Adama Science and Technology University

School of Graduate Studies

Department of Computing

Rule Based Expert System for Classifying HIV Patients Taking ART

At Asella and Adama Hospital

Yordanos Gebeyehu

June 2015

Adama Science and Technology University

School of Graduate Studies

Department of Computing

A Thesis Submitted to the Department of Computing of Adama Science and Technology University in Partial Fulfillment of the Requirements for Degree of Master of Science in Software Engineering

By

Yordanos Gebeyehu

June 2015

Adama, Ethiopia

Adama Science and Technology University

School of Graduate Studies

Department of Computing

Rule Based Expert System for Classifying HIV Patients Taking ART

At Asella and Adama Hospital

By

Yordanos Gebeyehu

Names and Signature of Members of the Examining Board

______________________	______________________
Chair Person, Examining Board	Signature
Dr. Farid Ahmad	______________________
Advisor	Signature
______________________	______________________
Examiner	Signature

DECLARATION

This thesis is my original work and has not been submitted for a degree in any university.

YordanosGebeyehu

June, 2015

ACKNOWLEDGMENTS

On the top I would like to present my deep gratitude thanks to my God and St. Marry for everything which is passed in this research time and for reaching this day. I would like to express my heart full thank to Ato Tinsae Gizachew next to my advisor Dr. Farid Ahemd.

My appreciation and gratitude thank is also extended to all respondents who gave their valuable time, voluntary support and shared their personal experience to make the research possible. I thank the community of Adama and Assela hospital for all supports.

I would like to thank Ato Memekiya and Dr.Tilaye who are ART office workers for their ultimate support starting from giving the patients record to the last of my work.

And also I would like to present my heart-full thank to the Adama Sciences and Technology University; computing department staff members.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

LIST OF ACRONYMS AND ABBREVIATIONS

AIDS	Acquired Immunodeficiency Syndrome
AI	Artificial Intelligence
ARV	Antiretroviral Drug
HIV	Human Immune Deficiency Virus
ART	Antiretroviral Therapy or Treatment
HAART	Highly Active Antiretroviral Therapy
MoH	Ministry of Health
VCT	Voluntary Counseling and Testing
PMTCT	Prevention of Mother to Child Transmission
HAPCO	HIV/AIDS Prevention and Control Office
SQL	Structure Query Language
MVC	Model View Controller
KBS	Knowledge-Based System
JESS	Java Expert System Shell
DSS	Decision Support System
LHS	Left Hand Side
RHS	Right Hand Side
WEKA	Waikato Environment for Knowledge Analysis
MVC	Model View Controller
XML	Extensible Markup Language
UI	User Interface
RIPPER	Repeated Incremental Pruning to Produce Error Reduction
PART	Projective Adaptive Resonance Theory

GUI	Graphical User Interface
USA	United Nation of America
RBES	Rule Based Expert System
UML	Unified Modeling Language
RBES	Rule Based Expert System
ICT	Information Communication Technology
IDE	Integrated Development Environment

ABSTRACT

Knowledge based system is an interactive computer based system intended to help decision makers and assist experts in use of available technologies like ICT. Even though there are different sectors in which decision support system are used some of them are academic, commercial, industries, healthcare sectors, agricultural areas etc.; in most of them the way of making decision is not smart/properly designed because of lack of organizing huge amount of data in proper format for decision makers. A properly designed rule based expert system is intended to help decision makers to compile useful information from a combination of raw data, documents, personal knowledge, or business models to identify and solve problems and make decisions. Now a day's decision makers suffer from making decision from vast amount of data.

In this study rule based expert system for classifying HIV patients who are taking ART treatment is developed by adopting the concepts of data mining machine learning methods and rule based expert system. The thirteen years data of patients who are in ART treatment is analyzed to visualize and discover the knowledge. To acquire the rule, the PART algorithm is used while for implementation of this rule based system Jess language is used. The MVC design pattern is applied in developing this rule based expert system.

Keywords: knowledge based system, data mining, and software architectural pattern.

CHAPTER 1

INTRODUCTION

1.1. BACKGROUND

Acquired immune deficiency syndrome (AIDS) is a disease of the human immune system, caused by the human immune deficiency virus (HIV). HIV is a retrovirus that normally affects the immune system. AIDS is one of the most serious health crises and AIDS-related illnesses are among the leading causes of death worldwide [1]. The first cases of AIDS were reported more than 32 years ago in the June 5, 1981 in USA. Since then, an estimated 36 million persons worldwide have died from HIV/AIDS; an estimated 35.3 million persons continue to live with HIV infection [32]. Sub-Sahara Africa remains the region most heavily affected by HIV. 1.9 million People living in sub-Sahara Africa become newly infected with HIV, bringing the total number of people living with HIV to 22.4 million. Moreover an estimated 1.4 million AIDS related deaths occur in sub-Sahara Africa [7] [30].

