data mining techniques. Knowledge acquisition using data mining technique eliminates or reduces the difficulty caused by the knowledge acquisition bottleneck of rule based systems and automates knowledge acquisition by obtaining low-cost, add better values on tradition method, and high-quality knowledge base.

Data mining used as extraction of implicit, previously unknown and potentially useful information from data. It uses machine learning, statistical and visualization techniques to discover and present knowledge in a form, which is easily comprehensible to humans. Knowledge Discovery is needed to make sense and use of data. Knowledge Discovery in Data is the non-trivial process of identifying valid, novel, potentially useful and ultimately understandable patterns in data (Tripathi, 2011).Data mining techniques consists of more than collection and managing data; it also includes analysis and prediction. People are often do mistakes while analyzing or, possibly, when trying to establish relationships between multiple features. This makes it difficult for them to find solutions to certain problems. Machine learning can often be successfully applied to these problems, improving the efficiency of systems and the designs of machines. There are several applications for Machine Learning, the most significant of which is data mining. Numerous machine learning applications involve tasks that can be set up as supervised.

3.4.2.1. DATA PREPROCESSING

There different data preprocessing techniques. Data cleaning can be applied to remove noise and correct inconsistencies in the data. Data integration merges data from multiple sources into a coherent data store. Data transformations, such as normalization, may improve the accuracy and efficiency of mining algorithms involving distance measurements. Data reduction can reduce the data size by aggregating, eliminating redundant features. These techniques are not mutually exclusive; it can work together for a better data quality. Data mining requires access to data. The data may be represented as volumes of records in several database files or the data may contain only a few hundred records in a single file (Inderpal, 2013). In this study the researcher used dataset from a baseline or a spreadsheet patient's datasets. The researcher performs preprocessing activities to make the data more suitable for data mining techniques.

i. Data Selection:

The first step in data mining is to select the types of data to be used. A total of **2640** diabetes patients' datasets were collected from three different hospitals. **1500** datasets were collected from Jimma University Specialized Hospital, Saint Paulos Millennium Medical Hospital **320** and **820** datasets from Adama Medical College Hospital. Cross fold validation 10 used as default and experiment design dividing the prepared data set into 70% training and 30% for testing then accordingly running the experiment is the task. This means that by deciding the number of iteration that the algorithm would iterate. 70% of the original data was selected for training purpose since the classifier learns more from large amount of data and increases its performance.

ii. Dataset Description

The dataset contains the following attributes are selected by discussing with domain experts

Table 3.4.2.1: Datasets attribute description

No	Attributes	Description	Values
1	Patients Status	Patients Gender and characteristics	Nominal values :{Male ,Femalepregnant and Female non pregnant}
2	Patients Age	Age of Patients	Numeric values
3	DmSymptoms	Diabetes symptoms of patients	Nominal values :{Healthy, Diabetes}
4	Dm_Riskfactor	riskfactor of diabetes	Nominal values :{Healthy, At Risk}
5	BMI	Body mass index of patients measure	Nominal values :{Overweight , Healthywieght, Underweight}
6	FBS1	Patients First Blood Sugar test	Numeric values
7	FBS2	patients Second Blood Sugar test	Numeric values
8	HbAC1	Patients HbAC1 test	Numeric values
9	BP(Blood Pressure)	Patients' hypertension measure test	Numeric values
10	HDL	Patients Cholesterols level test	Numeric values
11	TGL(triglyceride)	Patients rate of triglyceride level test	Numeric values
12	Class of Diagnosis and Treatment	Nominal Values:{ Healthy, Diabetes, TypeIDm, AtRiskTypeIDm, TypeIIDm,AtrikTypeIIDm, GestationalDm,AtriskgestationDm}	

iii. **Data Cleaning:**

Real-world data tend to be incomplete, noisy, and inconsistent. Data cleaning routines attempt to fill in missing values, smooth out noise while identifying outliers, and correct inconsistencies in the data (Zhuang etal, 2009). Cleaning the data is proved a nontrivial and tedious task data error identification is both an automated and a manual process, and required an iterative procedure that drew upon different experts (Inderpal, 2013). Data cleaning routines work to clean the data by filling in missing values, smoothing noisy data, identifying or removing outliers, and resolving inconsistencies. If users believe the data are dirty, they are unlikely to trust the results of any data mining that has been applied to it.

Furthermore, dirty data can cause confusion for the mining procedure, resulting in unreliable output. Although most mining routines have some procedures for dealing with incomplete or noisy data, they are not always robust. Instead, they may concentrate on avoiding over fitting the data to the function being modeled.

The researcher was cleaned the data that has been collected from each hospitals separately according to the diabetic patient's baseline data by discussing with domain experts. Datasets status are varies from patients to patients and incomplete, as a result the researcher corrected the entire data in the datasets before using in modeling. Since the data for all patients have been recorded based on patient identification card separately, first the researcher integrated these data into one to make the data cleaning process more convenient and easy. Therefore, the data that have been recorded for each patients within the sheets have integrated into one sheet and made ready for the next preprocessing step.

Some attributes which are confidential or irrelevant for the decision making process were removed in data preprocessing task. The removal of inappropriate and unnecessary attributes from the dataset is applied on this step of classification. Therefore, attributes described in table below has been no any significant values for mining purpose and are removed.

Table 3.4.2.2 Removed attributes

Removed Attributes From Datasets		
No	**Attribute Name**	**Reason**
1	Patients Card number	Not necessary
2	Patients names	Not necessary
3	Address	Not necessary
4	Marital status	Not necessary
5	Medicine material	Not necessary
6	Temperature	Not necessary
7	Medicine names and types	Not necessary

iv. **Data Integration:**

Data integration combines data from multiple sources into a coherent data store. Since the original dataset collected from different sources and the dataset has different attributes for different patients, entity identification problem and redundancy issues are solved by taking the common attributes for datasets that was collected.

v. **Data Transformation**

Data Transformation techniques can be used to reduce the number of values for a given continuous attribute by dividing the range of the attribute into intervals. Interval labels can be used to replace actual data values. Replacing numerous values of a continuous attribute by a small number of interval labels there by reduces and simplifies the original data. This leads to a concise and easy to use knowledge level representation of mining results.

vi. **Attribute Selection:**

In processing medical data, choosing the optimal subset of features is such important, not only to reduce the processing cost but also to improve the usefulness of the model built from the selected data (Inderpal, 2013). The goal of attribute subset selection is to find a minimum set of attributes such that the resulting probability distribution of the data classes is as close as possible to the original distribution obtained using all attributes. Mining on a reduced set of attributes has an additional benefit of reducing the number of attributes appearing in the discovered patterns, helping to make the patterns easier to understand.

To select the best attributes for data mining, the researcher uses information gain method which exists in WEKA data mining tool. Before calculating the information gain of the attributes the researcher was discussed with domain experts' for select the most significant attributes.

```
ttribute selection output

=== Attribute Selection on all input data ===

Search Method:
        Attribute ranking.

Attribute Evaluator (supervised, Class (nominal): 12 Class):
        Gain Ratio feature evaluator

Ranked attributes:
 1          3 Diabetes Symptoms
 0.59095    4 Diabetes Riskfactor
 0.29825    2 patients Age
 0.22755    6 FBS1
 0.22303    1 patients Status
 0.17646    9 BP
 0.13832    5 BMI
 0.09926    7 FBS2
 0.02613   11 TGL
 0.01674   10 HDL
 0.00806    8 HBA1C

Selected attributes: 3,4,2,6,1,9,5,7,11,10,8 : 11
```

Figure-3.1: Attribute Selection on all input data using Weka

vii. Data formatting

Weka needs data to be prepared in some formats and file types. The datasets provided to this software were prepared in a format that is acceptable for Weka software. Weka accepts records which attribute values are separated by commas and saved in an ARFF (Attribute-Relation File Format) file format (a file name with an extension of ARFF i.e. FileName.arff). At first the integrated dataset was in an excel file format. To feed the preprocessed datasets into the Weka tool data mining software the file is changed into other file format. The excel file was first changed into a comma delimited (CSV) file format. After changing the datasets into ARFF format the next step was opening the file with the Weka data mining software. The class label is the dependent attribute and the rest are independent and value of the class label attribute produced. In this study, the dataset has ten Classes namely: Healthy, Diabetes, PreDm, AtRiskPreDm, TypeIDm, AtriskTypeIDm, TypeIIDm, AtRiskTypeIIDm, GestationalDm and AtRiskgestationDm.

v. Data Mining and Model Selection:

To build the analytical model for diagnosis and treatment of diabetes, three classification algorithms namely J48, JRIP and PART are built. J48 is tree based classifiers in WEKA whereas JRip and PART are rule based classifiers. In supervised learning, the training data are accompanied by class labels indicating the class of the clarifications.

3.4.2.2. EXPERIMENT-1-J48 PRUNED TREE

Decision trees are data-mining methodologies applied in many real-world applications as a powerful solution to classification problems. In decision tree experiment, the performance of J48 classifier in predicting diabetes status of patients was evaluated. The experiment was conducted with the default parameters of WEKA. From the total dataset of **2640** records, **2512** were correctly classified and the remaining **128** instances were incorrectly classified. Decision tree is a graphical representation of the relations that exist between the data in the database. It is used for data classification. The result is displayed as a tree, hence the name of this technique. Decision trees are mainly used in the classification and prediction. It is a simple and a powerful way of representing knowledge.

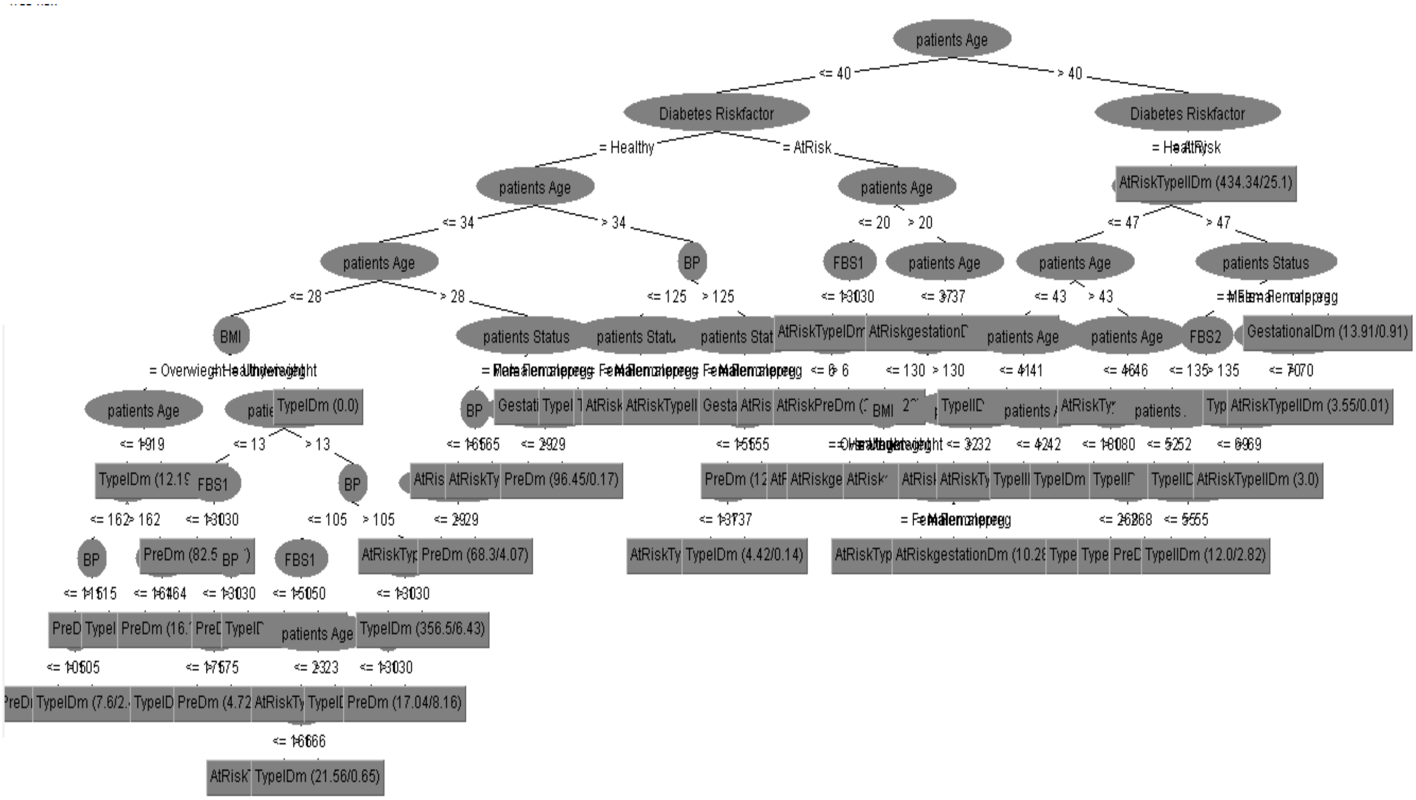

Figure-3.2: J48 pruned decision tree Weka output

The models obtained from the decision tree are represented as a tree structure. The instances are classified by sorting them down the tree from the root node to some leaf node. The nodes are branching based on if-then condition (Boris & Milan, 2012). This experiment conducted under percentage split test option with 90 % of the data set for training and the remaining for testing with default parameters of Weka and the algorithm generates a model as a decision tree with 59 Number of Leaves and 110 Size of the tree and Correctly Classified Instances are 2512 which means 95.1515 % and Incorrectly Classified Instances are 128 which means 4.8485 % from Total Number of Instances of 2640. The algorithm takes 0.31 seconds to develop the model.

=== Detailed Accuracy By Class ===

	TP Rate	FP Rate	Precision	Recall	F-Measure	ROC Area	Class
	0	0	0	0	0	0.972	Healthy
	0	0	0	0	0	0.783	Diabetes
	0.968	0.011	0.95	0.968	0.959	0.998	PreDm
	0	0	0	0	0	0.987	AtRiskPreDm
	0.975	0.018	0.95	0.975	0.962	0.998	TypeIDm
	0.915	0.002	0.972	0.915	0.943	0.997	AtRiskTypeIDm
	0.987	0.02	0.913	0.987	0.948	0.997	TypeIIDm
	0.916	0.005	0.984	0.916	0.949	0.997	AtRiskTypeIIDm
	1	0.003	0.9	1	0.947	1	GestationalDm
	0.918	0.002	0.957	0.918	0.938	0.994	AtRiskgestationDm
Weighted Avg.	0.952	0.012	0.949	0.952	0.95	0.997	

Figure-3.3: Detailed accuracy by J48 classification algorithms

=== Confusion Matrix ===

a	b	c	d	e	f	g	h	i	j	<-- classified as
0	0	1	0	0	0	1	0	1	0	a = Healthy
0	0	2	0	0	0	0	0	1	0	b = Diabetes
0	0	460	0	10	0	3	0	2	0	c = PreDm
0	0	0	0	0	2	0	0	0	2	d = AtRiskPreDm
0	0	12	0	661	3	0	2	0	0	e = TypeIDm
0	0	2	0	10	173	0	4	0	0	f = AtRiskTypeIDm
0	0	2	0	2	0	451	2	0	0	g = TypeIIDm
0	0	5	0	9	0	39	614	1	2	h = AtRiskTypeIIDm
0	0	0	0	0	0	0	0	63	0	i = GestationalDm
0	0	0	0	4	0	0	2	2	90	j = AtRiskgestationDm

Figure-3.4: Confusion Matrix J48

Classifier output

```
patients Age <= 40
|   Diabetes Riskfactor = Healthy
|   |   patients Age <= 34
|   |   |   patients Age <= 28
|   |   |   |   BMI = Overwieght
|   |   |   |   |   patients Age <= 19
|   |   |   |   |   |   FBS1 <= 162
|   |   |   |   |   |   |   BP <= 115
|   |   |   |   |   |   |   |   BP <= 105: PreDm (46.87/1.78)
|   |   |   |   |   |   |   |   BP > 105: TypeIDm (7.6/2.47)
|   |   |   |   |   |   |   BP > 115: PreDm (90.88/3.46)
|   |   |   |   |   |   FBS1 > 162
|   |   |   |   |   |   |   FBS1 <= 164: TypeIDm (8.82/0.25)
|   |   |   |   |   |   |   FBS1 > 164: PreDm (16.31/0.99)
|   |   |   |   |   patients Age > 19: TypeIDm (12.19/1.98)
|   |   |   |   BMI = Healthywieght
|   |   |   |   |   patients Age <= 13
|   |   |   |   |   |   FBS1 <= 130: PreDm (82.5/10.3)
|   |   |   |   |   |   FBS1 > 130
|   |   |   |   |   |   |   BP <= 130
|   |   |   |   |   |   |   |   FBS2 <= 175: TypeIDm (58.94/3.52)
|   |   |   |   |   |   |   |   FBS2 > 175: PreDm (4.72/0.66)
|   |   |   |   |   |   |   BP > 130: PreDm (9.21/0.34)
|   |   |   |   |   patients Age > 13
|   |   |   |   |   |   BP <= 105
|   |   |   |   |   |   |   FBS1 <= 150: TypeIDm (9.57/1.28)
|   |   |   |   |   |   |   FBS1 > 150
|   |   |   |   |   |   |   |   patients Age <= 23
|   |   |   |   |   |   |   |   |   FBS1 <= 166: AtRiskTypeIDm (8.21/1.1)
|   |   |   |   |   |   |   |   |   FBS1 > 166: TypeIDm (21.56/0.65)
|   |   |   |   |   |   |   |   patients Age > 23: AtRiskTypeIDm (23.88/3.74)
|   |   |   |   |   |   BP > 105
|   |   |   |   |   |   |   FBS1 <= 130
```

Figure-3.5: the sample rules generated by J48 classifier decision tree

3.4.2.3. EXPERIMENT -2-PART CLASSIFIER

PART is a rule-based classifier uses a set of IF-THEN rules for classification. An IF-THEN rule is an expression of the form IF condition THEN conclusion. The “IF”-part (or left-hand side) of a rule is known as the rule antecedent or pre condition. The “THEN”-part (or right-hand side) is the rule consequent (Chen, 2009). This experiment conducted under percentage splits technique using 90 % of instances for training and the remaining for testing with default parameters of Weka and the algorithm generates a model with 40 rules and Correctly Classified Instances are 2495 which means 94.5076 % and Incorrectly Classified Instances are 145 which means 5.4924 % from Total Number of Instances of 2640. The algorithm takes 0.69 seconds to develop the model.

PART decision list

```
Classifier output

PART decision list
------------------

patients Age > 40 AND
Diabetes Riskfactor = AtRisk: AtRiskTypeIIDm (434.34/25.1)

patients Age > 40 AND
patients Age > 47 AND
patients Status = Femalenonpreg AND
patients Age <= 70: TypeIIDm (131.06/8.03)

patients Age <= 30 AND
Diabetes Riskfactor = AtRisk AND
patients Age <= 20 AND
FBS1 > 130: AtRiskTypeIDm (126.69/6.37)

patients Age <= 32 AND
patients Age <= 28 AND
BMI = Overwieght AND
patients Age <= 19 AND
FBS1 <= 162 AND
BP > 115: PreDm (91.71/3.96)

patients Age > 40 AND
patients Status = Male AND
FBS2 > 127 AND
FBS1 > 93 AND
BP > 90 AND
FBS2 > 136 AND
patients Age <= 69 AND
FBS1 <= 166: TypeIIDm (205.01/11.96)
```

Figure -3.6: Rules generated by PART Classification algorithm

=== Detailed Accuracy By Class ===

	TP Rate	FP Rate	Precision	Recall	F-Measure	ROC Area	Class
	0	0	0	0	0	0.617	Healthy
	0	0	0	0	0	0.597	Diabetes
	0.971	0.015	0.933	0.971	0.951	0.998	PreDm
	0.5	0	0.667	0.5	0.571	0.994	AtRiskPreDm
	0.966	0.01	0.972	0.966	0.969	0.999	TypeIDm
	0.91	0.001	0.983	0.91	0.945	0.992	AtRiskTypeIDm
	0.954	0.021	0.906	0.954	0.93	0.997	TypeIIDm
	0.919	0.017	0.949	0.919	0.934	0.995	AtRiskTypeIIDm
	1	0.004	0.851	1	0.92	0.999	GestationalDm
	0.918	0	1	0.918	0.957	0.993	AtRiskgestationDm
Weighted Avg.	0.945	0.013	0.944	0.945	0.944	0.996	

Figure -3.7: Detailed accuracy by PART algorithms

=== Confusion Matrix ===

a	b	c	d	e	f	g	h	i	j	<-- classified as
0	0	2	0	0	0	0	0	1	0	a = Healthy
0	0	2	0	0	0	0	0	1	0	b = Diabetes
0	0	461	0	8	0	3	1	2	0	c = PreDm
0	0	0	2	0	2	0	0	0	0	d = AtRiskPreDm
0	0	16	0	655	1	0	6	0	0	e = TypeIDm
0	0	5	0	7	172	0	5	0	0	f = AtRiskTypeIDm
0	0	2	0	0	0	436	19	0	0	g = TypeIIDm
0	0	6	1	4	0	42	616	1	0	h = AtRiskTypeIIDm
0	0	0	0	0	0	0	0	63	0	i = GestationalDm
0	0	0	0	0	0	0	2	6	90	j = AtRiskgestationDm

Figure -3.8: Confusion Matrix by PART classifier

3.4.2.4. EXPERIMENT -3-JRIP CLASSIFIER

JRip is a rule based classifier that extracts rules from a large dataset. With JRip, IF-THEN rules are generated from the experimental diabetic dataset with the default parameters of WEKA and 10-fold cross-validation test mode. This experiment conducted under percentage splits technique using 90 % of instances for training and the remaining for testing with default parameters of Weka and the algorithm generates a model with 32 rules and Correctly Classified Instances are 2501 which means 94.7348 % and Incorrectly Classified Instances are 139 which means 5.2652 % from. Total Number of Instances of 2640. The algorithm takes 2.62 seconds to develop the model.

Classifier output

```
JRIP rules:
===========


(Diabetes Symptoms = Healthy) => Class=Healthy (3.0/0.0)
(HBA1C >= 6.5) => Class=AtRiskPreDm (4.0/0.0)
(patients Status = Femalepreg) and (Diabetes Riskfactor = Healthy) => Class=GestationalDm (59.0/1.0)
(patients Status = Femalepreg) and (BP <= 169) => Class=AtRiskgestationDm (69.0/0.0)
(BP >= 160) and (FBS1 <= 120) and (patients Age >= 36) => Class=AtRiskgestationDm (22.0/0.0)
(patients Status = Femalepreg) and (patients Age >= 38) => Class=AtRiskgestationDm (6.0/1.0)
(patients Age <= 29) and (Diabetes Riskfactor = AtRisk) and (patients Age <= 20) => Class=AtRiskTypeIDm (121.0/0.0)
(patients Age <= 30) and (patients Age >= 24) and (TGL >= 170) and (patients Age <= 29) => Class=AtRiskTypeIDm (19.0/0.0)
(patients Age <= 30) and (patients Age >= 25) and (FBS2 <= 155) and (patients Age <= 29) and (FBS1 >= 155) => Class=AtRiskTypeIDm (14.0/0.0)
(patients Age <= 30) and (patients Age >= 29) and (FBS2 >= 210) => Class=AtRiskTypeIDm (11.0/0.0)
(patients Age <= 16) and (patients Age >= 16) and (BP <= 100) and (BMI = Healthywieght) => Class=AtRiskTypeIDm (7.0/0.0)
(patients Age >= 41) and (Diabetes Riskfactor = Healthy) and (patients Age >= 48) and (FBS2 >= 144) => Class=TypeIIDm (226.0/0.0)
(patients Age >= 41) and (Diabetes Riskfactor = Healthy) and (FBS1 <= 140) and (patients Age >= 48) => Class=TypeIIDm (76.0/0.0)
(patients Age >= 41) and (Diabetes Riskfactor = Healthy) and (FBS1 >= 192) and (FBS1 <= 252) => Class=TypeIIDm (61.0/0.0)
(patients Age >= 41) and (patients Age <= 41) => Class=TypeIIDm (44.0/0.0)
(patients Age >= 48) and (patients Age >= 56) and (patients Age <= 67) and (patients Age >= 65) => Class=TypeIIDm (6.0/0.0)
(patients Age >= 48) and (patients Age <= 50) and (patients Age >= 49) => Class=TypeIIDm (6.0/0.0)
(patients Age <= 34) and (FBS1 <= 120) => Class=PreDm (186.0/4.0)
(patients Age <= 34) and (patients Age >= 30) and (Diabetes Riskfactor = Healthy) => Class=PreDm (152.0/0.0)
(FBS2 <= 126) and (FBS1 >= 155) and (FBS2 >= 120) and (patients Status = Male) => Class=PreDm (52.0/0.0)
(patients Age <= 13) and (BMI = Overwieght) => Class=PreDm (16.0/0.0)
(patients Age <= 39) and (patients Age >= 39) => Class=PreDm (12.0/0.0)
(patients Age <= 34) and (patients Age <= 13) and (patients Age >= 13) and (patients Status = Male) => Class=PreDm (11.0/1.0)
(patients Age <= 34) and (BMI = Overwieght) and (patients Age <= 19) => Class=PreDm (15.0/0.0)
(patients Age <= 34) and (patients Age >= 30) and (patients Status = Femalenonpreg) and (patients Age <= 32) => Class=PreDm (5.0/0.0)
(HBA1C >= 6) and (patients Status = Femalepreg) => Class=PreDm (2.0/0.0)
(Diabetes Riskfactor = AtRisk) => Class=AtRiskTypeIIDm (525.0/4.0)
(patients Age >= 33) and (BP >= 125) => Class=AtRiskTypeIIDm (38.0/0.0)
(patients Age >= 42) and (patients Age <= 42) => Class=AtRiskTypeIIDm (39.0/0.0)
(patients Age >= 47) => Class=AtRiskTypeIIDm (91.0/35.0)
```

Figure-3.9: Rules generated by JRIP Classification algorithm

=== Detailed Accuracy By Class ===

	TP Rate	FP Rate	Precision	Recall	F-Measure	ROC Area	Class
	1	0	1	1	1	1	Healthy
	0	0	0	0	0	0.797	Diabetes
	0.928	0.006	0.974	0.928	0.95	0.985	PreDm
	1	0	1	1	1	1	AtRiskPreDm
	0.988	0.045	0.883	0.988	0.932	0.978	TypeIDm
	0.905	0.001	0.988	0.905	0.945	0.972	AtRiskTypeIDm
	0.91	0	1	0.91	0.953	0.994	TypeIIDm
	0.958	0.017	0.95	0.958	0.954	0.991	AtRiskTypeIIDm
	0.921	0	0.983	0.921	0.951	0.979	GestationalDm
	0.98	0	0.99	0.98	0.985	0.987	AtRiskgestationDm
Weighted Avg.	0.947	0.017	0.95	0.947	0.947	0.985	

Figure -3.10: Detailed accuracy by JRip algorithms

=== Confusion Matrix ===

a	b	c	d	e	f	g	h	i	j	<-- classified as
3	0	0	0	0	0	0	0	0	0	a = Healthy
0	0	2	0	0	0	0	0	1	0	b = Diabetes
0	0	441	0	30	0	0	4	0	0	c = PreDm
0	0	0	4	0	0	0	0	0	0	d = AtRiskPreDm
0	0	8	0	670	0	0	0	0	0	e = TypeIDm
0	0	1	0	13	171	0	4	0	0	f = AtRiskTypeIDm
0	0	1	0	16	0	416	24	0	0	g = TypeIIDm
0	0	0	0	26	2	0	642	0	0	h = AtRiskTypeIIDm
0	0	0	0	3	0	0	1	58	1	i = GestationalDm
0	0	0	0	1	0	0	1	0	96	j = AtRiskgestationDm

Figure -3.11: Confusion Matrix JRip classifier

3.4.3. MODEL EVALUATION

All the selected algorithms allow generating rules from the data set. The results of the algorithms are evaluated based on prediction accuracy in classifying the instances of the dataset into Healthy, Diabetes, PreDm, AtRiskPreDm, TypeIDm, AtriskTypeIDm, TypeIIDm, AtRiskTypeIIDm, GestationalDm and AtRiskgestationDm. The performance of classifier algorithms is compared and the one which performed better is selected as prime choice for the knowledge acquisition step.

The accuracy, precision, recall and f-measure of each of the mentioned classifiers which are obtained from the experiment are shown in table.

Table 3.4.2.3: Performance of Classifiers

Model Evaluation	Correctly classified instances		Incorrectly classified instances		Time taken to build model	Precision	Recall	F Measure
Classifiers	Instances	Percentage	Instances	Percentage	Time/seconds			
J48	2512	95.1515 %	128	4.8485 %	0.31	0.949	0.952	0.95
PART	2495	94.5076 %	145	5.4924%	0.69	0.944	0.945	0.944
JRip	2501	94.7348 %	139	5.2652%	2.69	0.95	0.947	0.947

As shown in Table above, three experiments were carried out using decision tree classifiers (i.e. J48 pruned tree) and rule based classifiers (i.e. PART and JRip). From this experiment one can observe that the J48 classifier achieves best accuracy by classifying 2512 instances out of 2640 correctly comparing with JRip and PART. Results of JRip and PART show that nearly equal number of incorrectly classified instances. The highest incorrect classification is registered by JRip algorithm. Table 3.11.3.1. depicts the confusion matrix of the best performing classifier (i.e. PART classifier).Based on the results obtained by objective interestingness evaluation methods result, the researcher decided to use J48 classification algorithm model for further use in the development of rule base of the KBS because it registered better performance than JRip and PART Classification algorithms.

CHAPTER FOUR

DESIGN AND IMPLEMENTATION OF THE PROTOTYPE

4.1. INTRODUCTION

In the following sections, the implementation includes the real construction of the prototype system for diagnosis and treatment of diabetes using data mining techniques. After the necessary knowledge is represented using a rule-based knowledge representation technique, the next step is coding the represented knowledge using Prolog programming language into a suitable format that is understandable by the inference engine and integrating to micro visual studio 2013 for clear and easy user interface through assembly library.

4.2. ARCHITECTURE OF THE PROTOTYPE SYSTEM

An architecture is a blueprint showing how the components of the prototype self-learning knowledge-based system interacts and interrelates.

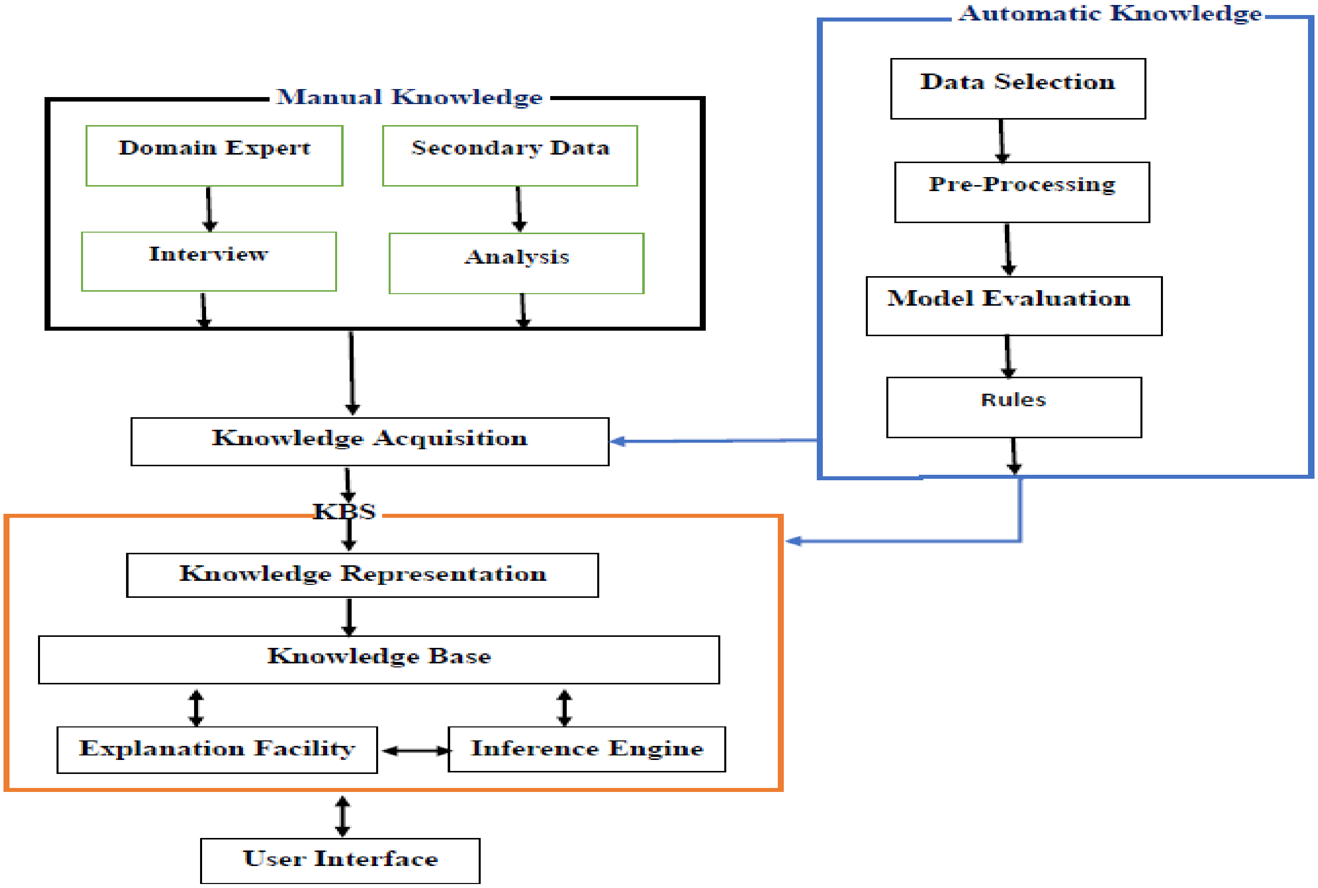

Figure-4.1: Architecture of KBS

4.3. SOFTWARE REQUIREMENT FOR DEVELOPMENT KBS

Clisp.net is integrating common lisp prolog with dot net frame work using library assembly plugin mechanism to visual studio development environment. For this study, Clisp.net integration tool is used to construct the prototype system. The KBS infers solution by running the knowledge base through an inference engine, a software program that interacts with the user and processes the results from the rules and facts in the knowledge base. In this study, Prototype KBS is developed by the software Clips.Net Prolog integration with Microsoft Visual Studio Ultimate 2013. This programming language is preferred due to its object oriented nature and its interactive capabilities with the users.

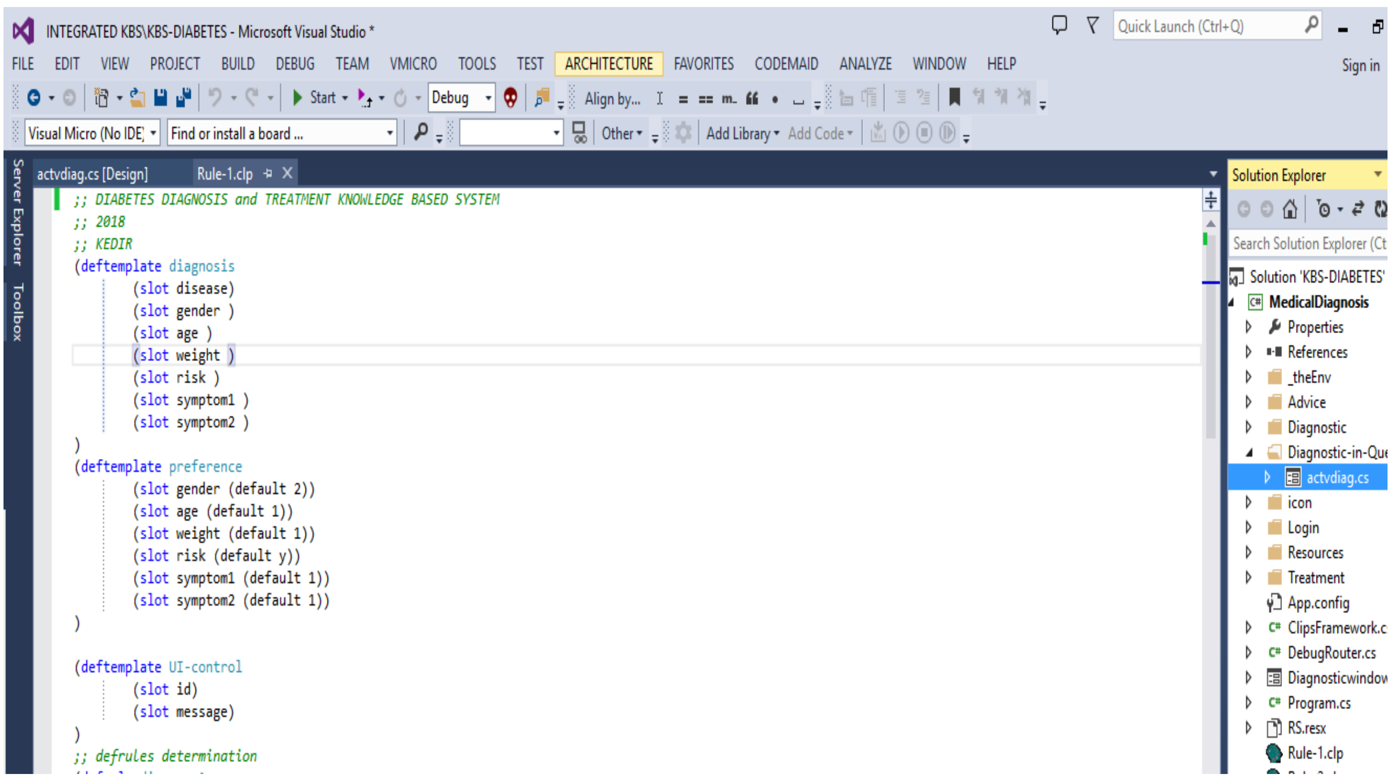

Figure-4.2: Prototype of KBS development environment

4.4. KNOWLEDGE BASE

Knowledge base is a set of rules or the encoded knowledge about diagnosis and treatment of diabetes of the prototype system. The validated knowledge is represented in the form of rules by rule-based representation technique and the rules are codified to the knowledge base of the prototype system using integrated Clisp.net Prolog programming language. The rules used for development of the prototype are generated using data mining techniques that the test instances scores more than 95 % of accuracy as discussed in Chapter 3.

4.5. KNOWLEDGE REPRESENTATION

The most commonly used methods of knowledge representations are production rule, frame and network. Knowledge captured from experts and other sources must be organized in such a fashion that a computer inference program enables to access this knowledge whenever needed and draw conclusions. In this prototype KBS, production rules are used, since it permits the relationships that make up the knowledge base to be broken down into manageable units.

A rule is a correlation found between the main variable (dependent) and the others (independent). IF this condition (or antecedent-condition or premise) occurs, THEN some action (or conclusion or consequence) will occur. The following rules in the knowledge base of the prototype are expressed with natural language rules IF … THEN … The corresponding rules extracted form J48 decision trees listed as follows.

Rule-1: if ((Diabetes Symptoms = Healthy or Diagnosis=Healthy): Free From Diabetes

Rule-2: if ((Diabetes Symptoms = Diabetes or Diagnosis=Diabetes): Got Diabetes

Rule: If ((patients Age <= 40) and (Diabetes Riskfactor = Healthy) and (patients Age <= 34) and (patients Age <= 28) and (BMI = Healthywieght) and (patients Age >13) and (BP >130) and (FBS1<=130) and (patients Age <=23) and (FBS1 > 150) and (FBS1 <= 166)): PreDm (17.04/8.16)

Rule-3: If ((patients Age <= 40) and (Diabetes Riskfactor = Healthy) and (patients Age <= 34) and (patients Age > 28) and (patients Status = Male) and (BP <= 165) and (patients Age <= 29)): AtRiskTypeIDm (4.01/0.01) or if (patients Age > 29): PreDm (68.3/4.07) or if (BP > 165): AtRiskTypeIDm (12.2/1.09)

Rule-4: If ((patients Age <= 40) and (Diabetes Riskfactor = Healthy) and (patients Age >34) and (BP <= 125) and (patients Status = Male)): TypeIDm (47.81/2.89) or if (BP> 25): AtRiskTypeIIDm (12.94/1.57)

Rule-5: If ((patients Age <= 40) and (Diabetes Riskfactor = AtRisk) and (patients Age >20) and (patients Age <= 37) and (FBS1 >130) and (patients Age <= 32) and (patients Status = male)): AtRiskTypeIIDm (13.41/2.19)

Rule-6: If (patients Age > 40) and (Diabetes Riskfactor = Healthy) and (patients Age > 47) and (patients Status = male) and (FBS2 <= 135 and (patients Age <= 52)): TypeIIDm (28.9/0.85)

Rule-7: If (patients Age > 40) and (Diabetes Riskfactor = Healthy) and (patients Age > 47) and (patients Status = Male) and (FBS2 <= 135) and (patients Age >52) and (patients Age <= 55)): PreDm (39.73/3.24) or if (patients Age > 55): TypeIIDm (12.0/2.82)

Rule-8: If (patients Age > 40) and (Diabetes Riskfactor = Healthy) and (patients Age > 47) and (patients Status = Male) and (FBS2 > 135) and (patients Age <= 69)): TypeIIDm (211.69/12.87) or if (patients Age > 69): AtRiskTypeIIDm (3.0)

Rule-9: If ((patients Age <= 40) and (Diabetes Riskfactor = Healthy) and (patients Age >34) and (BP > 125) and (patients Status = Femalenonpreg) and (FBS1 <= 155) and (FBS1 <= 137)): AtRiskTypeIIDm (9.05/0.86) or (FBS1 > 137)): TypeIDm (4.42/0.14) or FBS1 > 155: PreDm (12.17/1.04)

Rule-10: If ((patients Age <= 40) and (Diabetes Riskfactor = AtRisk) and (patients Age >20) and (patients Age <= 37) and (FBS1 >130) and (patients Age <= 32) and (patients Status = Femalenonpreg)): PreDm (2.19/1.1)

Rule-11: If (patients Age > 40) and (Diabetes Riskfactor = Healthy) and (patients Age > 47) and (patients Status = Femalenonpreg) and (patients Age <= 70)): TypeIIDm (131.14/8.11) or if (patients Age > 70): AtRiskTypeIIDm (3.55/0.01)

Rule-12: If ((patients Age <= 40) and (Diabetes Riskfactor = Healthy) and (patients Age <= 34) and (patients Age > 28) and (patients Status = Femalepreg)): GestationalDm (14.62/1.11)

Rule-13: If ((patients Age <= 40) and (Diabetes Riskfactor = Healthy) and (patients Age >34) and (BP > 125) and (patients Status = Femalepreg)): AtRiskgestationDm (4.43/2.02) or if (patients Status = Femalepreg): GestationalDm (34.73/1.9)

Rule-14: If ((patients Age <= 40) and (Diabetes Riskfactor = AtRisk) and (patients Age >20) and (patients Age <= 37) and (FBS1 >130) and (patients Age > 32) and (patients Status = Femalepreg)): AtRiskgestationDm (10.28/0.22)

Rule-15: If (patients Age > 40) and (Diabetes Riskfactor = Healthy) and (patients Age > 47) and (patients Status = Femalepreg): GestationalDm (13.91/0.91)

In the above rule 1, the system requests the identifying by using prefilling information about patients and submit it to diagnosis the patients whether patients have diabetes or healthy.

4.6. DIAGNOSIS AND TREATMENT

Diagnosis of diabetes is based on several cases like patient‘s physical exam, presence or absence of symptoms, medical history, risk factors, blood test reports etc. Blood tests can be used to confirm a diagnosis of diabetes, based on the amount of glucose found. Urine test can also be used to check protein in the urine that may help diagnose diabetes. These tests also can be used to monitor the disease once the patient is on a standardized diet, physical exercise, oral medications, or insulin therapy. The system can provide necessary information about the indications, diagnosis and primary treatment advices to the diabetics. Prototype KBS developed using both automatic extracted and expert based knowledge used for diagnosis and treatment of diabetes as shown below using Mockler Chart. This Mockler Chart of diagnosis and treatment has been drawn to show the relation of components tests, patient's situation, patient's age, Body Mass Index (BMI), symptoms and risk factors. The used Mockler Chart of symptoms, the questions and choices related to determining of the patient's symptoms which concluded diabetes or further analysis of the patients.

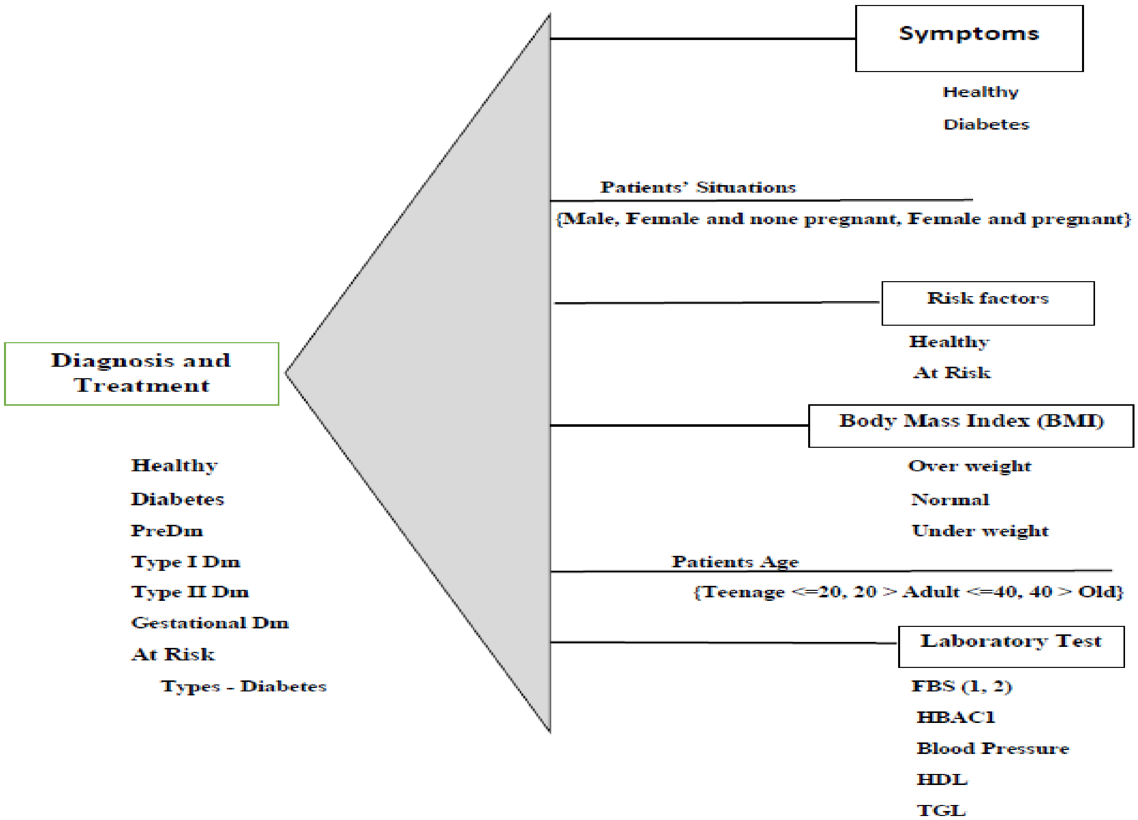

Figure 4.3: Mockler Chart shows relations how KBS used for diagnosis and treatment of diabetes mellitus

4.7. THE INFERENCE ENGINE

The inference engine simulates the domain expert reasoning process. It matches the goal by searching through knowledge base to find rules whose premises match with the given facts in knowledge base. The searching process continues until the inference engine unable to match any premise with the facts in the working memory. As the researcher discussed before, the system have used forward chaining reasoning mechanism. Forward chaining is one of the two main methods of reasoning when using an inference engine and can be described logically as repeated application of modus ponens. Forward chaining is a popular implementation strategy for knowledge base systems.

Forward chaining starts with the available symptoms and uses inference rules to extract more data (from an end user) until a goal of the diagnosis result is reached. An inference engine using forward chaining searches the inference rules until it finds one where the symptoms are known to be true. When such a rule is found, the engine can conclude, or infer, the diagnosis result, resulting in the addition of new information to its symptoms and decide treatment.

3.8. USER INTERFACE

It is the means by which the user and a computer system interact, in particular the use of input devices and software. The acceptability of a knowledge based system depends on the quality of the user interface. The user interface is used as the means of interact user and the knowledge based system. In this study user interface interact facilitate diagnosis and treatment based on predefined Knowledge based contain the rules.

Explanation Facility

The prototype system can describe “what” a request to repeat for clarification before it reached on its conclusions. This ability is usually important since the type of problems to which knowledge-based systems are carried out need an explanation of the result delivered to the end-users. It has also the ability of justifying “why” a certain problem is being questioned in order to reason out it mean and its benefits. The developer of the system included “what” and “why” explanation facilities in problem solving. Moreover, the developer considers the explanation facilities included in the system are easily understandable by the end-users. The developed prototype KBS facilitate reasoning and describing mechanism which made the system more user friend in addition to graphical user interface.

Login Form: The system provide the user login screen form automatically when user initiated the system. This serves as an introductory screen to the application where user is expected to supply username and password for authentication. After the user fill required information of authentication then click ok as shown figure 4.4.The system automatically display the home screen as shown figure 4.5.

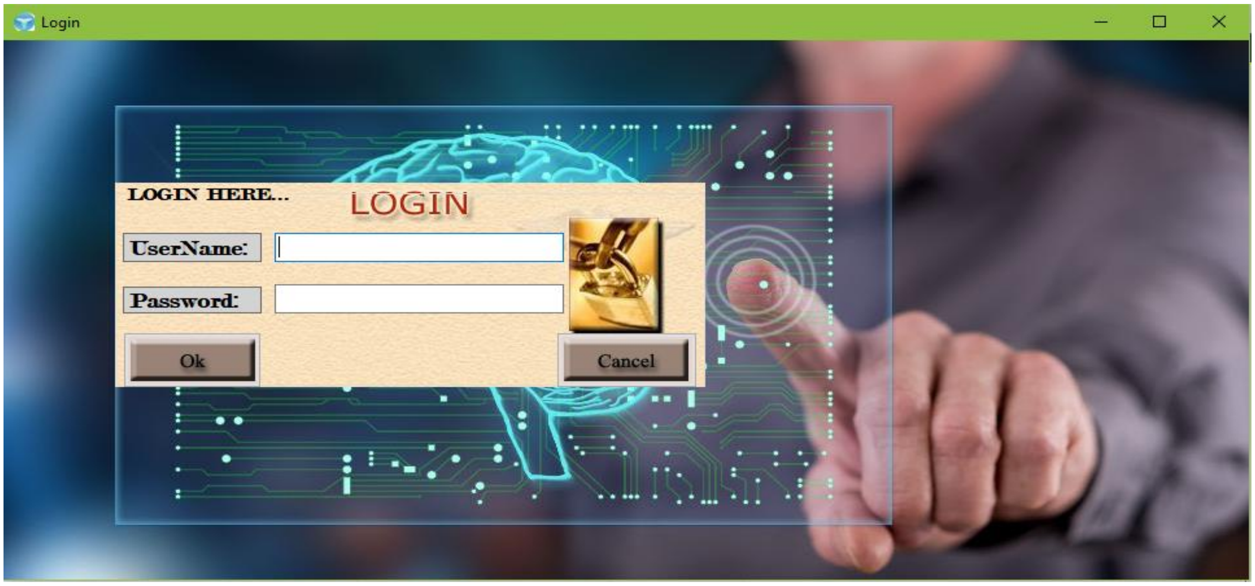

Figure-4.4: login screen user interface

Home Form: home screen provide different tasks that help the user for diagnosis and treatment of the diabetes. This allows the user by providing what user want to do by referencing it.

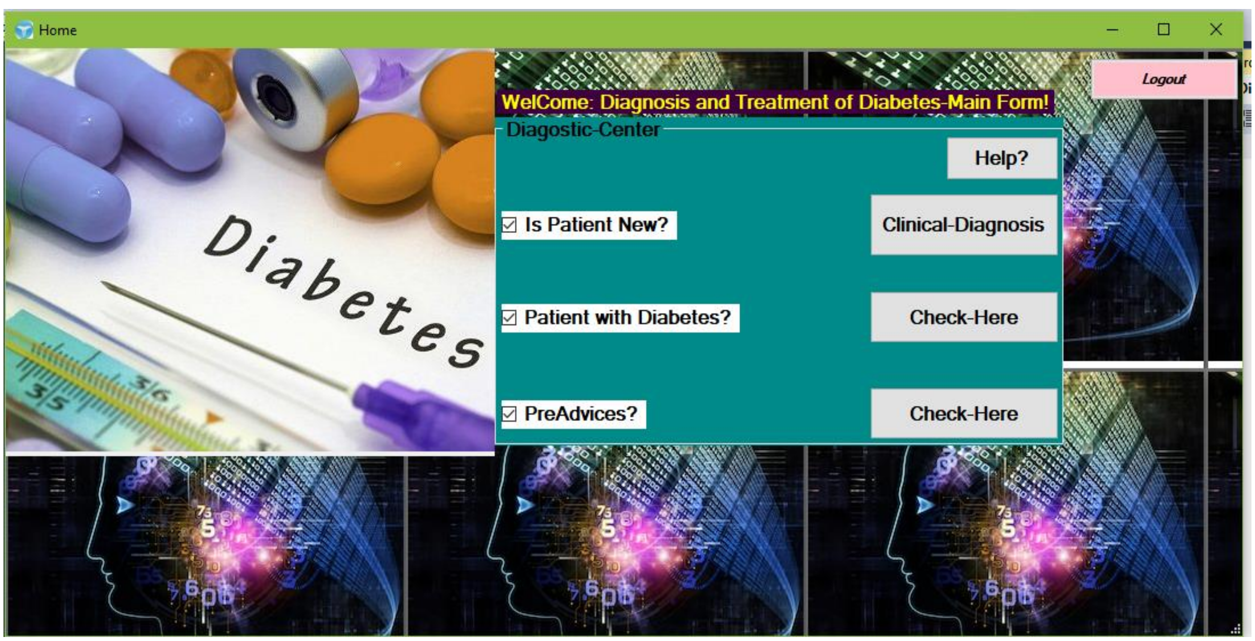

Figure-4.5: home screen user interface of the proposed system

From home screen the user can test healthy before advices the normal person. Pre-advices used healthy diagnosis result for protection and advice as shown figure below.

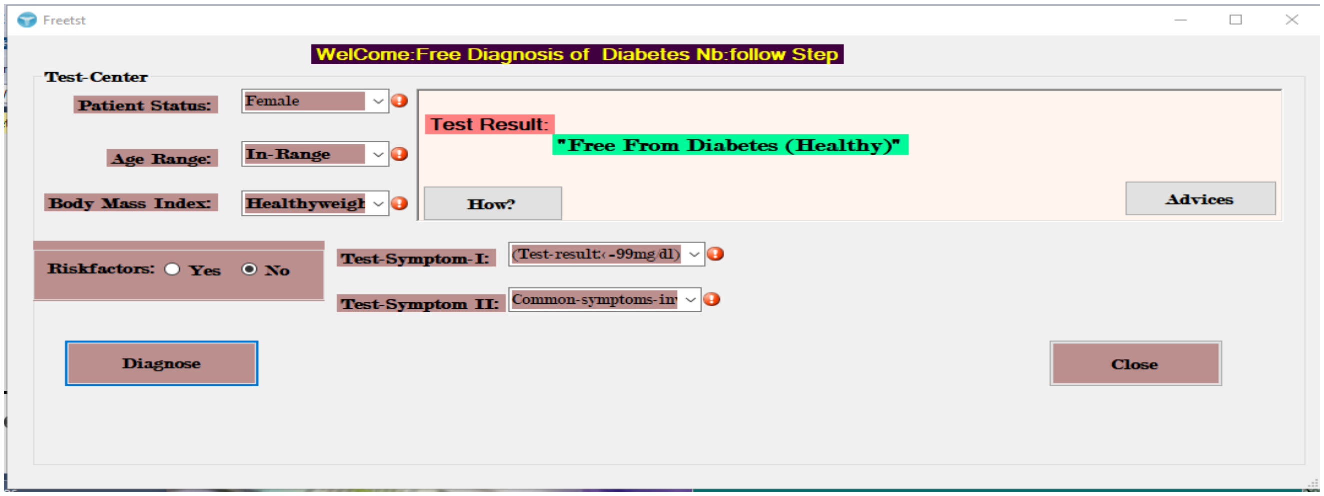

Figure-4.6: user interface used identify whether patients healthy

As observe from the figure 4.5 here above, the user has three options on home user interface. To enter her/his option from the available three options, the only thing that the user has to do is check the checkbox and click on the button which is next side of it and then after based on the user choice the system displays the next step. For instance, if the user wants to conduct clinical diagnosis, Figure 4.7 will be displayed as follows:

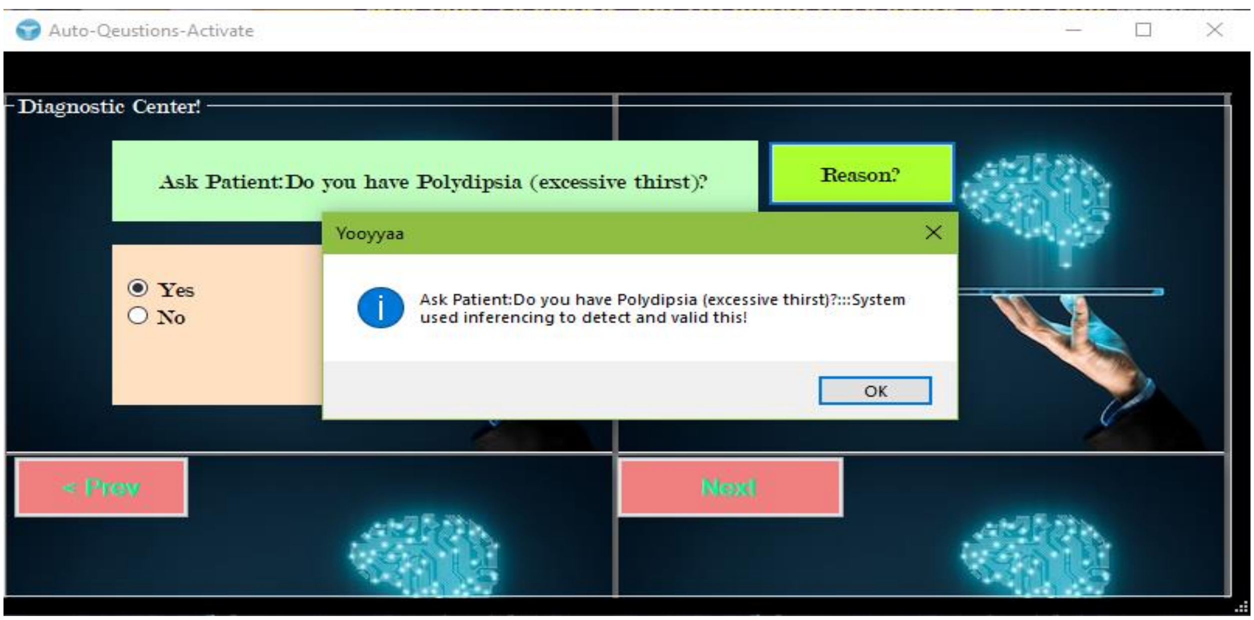

Figure-4.7: user interface used to identify patients diabetes using questions

The system display questions in banner label on form automatically in form of question as sequences manner by including symptoms associated with a diabetic patient. The form has been kept simple in terms of description of symptoms and response to each symptom in Yes/No format then next will provided for the user by end Yes. As shown figure 4.7 above the developed prototype system functions by asking questions to the new patient who came for diagnosis and treatment of diabetes.

First System ask questions as above figure shown in order to identify whether patient had diabetes before send patient for a laboratory blood test to decide whether the patient is diabetic. For diabetic patient, in order to identify type of diabetes the system order further diagnosis by activated further test message before restart itself and ready for new automatic diagnosis progress ready.

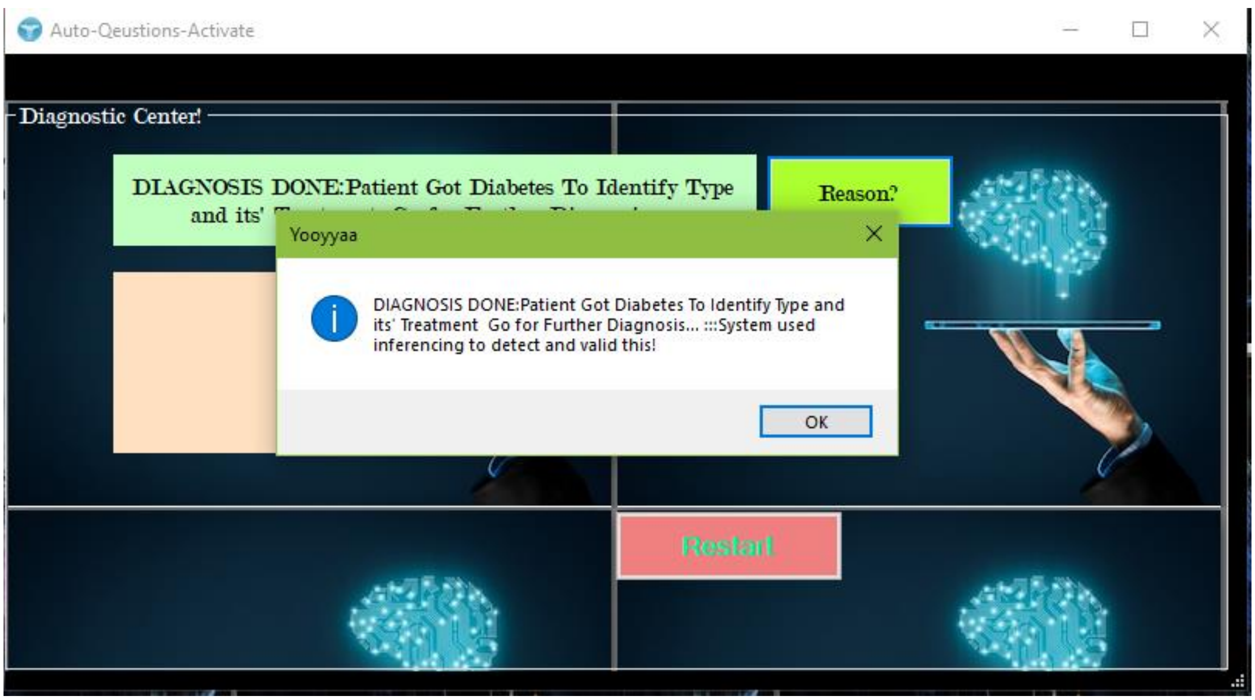

Figure-4.8: User interface detects patients got diabetes using questions

The prototype system to decide whether the patient has specifically prediabetes, Type I Diabetes Mellitus, Type II Diabetes Mellitus, Gestational Diabetes Mellitus or other type of diabetes. The system automatically provide user interface of the patients' symptoms in order to made decision more clear by remembering the user from what previously automatically diagnosed by asking questions provided.

As shown below figure 4.9 the system provide information about diabetes symptoms used to identifying diabetes type in detail with reason before order users proceeded to laboratory test.This help the user to easily identify and confident in decision making of targeted diabetes. After the user checked radio button (for instance Test-II-Symptoms) as realized from figure below then press (click) button (i.e. Lab-Test) automatically system provide/display diagnostic page as shown figure 4.10.

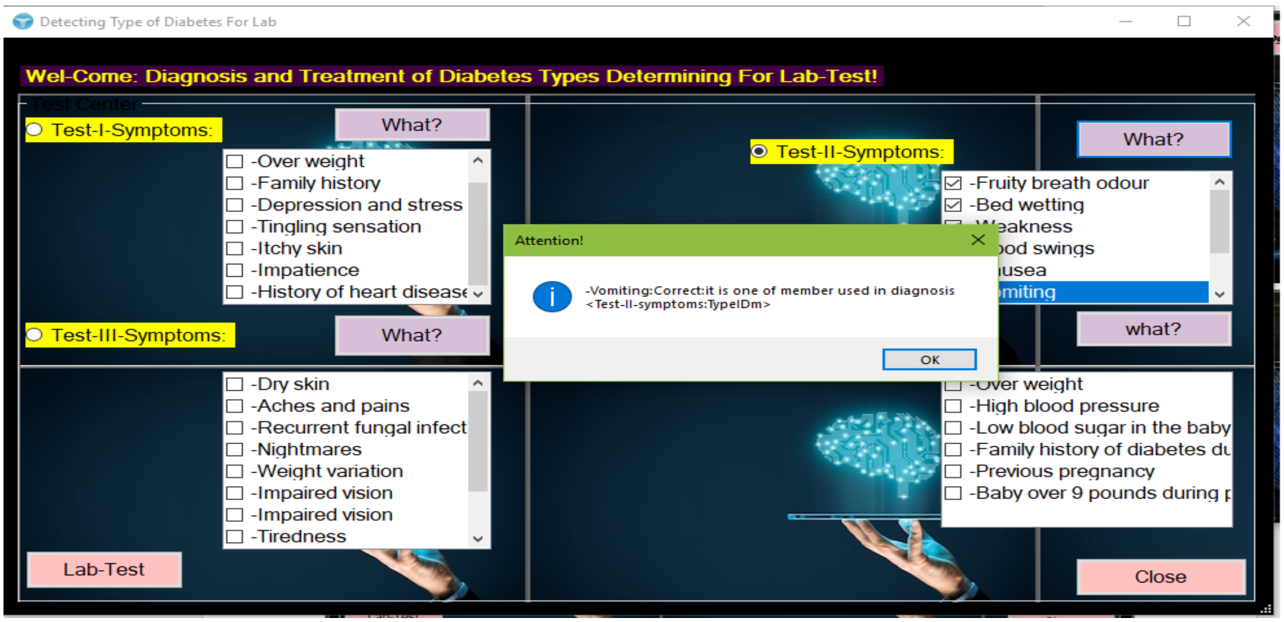

Figure-4.9: the user interface provide information about diabetes symptoms before lab-test

For laboratory test the prototype KBS contemplates the Patient status, concept of age, Body Mass index, risk factors ,criteria of diabetes such as the, Fasting Blood Sugar, 2 h glucose oral glucose tolerance test,HbAc1, family history of diabetes and/or obesity, and ketone and/or autoantibodies etc..

Besides, if the previous diagnosed test result of a certain patient shows the patient is type I diabetes and when the patient wants to diagnose again, the test result shows the patient has diabetes. As shown figure below the result displayed after all requirement information fulfilled otherwise the result is not perfect or error loaded. After “Test Result” displayed the prototype system provide “How” which answer how “Test Result”? displayed and Treatment which proceed next to diagnosed diabetes as shown figure 4.11 it recommends treatments for patients based on the type of diabetes that the patient has been diagnosed.

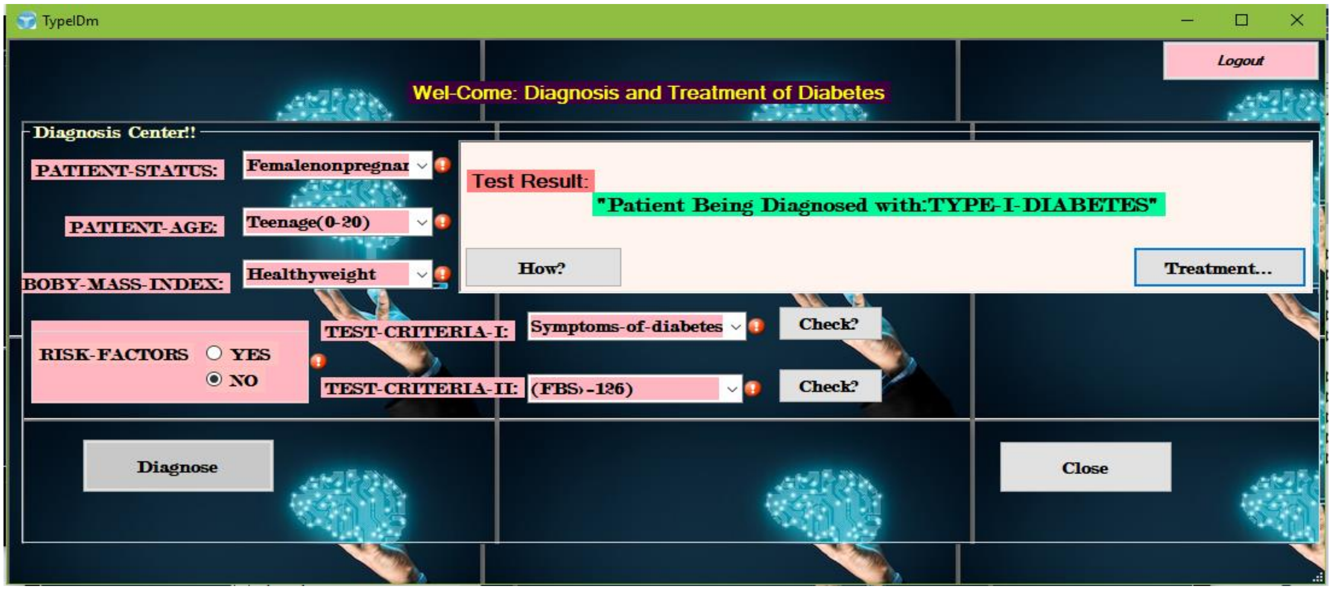

Figure-4.10: Graphical User Interface for Clinical Diagnosis

After identifying the diabetes type (i.e. similar to diagnosed Type I Diabetes), the system recommends the treatment facilities such as the diet information, medication therapy, exercise and foods to avoid or limit. Moreover, the system advices the patient to monitor his/her glucose level either using FPG or 2h OGTT or hemoglobin A1C test methods. FPG test enables the patient to control his/her glucose level daily and hemoglobin A1C is used to measure the average sugars level that is accumulated in the patient's blood for the last 2-3 months by looking at the level of sugar in the patient's hemoglobin. As a result, if diabetic patients get proper treatments, they can control their sugar level, which will further enable them control more complications of diabetes.

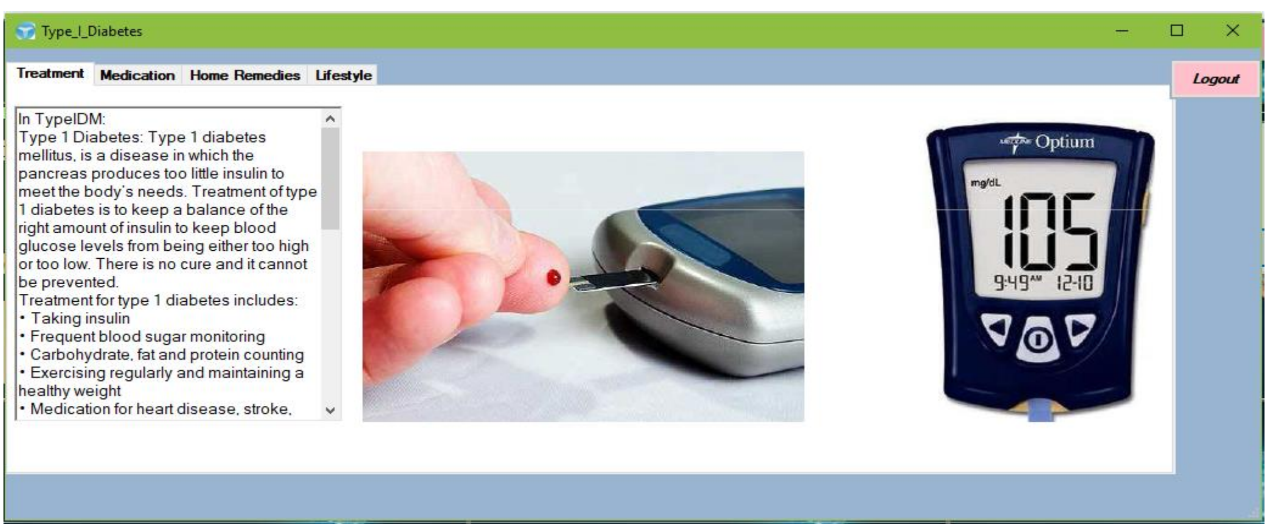

Figure-4.11: Graphical User Interface for clinical treatment

CHAPTER FIVE
TESTING AND PERFORMANCE EVALUATION OF THE PROTOTYPE

5.1. OVER VIEW OF SYSTEM TESTING AND EVALUATION

Testing and performance evaluation is an important issue for every intelligent system (i.e. Knowledge Based System). The purpose of the evaluation process is to get the end user‘s views on the significance or usefulness of the system. The evaluation and testing issue of the system answers the question “To what degree the effective and efficient prototype knowledge based system give acceptable and validate diagnosis and treatment service to users?” To answer this question, system is valid and user acceptance methods are used.

5.2. SYSTEM VALIDITY TESTING USING TEST CASES

System performance testing is applied to evaluate the performance of the KBS which helps to compare and contrast the domain expert’s judgment and the proposed system’s response. Test cases are one of the predominant evaluation mechanisms for evaluating the performance of the proposed system which helps to compare and contrast the domain experts judgment and the proposed system’s response so that conclusions could be made on whether the proposed system could work in the absence of domain expert or not.

For the purpose of testing the prototype Knowledge based system (KBS), four domain experts are selected from Jimma Specialized Hospital. The selected experts are professionals who work in the diabetes related disease and participated in knowledge acquisition phase as well as in the visual interaction of the system. From the total of 25 cases collected from Jimma Specialized Hospital, 23 patient’s cases are diagnosed correctly and 2 patient cases are diagnosed incorrectly by the KBS prototype system of the prototype KBS. Diabetes patient test cases were distributed equally for each expert evaluator. Evaluators start evaluation by testing all possible inputs and validate their subsequent outputs. The expert evaluators identify correctly and incorrectly diagnosed cases by comparing the outputs made by the KBS system with that of the experts’ decision on the same patient cases. Based on the Collected and applied to the system the diagnosis performance result:

$$\text{Diagnosis Accuracy } = \frac{\textit{Total number of Correct Test Cases}}{\textit{Total number of Test Cases}} \times 100$$

$$\text{Diagnosis Accuracy } = \frac{23}{25} \times 100$$

$$\text{Diagnosis Accuracy} = 0.92 * 100 = 92\%$$

5.3. USER ACCEPTANCE TESTING

For testing the user acceptance of a system, a questionnaire is prepared to evaluate the user acceptance of a system and the evaluators fill the questionnaire after they have used the system. Testing is used to identify whether the developed system is achieve its minimal goal or not, because one of the objective of this research is to test the user acceptance and the performance (i.e. validation) of the system. The system testing is made on its accuracy for making KBS identification, so it is measured by the domain experts in the field of domain area. So the researcher has developed some questions that are answered by the domain experts during interacting with the system. For evaluation purpose 4 domain experts have been selected purposively from Jimma Specialized hospital, and they have tried to evaluate the system by discussing with themselves and have given their recommendation on the system. Totally questions have been prepared and there are five levels for the evaluation; these are **poor, fair, good, very good, and excellent,** and have **1, 2, 3, 4** and **5** respective values. In terms of percentage, 5 is given 100%, 4 is equal to 80%, 3 is equal to 60%, 2 is equal to 40% and 1 is equal to 20%. These values have given by the domain experts, during interaction with the system.

Table 5.3 user acceptance evaluation form

Questions	1-Poor	2-Fair	3-Good	4-Very good	5-Excellent	Percent (100%)
Is the system is easy to understand and use?					✓	100
Is the user interface is attractive for use?					✓	100
Is it accurate in prioritizing the patients based on their symptoms?					✓	100
Does the system is important for the domain experts?				✓		80
Does the system have significance contribution for the domain area?				✓		80
Is the system effective in time?					✓	100
How do you get the system will be useful as a training tool?				✓		80
Total Average Evaluation Result						**91.43%**

As the domain experts have evaluated the system on their point of view, they have rated **91.43%** effective and efficient. At end, they have given their own recommendation both on the further improvement of the system and on the strength of the system. According to the system evaluators mention on prototype knowledge based system as follows:

- The system can used in health sector where there is scarcity of domain experts.
- The system has reduced the required manpower in the health department, reduce the time lost by the patient, serve where lack of domain expert and minimize the crowdedness of patients and consistent decision have been given by the system for similar symptoms, it does not give different decision for the same case. It was better if the system update itself automatically when new case (symptoms) added and support local language.

5.4. DISCUSSION

The purpose of this research is to develop prototype KBS for diagnosis and treatment of diabetes using data mining techniques. The use of data mining techniques to build the knowledge base of the KBS can be taken as strong features of the system. The findings are discussed in this section. The main aim of data mining is classifying the attribute based on the given attribute. This is achieved by decision trees even though three algorithms are selected for this purpose. Classification maps data into predefined groups. It is often referred to as supervised learning as the classes are determined prior to examining the data. Classification Algorithms usually require that the classes be defined based on the data attribute values. They often describe these classes by looking at the characteristics of data already known to belong to class. Pattern Recognition is a type of classification where an input pattern is classified into one of the several classes based on its similarity to these predefined classes(Rajesh & Sangeetha, 2012). The rule that is discovered from (J48) decision tree is the attempt of finding the rules which used to develop prototype KBS. For this sake the previous rule which mined out by the J48 Classifier is the result of this study which correct out the knowledge used for prototype KBS to diagnosis and treatment of diabetes.

On this study different types of situation were conducted for the purpose of developing prototype KBS for diagnosis and treatment of diabetes, as well which rules(knowledge), which model and which algorithm would perform very well and it is approved under this study.

For this study three algorithm were selected to test on the diabetic datasets in order to generate rules i.e. J48, PART and JRip algorithms. Three of them are raised under the methodology of this study. Therefore, analyzing one by one and seeing the result that they performed during the previous experiment has been tabularized accordingly. Additionally the J48 algorithm is the most performing model more than the rest of the algorithm. And the other algorithm had been resulted according to the nature and ability of evaluation based on the algorithm is set by default.

The J48 algorithm is the most accurate model from the other due the result that this algorithm demonstrated in case of performance, time, labeling, specificity and confusion Matrix. From the previous situation the J48 algorithm had scored a time of 0.31 seconds to classy the 2512 records according the class they belongs too. Beside this, the model also showed the good performance more than the other. The ROC which this model displayed is almost approximate to one which is 0.997 and the result of precision and recall (0.949 and 0.952) also pretty well more than the left model. This model showed the most performing one in case of diabetic datasets classifying.

This is proved on the Table 5.3 result of J48 model. This model scored the accuracy of 95.1515 % to classify the datasets. The result is without any bias means the performances accuracy that this model scored proves how well the J48 algorithm. The rule that is generated from the J48 algorithm is the best rule which used for development of prototype KBS for diagnosis and treatment. Decision Trees are built from nodes, branches and leaves that indicate the variables, conditions, and outcomes, respectively.

The Second most performing model is the JRip Classier or model which is the second one according to the above criteria (i.e. performance). This model performed the third promising result next to the JRip algorithm. This model scored the 94.7348 % accuracy on the general data to classify the status of diabetic patient datasets. The time taken to perform the general data by this algorithm is 2.69 seconds and to classify the 2501 instances of the records. The precision (0.95) and recall (0.947).This result is the most promising result next to J48 algorithm by understanding the experiment result of the model.

The third most performing model is the PART pruned model which is the third one according to the above criteria (i.e. performance) which is almost very close to the JRip classifier. This model performed the third promising result next to the JRip algorithm. This model scored the 94.5076 % accuracy on the general data to classify the status of diabetic patient datasets. The time taken to perform the general data by this algorithm is 0.69 seconds and to classify the 2495 instances of the records. The precision (0.944) and recall (0.945).This result is the most promising result next to JRip algorithm by understanding the experiment result of the model.

Generally, the J48 model is the most performing model with a good accuracy of results. The JRip rule induction is the second most performing model next to J48 model whereas the PART is the last performed classifier. Among these algorisms J48 algorithm is the best performing model by classifying diabetic patient datasets and generate rules. As discussed the performance of J48 is better than JRip and PART algorism as a result the rules generated by J48 model was used for prototype KBS for diagnosis and treatment of diabetes. Tacit knowledge about the diagnosis and treatments of diabetes is extracted from the domain experts using interviewing method in order to have detail understanding of the domain knowledge, it is challenging to extract the necessary knowledge due to the domain expert's personal nature of tacit knowledge. In order to handle this challenge data mining model was used by applying the model on diabetic patient dataset that was collected from listed above hospital. By compare and contrast the model performance accuracy the model with high performance during extract knowledge used for prototype KBS develop. However, medical datasets nature, occurrences of errors in data, poor organized, available of data etc. are preprocessing datasets challenge.

Throughout coding the represented knowledge about diabetes using the Clisp.net, the facts base of the prototype system is able to update its knowledge automatically by get and set method. However, the researcher encountered a challenge to update the rules of the knowledge base of the prototype system dynamically because of rule generated automatically and manually (i.e. domain experts and referring documents) was based on the datasets which was from diabetic patients' baseline data which made differences. Training was given to the domain experts on how the system functions and on how to use and interact with the system. However, from four evaluators, all of them were fulfilled by the prototype system. The two evaluators commented that, better if system use local language and also want the decisions provided by the system, if local language so as to understand the decisions made by the system or option. In general, the testing and evaluation outcomes of the prototype system has achieved the objectives of the study. However, additional study is needed to bring complete implementation and use of effective and efficient knowledge-based system for diagnosis and treatment of diabetes.

CHAPTER SIX

CONCLUSION AND RECOMMENDATIONS

The previous chapters have brought to light some significance issues in the design of the prototype knowledge-based system for diagnosis and treatment of diabetes. In this part, the researcher concludes the study work and gives recommendation for future investigation in the health and medicine area.

6.1. CONCLUSION

Diabetes is a metabolic condition that leads to high blood sugar levels. It is a kind of disease in which the body does not produce or properly use insulin. The amount of glucose in blood is too high because the body cannot use it properly. This is because pancreas does not produce any insulin, or not enough, to help glucose enter patient body's cells or the insulin that is produced does not work properly. Insulin is the hormone produced by the pancreas that allows glucose to enter the body's cells, where it is used as fuel for energy so we can work, play and generally live our lives.

Due to this reason, patients need diagnosis and consistent treatment. But, in our country, there are no sufficient numbers of domain experts. This situation leads to disproportional numbers of experts and patients also available resource, urbanization and high obesity, lack of know how. As a result, diabetic patients are not getting enough diagnosis and treatment. Hence, in this study an effort has been made to design and develop a prototype of effective and efficient knowledge-based system that can provide advice for experts and patients to facilitate the diagnosis and treatment of patients living with diabetes.

Knowledge based systems can help a great deal in decision making through a display of intelligent behavior that may include learning and reasoning. In developing the prototype system, knowledge is acquired using both structured and unstructured interviews with domain experts and from relevant documents by using documents analysis method to find the solution of the problem. In this research, effective and efficient KBS that supports diagnosis and treatment of diabetes disease was developed by using data mining techniques as a knowledge acquisition step. The aim of using data mining techniques with the KBS is to reduce the difficulty caused by the 'knowledge acquisition bottleneck' and to obtain low-cost and high-quality knowledge base. The generated

knowledge is represented using rule-based representation technique and codified using clisp.net editor tool for building the knowledge-based system for diagnosis and treatment of diabetes.
For testing and evaluation of the prototype system, 25 cases of patients are selected using purposive sampling method in order to test the accuracy of the prototype system. The correct and incorrect results are identified by comparing decisions made by the domain experts on the cases of patients and with the conclusions of the prototype system. And also the process of ensuring that the prototype system satisfies the requirements of its end-users is validated. This permits end-users to test the prototype system by actually using it and evaluating the benefits received from its use. As the testing result show, the overall performance of the prototype system registers 92%. The use of data mining techniques to build the knowledge base of the KBS and graphical user interface based is strong features use of the system. However, the system lacks to update rules in the knowledge base of the KBS dynamically and automatic from database is feature work.

6.2. RECOMMENDATIONS

Data mining techniques was applied on diabetic patients' baseline datasets in order to generate rules used for developing prototype KBS for diagnosis and treatment. But, diabetic datasets are manually stored witch made preprocessing datasets difficult. For future if such challenge solved automatically rule generating from database and integrating to knowledge base the rule generated with data mining techniques become easy. A method must be investigated on how to integrate the prototype system with the existing health information systems. This would lead to the development of standards applicable to all, enabling suitable information exchange and planning for additional improvement of functionality. Since such types of systems developed in English language, the applicability of the system is limited, so the researcher recommended that the system will have more applicability if it can be developed in local languages.

REFERENCES

Abetwe, D. A., & A. (2009). Addis Ababa University Faculty of Informatics Department of Information Science Application of Knowledge Based System for Woody Plant Species Identification Dejen Alemu Abetwe A thesis submitted to the school of Graduate Studies of Addis Ababa University.

Abuhay, T. M., & Tesema, T. B. (2015). International Journal of Advanced Research in Computer Science and Software Engineering Integrating Data Mining Results with the Knowledge Based System for Diagnosis and Treatment of Visceral Leishmaniasis. Researchgate, 5(5), 128–142.

Adlassnig, K. (2001). The Section on Medical Expert and Knowledge-Based Systems at the Department of Medical Computer Sciences of the University of Vienna Medical School, 21, 139–146.

Ahmed, I. M., Mahmoud, A. M., Aref, M., & Salem, A. M. (2012). A study on Expert Systems for Diabetic Diagnosis and Treatment. Advances in Information Science, (1), 363–367.Alphafarm. (2009). Talking diabetes No.29.

Alemu Degen, 2009.Application of Knowledge Based System for Woody Plant Species Identification, Addis Ababa: AAU.

Asghar, S. & Iqbal, K. (2009). Automated data mining techniques: A critical literature review. IEEE Proceedings of the International Conference on Information Management and Engineering, , IEEE Xplore Press, Kuala Lumpur, pp: 75-79. DOI: 10.1109/ICIME.

Avneet Pannu. (2015). Artificial Intelligence and its Application in Different Areas. Engineering and Innovative Technology, 4(10), 79–84.

Bhambri, V. (2011). Application of data mining in banking sector. International journal of computer science and technology, 2(2), 199-202. Retrieved 20June, 2015, from https://www.academia.edu/1866998/Application_of_Data_Mining_in_Banking_Sector.

Chala Diriba, Million Meshesha, D. T. (2016). Computer Science. IJCSI International Computer Science, 13(4), 108–112. https://doi.org/10.20943/01201604.108112

Choubey, D. K., Paul, S., & Dhandhenia, V. K. (2017). Rule based diagnosis system for diabetes . Biomedical Research, 28(12), 5196–5209.

Christopher, J. J., Nehemiah, K. H., & Arputharaj, K. (2016). Knowledge-based Systems and Interestingness Measures : Analysis with Clinical Datasets. Computing and Information Technology, 24(1), 65–78.

Dagnachew, M. (2016). Premedical Triage Using Fuzzy Inference Rule Based Approach : The Case Of Menelik Ii Hospital School of Graduate Studies , College of Information Science Department of Information Science. AAU, (January).

El-korany, A., & Rafea, A. (1998). A Structured Testing Methodology for Knowledge-Based Systems, 1–15.

Garber, A. J., Handelsman, Y., Einhorn, D., Bergman, D. A., Bloomgarden, Z. T., Fonseca, V.,Nesto, R. W. (2008). ACE / AACE Consensus Statement Diagnosis and Management of Prediabetes in the Continuum of Hyperglycemia — When Do the Risks of Diabetes Begin ? A Consensus Statement From the American College of Endocrinology and the American Association of Clinical Endo. In Statement Diagnosis and Management of Prediabetes (Vol. 14, pp. 933–946).

Gary R. Morrison. (2014). Experimental Research Methods, (January 2003).

Gizaw, M., Harries, A. D., Ade, S., Tayler-Smith, K., Ali, E., Firdu, N., & Yifter, H. (2015). Diabetes mellitus in Addis Ababa, Ethiopia: admissions, complications and outcomes in a largereferral hospital. Public Health Action, 5(1), 74–78. https://doi.org/10.5588/pha.14.0107

Hernández-medrano. (2011). Knowledge-Based System Model to Support Diabetes Research and Clinical Process Modelo de Sistema Basado en Conocimiento para Apoyo en la Investigación y Proceso Clínico de la Diabetes, XIII(2).

Ian.H.Witten, E. F. 2011, Data Mining: Practical Machine Learning Tools and Techniques,
Newyork: Morgan Kaufmann.

Ionu Taranu. (2015). Data mining in healthcare: decision making and precision. Database Sysetm, VI(4), 33–40.

Jiawei Han andMicheline Kamber. (2006). DataMining: Concepts and Techniques, Second Edition JiaweiHan.

Kaiser. (2017). Type 1 Diabetes Treatment Guideline Major Changes as of May 2017.

Kulani Makhubele. (2012). Report A Knowledge Based Expert System for Medical.

Kumar, R., & Verma, D. (2012). Classification Algorithms for Data Mining: A Survey. International Journal of Innovations in Engineering, 7–14. Retrieved from http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.305.2167&rep=rep1&type=pdf

Lingaraj, H., Arts, K., Devadass, R., Arts, K., Gopi, V., Arts, K., & Devadass, R. (2015). Prediction Of Diabetes Mellitus Using Data Mining Techniques : A Review. Bioinformatics & Cheminformatics, (March).

Llorens, E., Comas, J., Martí, E., Lluís, J., Sabater, F., & Poch, M. (2009). Integrating empirical and heuristic knowledge in a KBS to approach stream eutrophication, 220, 2162–2172. https://doi.org/10.1016/j.ecolmodel.2009.06.012

Lokanayaki, K. (2013). Exploring on Various Prediction Model in Data Mining Techniques for Disease Diagnosis. International Journal of Computer Applications, 77(5), 26–29.

Management, C., To, A., & It, I. (2008). Management Approach To (Vol. 1).

Ncr, P. C., Ncr, R. K., Spss, J. C., Spss, T. K., Daimlerchrysler, T. R., & Daimlerchrysler, R. W. (1999). The CRISP-DM Process Model. In The CRISP-DM Process Model.

Marta A., Argimiro A. and Ramon X. Forecasting with twitter data. ACM Transactions on Intelligent Systems and Technology (TIST) Special Section on Intelligent Mobile Knowledge Discovery and Management Systems and Special Issue on Social Web Mining, Volume 5 Issue 1, Article No. 8, 2013

M.S. Abdullah, I. Benest, A. Evans, and C. Kimble. 2002. Knowledge Modelling Techniques For Developing Knowledge Management Systems, 3rd European Conference on Knowledge Management, Dublin, Ireland, pp. 15-25.

Nega, Adane, A. K. (2017). Data Mining Based Hybrid Intelligent System for Medical Application. Information Engineering and Electronic Business, (July), 38–46. https://doi.org/10.5815/ijieeb.2017.04.06

Nurses Learning Network, “Laboratory Procedures”, Available at:

http://www.nurseslearning.com/courses/nrp/NRP1605/course/section3/index.htm, Accessed date: December 15, 2012.

P.M. Beulah Devamalar, V. Thulasi Bai, S. K. S. (2008). The new telemedicine paradigm : Fully automated real time web-centric expert system to support diabetes diagnosis. Healthcare Technology Management, 9(January). https://doi.org/10.1504/IJHTM.2008.020203

Perveen, S., Shahbaz, M., Guergachi, A., & Keshavjee, K. (2016). Performance Analysis of Data Mining Classification Techniques to Predict Diabetes. Procedia Computer Science, 82, 115–121. https://doi.org/10.1016/j.procs.2016.04.016

Pillay, K., Maunder, E., & Naidoo, K. (2009). Dietary intake and metabolic control of children aged six to ten with type 1 diabetes mellitus in KwaZulu-Natal. South Africa Journal Clinical Nutrition, 22(2), 95–98.

R P Ambilwade Central, Ramesh Manza, B. P. G. D. (2014). Medical Expert Systems for Diabetes Diagnosis : A Survey International Journal of Advanced Research in Medical Expert Systems for Diabetes Diagnosis : A Survey. Computer Science and Software Engineering, (August 2016).

Rajesh, K., & Sangeetha, V. (2012). Application of data mining methods and techniques for diabetes diagnosis. International Journal of Engineering and Innovative Technology (IJEIT) Volume, 2(3), 224–229.

Roche, M. M., & Wang, P. P. (2014). Factors associated with a diabetes diagnosis and late diabetes diagnosis for males and females. Journal of Clinical and Translational Endocrinology, 1(3), 77–84. https://doi.org/10.1016/j.jcte.2014.07.002

Sajja, P. S., & Akerkar, R. (2010). Knowledge-Based Systems for Development, 1, 1–11.

Solomon Gebremariam. (2013). School of Graduate Studies School of Information Science a Self-Learning Knowledge Based System for School of Graduate Studies. Aau, (January).

Srivastava, S. (2015). Diabetes Mellitus : Diagnosis and Management Guidelines. JIMSA, 28(1), 47–50.

Studer, R., Benjamins, V. R., & Fensel, D. (n.d.). Knowledge Engineering : Principles and

Methods.Science,1–38.Retrievedfrom
http://www.swi.psy.uva.nl/usr/richard/home.html%0AAbstract

Sugiyama, T. (2011). Management of Gestational Diabetes Mellitus. JMAJ, 139(10), 293–300.

Tagel Aboneh. (2013). School of Information Science Knowledge Based System for Pre-Medical Triage Treatment At Adama University Asella School of Graduate Studies.

Tayade, M. (2014). Role of Data Mining Techniques in Healthcare sector in India Scholars Journal of Applied Medical Sciences (SJAMS) Review Article Role of Data Mining Techniques in Healthcare sector in India. Researchgate, (January).

Tripathi, K. P. (2011). A Review on Knowledge-based Expert System : Concept and Architecture. IJCA, 19–23.

Tunmibi, S., Adeniji, O., Aregbesola, A., Dasylva, A., Technology, I., State, K., … State, L. (2013). Manuscript Info. Advanced Research, 1(2320), 343–348.

V. K. D. L. D. M. M. Abbod MF, "Survey of utilization of fuzzy technology in medicine and Healthcare," Fuzzy Sets System, vol. 120, p. 331–349, 2001.

Weiss, S. M., Buckley, S. J., Kapoor, S., & Damgaard, S. (2003). Knowledge-Based Data Mining. IBM.

William T. Cefalu, M. (2017). Standards of Medical Care in Diabetes — 2017 Standards of Medical Care in Diabetes d 2017. The Journal of Clinical and Applied Research and Education, 40(January).

Zeki, T. S., Malakooti, M. V, & Ataeipoor, Y. (2012). An Expert System for Diabetes Diagnosis. American Academic & Scholarly Research Journal Vol., 4(5), 1–13.

https://www.medicinenet.com/script/main/art.asp?articlekey=3590
https://www.britannica.com/topic/information-system

APPENDIXES

Appendix I

Interview questions to Domain Experts

The importance of this interview questions is to extract tacit knowledge from domain expert that will support for the development of a knowledge based system for diagnosis and treatment of diabetes**.** The interviewer records the respondents' response using pen, pencil and paper. I thank you in advance for your willingness and valuable time.

1. What is diabetes?
2. What are types of diabetes?
3. What are the cause and risk factors of diabetes?
4. What are the complications of diabetes?
5. What are the common symptoms of diabetes?
6. What are the most fundamental symptoms that you consider when making diagnosis of diabetes?
7. Is there diagnosis and treatment standard guide line for diabetes?
8. What are the laboratory testing methods used for diagnosis of diabetes?
9. Which one is the most common laboratory testing method used for diagnosis of diabetes in our country?

Appendix II

Developing Knowledge Based System for Diagnosis and Treatment of Diabetes using data mining techniques Prolog Code:

```
;;;=================================================================
;;;CLIPS Integration with Microsoft Visual Basic 2013,2018
;;;=================================================================
;;;------------------ By Kedir Eyasu Abdulkadir,Thesis 2018--------------
;;;~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
;;; * DEFTEMPLATES & DEFFACTS *
;;;~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
(deftemplate UI-state
   (slot id (default-dynamic (gensym*)))
   (slot display)
   (slot relation-asserted (default none))
   (slot response (default none))
```

```
  (multislot valid-answers)
  (slot state (default middle)))
(deftemplate state-list
  (slot current)
  (multislot sequence))
(deffacts startup
  (state-list))
;~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
;Diagnosis Starts Rules with System banner
;~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
(defrule system-banner ""
 =>
 (assert (UI-state (display WelcomeMessage)
                   (relation-asserted start)
                   (state initial)
                   (valid-answers))))
;;;~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
;;;Determining Query Based Diagnos is_engine!;
;;~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
(defrule determine-engine-state ""
  (logical (start))
  =>
  (assert (UI-state (display StartQuestion)
                    (relation-asserted engine-starts)
                    (response No)
                    (valid-answers Yes No))))
(defrule determine-runs-normally ""
  (logical (engine-starts No))
  =>
  (assert (UI-state (display RunQuestion)
                    (relation-asserted runs-normally)
                    (response No)
                    (valid-answers Yes No))))
(defrule determine-Poly-state ""
  (logical (engine-starts Yes))
  =>
  (assert (UI-state (display PolydipsiaQuestion)
                   (relation-asserted engine-rotates)
                    (response No)
                    (valid-answers Yes No))))
```

```
(defrule determine-Weight ""
  (logical (runs-normally Yes))
  =>
  (assert (UI-state (display WeightQuestion)
                    (relation-asserted engine-sense)
                    (response No)
                    (valid-answers Yes No))))
(defrule determine-Fract ""
  (logical (runs-normally Yes))
  =>
  (assert (UI-state (display FractionQuestion)
                    (relation-asserted engine-act)
                    (response No)
                    (valid-answers Yes No))))
(defrule determine-Infect ""
  (logical (runs-normally Yes))
  =>
  (assert (UI-state (display InfectionsQuestion)
                    (relation-asserted engine-call)
                    (response No)
                    (valid-answers Yes No))))
(defrule determine-Experience ""
  (logical (runs-normally Yes))
  =>
  (assert (UI-state (display ExperienceQuestion)
                    (relation-asserted engine-output-low)
                    (response No)
                    (valid-answers Yes No))))
(defrule determine-Tiredness ""
  (logical (engine-starts Yes))
  =>
  (assert (UI-state (display TirednessQuestion)
                    (relation-asserted patient-has-symptom)
                    (response No)
                    (valid-answers Yes No))))
(defrule determine-Conscious ""
  (logical (engine-rotates Yes))
  =>
  (assert (UI-state (display ConsciousnessQuestion)
                    (relation-asserted petients-has-fill)
```

```
                   (response No)
                   (valid-answers Yes No))))
(defrule determine-point-Nausea ""
   (or (logical (engine-starts Yes)))
   =>
   (assert (UI-state (display NauseaQuestion)
                   (relation-asserted point-surface-state)
                   (response No)
                   (valid-answers Yes No))))
(defrule determine-Polyphagia-test ""
   (logical (engine-starts Yes)
          (engine-rotates Yes)
          (petients-has-fill Yes))
   =>
   (assert (UI-state (display PolyphagiaQuestion)
                   (relation-asserted Polyphagia-test-positive)
                   (response No)
                   (valid-answers Yes No))))
;;;~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
;;;* Diagnosis Order Error Correction and Final Result Suggestions *
;;;~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
(defrule normal-engine-state-conclusions ""
   (logical (runs-normally No))
   =>
   (assert (UI-state (display Mistakemessage)
                   (state final))))
(defrule Polyphagia-test-positive-yes ""
   (logical (Polyphagia-test-positive Yes))
   =>
   (assert (UI-state (display Diagnosisresult)
                   (state final))))
(defrule determine-Experience ""
   (logical (runs-normally Yes))
    =>
   (assert (UI-state (display Diagnosisresult)
                   (state final))))
(defrule Is-Error ""
   (declare (salience -10))

   (logical (UI-state (id ?id)))
```

```
  (state-list (current ?id))
  =>
  (assert (UI-state (display Mistakemessage)
                    (state final))))
;~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
;@ Graphical User Interface Control rules *
;~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
(defrule ask-question
  (declare (salience 10))
  (UI-state (id ?id))
  ?f <- (state-list (sequence $?s&:(not (member$ ?id ?s))))
  =>
  (modify ?f (current ?id)
             (sequence ?id ?s))
  (halt))
(defrule handle-next-no-change-none-middle-of-chain
  (declare (salience 10))
  ?f1 <- (next ?id)
  ?f2 <- (state-list (current ?id) (sequence $? ?nid ?id $?))
  =>
  (retract ?f1)
  (modify ?f2 (current ?nid))
  (halt))
(defrule handle-next-response-none-end-of-chain
  (declare (salience 10))
  ?f <- (next ?id)
  (state-list (sequence ?id $?))
  (UI-state (id ?id)
            (relation-asserted ?relation))
  =>
  (retract ?f)
  (assert (add-response ?id)))
(defrule handle-next-no-change-middle-of-chain
  (declare (salience 10))
  ?f1 <- (next ?id ?response)
  ?f2 <- (state-list (current ?id) (sequence $? ?nid ?id $?))
  (UI-state (id ?id) (response ?response))
  =>
  (retract ?f1)
```

```
  (modify ?f2 (current ?nid))
  (halt))
(defrule handle-next-change-middle-of-chain
  (declare (salience 10))
  (next ?id ?response)
  ?f1 <- (state-list (current ?id) (sequence ?nid $?b ?id $?e))
  (UI-state (id ?id) (response ~?response))
  ?f2 <- (UI-state (id ?nid))
  =>
  (modify ?f1 (sequence ?b ?id ?e))
  (retract ?f2))
(defrule handle-next-response-end-of-chain
  (declare (salience 10))
  ?f1 <- (next ?id ?response)
  (state-list (sequence ?id $?))
  ?f2 <- (UI-state (id ?id)
             (response ?expected)
             (relation-asserted ?relation))
 =>
  (retract ?f1)
  (if (neq ?response ?expected)
    then
    (modify ?f2 (response ?response)))
  (assert (add-response ?id ?response)))
(defrule handle-add-response
  (declare (salience 10))
  (logical (UI-state (id ?id)
  (relation-asserted ?relation)))
  ?f1 <- (add-response ?id ?response)
  =>
  (str-assert (str-cat "(" ?relation " " ?response ")"))
  (retract ?f1))
(defrule handle-add-response-none
  (declare (salience 10))
  (logical (UI-state (id ?id)
  (relation-asserted ?relation)))
  ?f1 <- (add-response ?id)
  =>
  (str-assert (str-cat "(" ?relation ")"))
  (retract ?f1))
```

```
(defrule handle-prev
  (declare (salience 10))
  ?f1 <- (prev ?id)
  ?f2 <- (state-list (sequence $?b ?id ?p $?e))
  =>
  (retract ?f1)
  (modify ?f2 (current ?p))
  (halt))
```

Graphical User Interface Code

```
using System;
using System.Collections.Generic;
using System.ComponentModel;
using System.Data;
using System.Drawing;
using System.Linq;
using System.Text;
using System.Windows.Forms;
using Mommosoft.ExpertSystem;
using System.Diagnostics;

namespace MDES
{
   public partial class Form3 : Form
   {
        private Mommosoft.ExpertSystem.Environment _theEnv = new
Mommosoft.ExpertSystem.Environment();
       public Form3()
       {
          ConsoleTraceListener tl = new ConsoleTraceListener();
          InitializeComponent();
          _theEnv.AddRouter(new DebugRouter());
          _theEnv.Load("Rulea.clp");
          _theEnv.Reset();
       }
       protected override void OnLoad(EventArgs e)
       {
          base.OnLoad(e);
          NextUIState();
       }
```

```
        private void OnClickButton(object sender, EventArgs e)
        {
           Button button = sender as Button;
           // Get the state-list.
           String evalStr = "(find-all-facts ((?f state-list)) TRUE)";
           using(FactAddressValuef=
(FactAddressValue)((MultifieldValue)_theEnv.Eval(evalStr))[0])
           {
              string currentID = f.GetFactSlot("current").ToString();

              if (button.Tag.Equals("Next"))
              {
                 Proceed.Visible = false;
                 if (GetCheckedChoiceButton() == null) { _theEnv.AssertString("(next " + currentID
+ ")"); }
                 else
                 {
                    _theEnv.AssertString("(next " + currentID + " " +
(string)GetCheckedChoiceButton().Tag + ")");

                 }
                 NextUIState();
              }
              else if (button.Tag.Equals("Restart"))
              {
                 _theEnv.Reset();
                 NextUIState();
                 Proceed.Visible = true;
                 }
              else if (button.Tag.Equals("Prev"))
              {
                 _theEnv.AssertString("(prev " + currentID + ")");
                 NextUIState();
              }
           }
        }

        private void NextUIState()
        {
           nextButton.Visible = false;
```

```
            prevButton.Visible = false;
            choicesPanel.Controls.Clear();
            _theEnv.Run();

            // Get the state-list.
            String evalStr = "(find-all-facts ((?f state-list)) TRUE)";
            using                    (FactAddressValue                    allFacts                    =
(FactAddressValue)((MultifieldValue)_theEnv.Eval(evalStr))[0])
            {
               string currentID = allFacts.GetFactSlot("current").ToString();
               evalStr = "(find-all-facts ((?f UI-state)) " + "(eq ?f:id " + currentID + "))";
            }

            using                    (FactAddressValue                    evalFact                    =
(FactAddressValue)((MultifieldValue)_theEnv.Eval(evalStr))[0])
            {
               string state = evalFact.GetFactSlot("state").ToString();
               if (state.Equals("initial"))
               {
                  nextButton.Visible = true;
                  nextButton.Tag = "Next";
                  nextButton.Text = "Next";
                  prevButton.Visible = false;
               }
               else if (state.Equals("final"))
               {
                  nextButton.Visible = true;
                  nextButton.Tag = "Restart";
                  nextButton.Text = "Restart";
                  prevButton.Visible = false;

               }
               else
               {
                  nextButton.Visible = true;
                  nextButton.Tag = "Next";
                  prevButton.Tag = "Prev";
                  prevButton.Visible = true;
               }
```

```
            using (MultifieldValue validAnswers = (MultifieldValue)evalFact.GetFactSlot("valid-
answers"))
            {
                String selected = evalFact.GetFactSlot("response").ToString();
                for (int i = 0; i < validAnswers.Count; i++)
                {
                    RadioButton rb = new RadioButton();
                    rb.Text = (SymbolValue)validAnswers[i];
                    rb.Tag = rb.Text;
                    rb.Visible = true;
                    rb.Location = new Point(10, 20 * (i + 1));
                    choicesPanel.Controls.Add(rb);
                }
            }
            messageLabel.Text = GetString((SymbolValue)evalFact.GetFactSlot("display"));
        }
    }

    private void ShowChoices(bool visible)
    {
        foreach (Control control in choicesPanel.Controls)
        {
            control.Visible = visible;
        }
    }

    private RadioButton GetCheckedChoiceButton()
    {
        foreach (RadioButton control in choicesPanel.Controls)
        {
            if (control.Checked)
            {
                return control;
            }
        }
        return null;
    }
    private string GetString(string name)
    {
        return RS.ResourceManager.GetString(name);
```

```
        }

        private void b2_Click(object sender, EventArgs e)
        {
            MessageBox.Show(messageLabel.Text + ":::System used inferencing to detect and valid this!", "Yooyyaa", MessageBoxButtons.OK, MessageBoxIcon.Information);
        }

        private void Proceed_Click(object sender, EventArgs e)
        {
            Diagnosticwindow Diagnosticwindow = new Diagnosticwindow();
            Diagnosticwindow.ShowDialog();
        }

    }
}
```

Clisp Rule to Diagnosis and Treatment Type I Diabetes

```
;; DIABETES DIAGNOSIS and TREATMENT KNOWLEDGE BASED SYSTEM
;; 2018
;; KEDIR
(deftemplate diagnosis
        (slot disease)
        (slot gender )
        (slot age )
        (slot weight )
        (slot risk )
        (slot symptom1 )
                (slot symptom2 )
)
(deftemplate preference
        (slot gender (default 2))
        (slot age (default 1))
        (slot weight (default 1))
        (slot risk (default y))
        (slot symptom1 (default 1))
                (slot symptom2 (default 1))
)

(deftemplate UI-control
```

```
        (slot id)
        (slot message)
)
;; defrules determination
(defrule disease-1
        (preference (gender ?g)(risk n)(age 1)(weight ?w)(symptom1 ?s1)(symptom2 ?s2))
                (test(eq ?g 1))
        (test(eq ?w 2))
        (test(eq ?s1 1))
                (test(eq ?s2 1))
        =>
        (assert (UI-control (id 1)(message "Patient Being Diagnosed with:TYPE-I-DIABETES")))
        ;(printout t "disease 1" crlf)
)
(defrule disease-2
        (preference (gender ?g)(age 2)(weight ?w)(risk n)(symptom1 ?s1)(symptom2 ?s2))
        (test(eq ?g 1))
        (test(eq ?w 2))
        (test(eq ?s1 3))
                (test(eq ?s2 2))
        =>
        (assert (UI-control (id 1)(message "Patient Being Diagnosed with:TYPE-I-DIABETES")))
        ;(printout t "disease 2" crlf)
)
(defrule disease-3
        (preference (gender ?g)(age 2)(weight ?w)(risk n)(symptom1 ?s1)(symptom2 ?s2))
        (test(eq ?g 2))
        (test(eq ?w 2))
        (test(eq ?s1 4))
                (test(eq ?s2 1))
        =>
        (assert (UI-control (id 1)(message "Patient Being Diagnosed with:TYPE-I-DIABETES")))
        ;(printout t "disease 3" crlf)
)
(defrule disease-4
        (preference (gender ?g)(age 2)(weight ?w)(risk n)(symptom1 ?s1)(symptom2 ?s2))
        (test(eq ?g 2))
        (test(eq ?w 2))
        (test(eq ?s1 3))
                (test(eq ?s2 4))
```

```
    =>
    (assert (UI-control (id 1)(message "Patient Being Diagnosed with:TYPE-I-DIABETES")))
    ;(printout t "disease 3" crlf)
)
```

Graphical User Interface to Diagnosis and Treatment Type I Diabetes

```
using System;
using System.Collections.Generic;
using System.ComponentModel;
using System.Data;
using System.Drawing;
using System.Linq;
using System.Text;
using System.Threading.Tasks;
using System.Windows.Forms;

namespace MDES
{
  public partial class Form5 : Form
  {
    #region Properties
    public string fileName = "Rule-1.clp";
    public string assertString { get; set; }
    public string Gender { get; set; }
    public string Age { get; set; }
    public string Weight { get; set; }
    public string Risk { get; set; }
    public string Symptom1 { get; set; }
    public string Symptom2 { get; set; }
    #endregion
     ClipsFramework cw;
        public Form5()
    {
      InitializeComponent();
      cw = new ClipsFramework();
      cw.Load(fileName);
      button2.Visible = false;
      treatment.Visible = false;
    }

    private void button1_Click(object sender, EventArgs e)
```

```
{

    lblResult.Text = "";
    GetFormValues();
    assertString = string.Format("(preference (gender {0})(age {1})(weight {2})(risk
{3})(symptom1 {4})(symptom2 {5}))", Gender, Age, Weight, Risk, Symptom1, Symptom2);
    cw.Reset();
    cw.CreateAssert(assertString);
    cw.Run();
    string response = cw.GetResponse();
    if (!string.IsNullOrEmpty(response))
        lblResult.Text = response;
    else
    {
        ApplyStrategy();
    }
    button2.Visible = true;
   treatment.Visible = true;
}

private void ApplyStrategy()
{
    cw.Reset();
    LoadAssertsrule();
    var results = LoadResultsByDimensions();
    var response = TranslateFinalResult(results);
    lblResult.Text = "Diagnosis Shows: "+ response;
}

private Dictionary<string, int> LoadResultsByDimensions()
{
    List<string> _results = new List<string>();
    Dictionary<string,int> _resultFinal = new Dictionary<string, int>();
    _results.Add(ByGender(Gender));
    _results.Add(ByWeight(Weight));
    _results.Add(BySymptom(Symptom1));
    _results.Add(BySymptom(Symptom2));
    _results.Add(ByAge(Age));
    _results.Add(ByFever(Risk));
```

```
    int cM1 = _results.Count(x => x == "1");
    int cM2 = _results.Count(x => x == "2");
    int cM3 = _results.Count(x => x == "3");
    int cM4 = _results.Count(x => x == "4");
    int cM5 = _results.Count(x => x == "5");
    int cM6 = _results.Count(x => x == "6");

    _resultFinal.Add("Unknown-Test-Detected-A", cM1);
    _resultFinal.Add("Unknown-Test-Detected-B", cM2);
    _resultFinal.Add("Unknown-Test-Detected-C", cM3);
    _resultFinal.Add("Unknown-Test-Detected-D", cM4);
    _resultFinal.Add("Unknown-Test-Detected-E", cM5);
    _resultFinal.Add("Unknown-Test-Detected-F", cM6);

    return _resultFinal;
}
private string TranslateFinalResult(Dictionary<string, int> results)
{
    return results.OrderByDescending(x => x.Value).FirstOrDefault().Key;
}


private void GetFormValues()
{

    Gender = "2";
    if (comboBoxGender.SelectedItem.ToString() == "Femalenonpregnant")
    {
        Gender = "1";
    }else if (comboBoxGender.SelectedItem.ToString() == "Male")
    {
        Gender = "2";
    }

    Age = "2";
    if (comboBoxAge.SelectedItem.ToString() == "Teenage(0-20)")
    {
        Age = "1";
    }
    else if (comboBoxAge.SelectedItem.ToString() == "Adult(20-40)")
```

```
{
   Age = "2";
}

Weight = "1";
if (comboBoxWeight.SelectedItem.ToString() == "Overweight")
{
   Weight = "1";
}
else if (comboBoxWeight.SelectedItem.ToString() == "Healthyweight")
{
   Weight = "2";
}

Risk = rbFever.Checked ? "y" : "n";

Symptom1 = "1";
switch (comboBoxSymptom1.SelectedItem.ToString())
{
   case "Symptoms-of-diabetes":
      Symptom1 = "1";
      break;
   case "Symptoms of Hyperglycemia crisis":
      Symptom1 = "2";
      break;
   case "(Rate of Triglycerides>=200)":
      Symptom1 = "3";
      break;
   case "Complication-symptoms":
      Symptom1 = "4";
      break;
   case "5":
      Symptom1 = "5";
      break;
   default:
      Symptom1 = "1";
      break;
}

Symptom2 = "1";
```

```
            switch (comboBoxSymptom2.SelectedItem.ToString())
            {
                case "(FBS>=126mg/dl)":
                    Symptom2 = "1";
                    break;
                case "(HbA1C>=6.5%)":
                    Symptom2 = "2";
                    break;
                case "Cuase-symptoms":
                    Symptom2 = "3";
                    break;
                case "common-symptoms":
                    Symptom2 = "4";
                    break;
                case "5":
                    Symptom2 = "5";
                    break;
                default:
                    Symptom2 = "1";
                    break;
            }
          }

        private void LoadAssertsrule()
        {
            string disease1 = "(diagnosis(disease 1)(gender 2)(age 1)(weight 1)(risk n)(symptom1
1)(symptom2 1))";
            string disease2 = "(diagnosis(disease 2)(gender 2)(age 1)(weight 1)(risk y)(symptom1
2)(symptom2 2))";
            string disease3 = "(diagnosis(disease 3)(gender 2)(age 2)(weight 2)(risk y)(symptom1
3)(symptom2 3))";
            string disease4 = "(diagnosis(disease 4)(gender 1)(age 1)(weight 2)(risk y)(symptom1
4)(symptom2 4))";
            string disease5 = "(diagnosis(disease 5)(gender 1)(age 2)(weight 1)(risk y)(symptom1
5)(symptom2 5))";
            string disease6 = "(diagnosis(disease 5)(gender 1)(age 2)(weight 1)(risk n)(symptom1
1)(symptom2 5))";
            cw.CreateAssert(disease1);
            cw.CreateAssert(disease2);
            cw.CreateAssert(disease3);
```

```
        cw.CreateAssert(disease4);
        cw.CreateAssert(disease5);
        cw.CreateAssert(disease6);
    }

    private string FindByOtherStrategy(string StrategyProperty, string valueToSearch, string
compOperator)
    {
        return cw.GetFact("diagnosis", StrategyProperty, valueToSearch,"disease",
"0",compOperator);

     }

    #region Little Strategy finding in facts
    private string ByGender(string value)
    {
        return FindByOtherStrategy("gender", value, "<=");
    }
    private string ByAge(string value)
    {
        return FindByOtherStrategy("age", value, "eq");
    }
    private string ByWeight(string value)
    {
        return FindByOtherStrategy("weight", value, ">=");
    }
    private string BySymptom(string value)
    {
        return FindByOtherStrategy("symptom", value, "<=");
    }
    private string ByFever(string value)
    {
        return FindByOtherStrategy("risk", value, "eq");
    }
    #endregion


    private void groupBox2_Enter(object sender, EventArgs e)
    {
```

```
        }

        private void button4_Click(object sender, EventArgs e)
        {
            MessageBox.Show(comboBoxSymptom1.SelectedItem + ":Correct:it is one of member used in diagnosis", "Attention!", MessageBoxButtons.OK, MessageBoxIcon.Information);
        }

        private void button5_Click(object sender, EventArgs e)
        {
            MessageBox.Show(comboBoxSymptom2.SelectedItem + ":Correct:it is one of member used in diagnosis", "Attention!", MessageBoxButtons.OK, MessageBoxIcon.Information);
        }

        private void button2_Click(object sender, EventArgs e)
        {
            MessageBox.Show(lblResult.Text + "Reason:System Suggested using inference method", "How?", MessageBoxButtons.OK, MessageBoxIcon.Question);

        }

        private void treatment_Click(object sender, EventArgs e)
        {
            MDES.Treatment.Type_I_Diabetes Type_I_Diabetes = new MDES.Treatment.Type_I_Diabetes();
            Type_I_Diabetes.ShowDialog();
        }

        private void button3_Click(object sender, EventArgs e)
        {
            this.Close();
        }
    }
}
```

Appendix III

Prototype Evaluation form for the Domain Expert

The importance of this evaluation form is to evaluate to what extent the prototype system is usable by the end-users in the domain area. I would like to thank you in advance for your willingness and valuable time.

Instruction: Please, tick on the appropriate value for the corresponding parameter of the knowledge based system for diagnosis and treatment of diabetes.

Parameter	Performance Value				
	1-Poor	2-Fair	3-Good	4-Very good	5-Excellent
Is the system is easy to understand and use?					
Is the user interface is attractive for use?					
Is it accurate in prioritizing the patients based on their symptoms?					
Does the system is important for the domain experts?					
Does the system have significance contribution for the domain area?					
Is the system effective in time?					
How do you get the system will be useful as a training tool?					