Ethiopia is one of the countries in Sub-Saharan -African region in which, the HIV epidemic has remained a major public health problem, mainly affecting people of prime productive and reproductive age. In Ethiopia, the first HIV/AIDS case was reported in 1986. With an estimated adult prevalence of 1.5%, it has a large number of people living with HIV (approximately 800,000 from total population size of 87 millions); and about 1 million AIDS orphans [13].

Having this in mind; in 2001, the Strategic Framework for National Response to HIV/AIDS in Ethiopia was launched. Recognizing the seriousness of the epidemic and its multi-faceted impact, in 2002, the government declared AIDS as a national public health emergency. In 2002, the Federal HIV/AIDS Prevention and Control Office (HAPCO) with mandates to coordinate and lead implementation of the national HIV/AIDS policy, was established. The Federal HAPCO took a lead role in organizing the National HIV/ADS Council, National and Regional HIV/AIDS Secretariats and Advisory Boards, and the National Partnership and Donors' Forum against HIV/AIDS [13].

Furthermore, to facilitate implementation of the national policy, several strategies and guidelines were developed, implemented and further revised as necessary, including on voluntary counseling and testing (VCT), antiretroviral therapy (ART) and on prevention of Mother-to-Child Transmission (PMTCT) of HIV. Furthermore, a road map for accelerated access to HIV prevention, treatment and care in Ethiopia [16], and the Plan of action for universal access to HIV prevention, treatment, care and support were also developed

Even though the system of delivering this treatment is developed and run for more than 13 years and HER data is available, this data is not well exploited to discover new knowledge that could support decision makers. The way of collecting data and generating information is the crucial for decision makers [46]. Having this in mind, the interesting patterns and trends extracted from vast amount of organizational data sources through data mining can be meaningful to obtain valuable knowledge and this knowledge is imperative for decision making process in the form of recommendations reflecting domain expertise or classifications. But when we come to reality a large gap remains between the results a data mining system can provide and taking actions based on them [39] [15]. Besides that, in the field of software engineering, gaining qualified information is essential for the system to be developed.

Knowledge based system is a computer program that emulates the behavior of human expert to solve problems associated with a particular domain. Its' primary goal is to make expertise available to decision makers and technicians [3].

For the development of efficient knowledge based system, the knowledge base has to be complete, efficient and non-redundant. So, knowledge acquisition is one bottleneck for the development of knowledge based system when the knowledge engineer doesn't use different techniques of knowledge acquisition [40]. In alleviating this bottleneck different knowledge acquisition techniques could be applied. Among these techniques data mining is the one.

Integration of data mining with knowledge based system is essential to have efficient knowledge based system. In addition to this, with the increased demand of complex, scalable and maintainable applications; the MVC architectural pattern is popular. From the perspective of software engineering theories, software architectural pattern is MVC is a well-known architectural pattern to be used; that provides clear separation between "how data is modeled", "how the model is shown in the output", and "how input events affect the interstate of the model”. There for in deploying the data mining technique in to knowledge based system, the MVC architectural pattern is used.

The aim of this study is towards the applicability of MVC architectural pattern in designing rule based expert system for classifying HIV/AIDS patients who take ART treatment by their clinical and socio demographic categories.

1.2. ANTIRETROVAL THERAPHY (ART)

ART refers to the use of a combination of three or more ARV to achieve viral suppression. This generally refers to lifelong treatment. Synonyms are combination ART and highly active ART (HAART) [50].The primary goals for initiating ART are to:

- reduce HIV-associated morbidity and prolong the duration and quality of survival,
- restore and preserve immunologic function,
- Maximally and durably suppress plasma HIV viral load (see Plasma HIV RNA Testing), and
- Prevent HIV transmission.

ART has reduced HIV-related morbidity and mortality and behavior-associated transmission of HIV. HIV suppression with ART may also decrease inflammation and immune activation thought to contribute to higher rates of cardiovascular and other end-organ damage reported in HIV-infected cohorts [50]. So, having these benefits this treatment is given to all patients living with HIV AIDS and the 2013 Guidelines Development Group recommends that national HIV program provide ART to all people with a confirmed HIV diagnosis as follows:-

- CD4 count of 500 cells/mm3 or less, giving priority to initiating ART among those with severe/advanced HIV disease
- CD4 count of 350 cells /mm3 or less.
- It is also recommended to initiate ART in people with active TB disease and HBV co-infection with severe chronic liver disease
- all pregnant and breastfeeding women with HIV
- all children younger than five years living with HIV and
- all individuals with HIV in serodiscordant relationships, regardless of CD4 cell count

1.3. ART IN ETHIOPIA

ART is the one from several strategies developed and implemented by the federal HAPCO as mentioned in the last paragraph of background section. Recognizing the benefits, the Ethiopian government has made concerted and sustained effort over the last several years to introduce and scale-up counseling and testing services and use of antiretroviral drugs both for treatment and prophylaxis [16]. Be familiar with the devastating effect of HIV/AIDS on its population and the positive impact of ART, the Ethiopia government has responded to the epidemic as a national emergency and imperative to scale up the ART program since 2003, trained health providers in 37 hospitals throughout the country are providing fee based ART services to over 13,000 persons. The MoH has been working towards the provision of safe, effective, equitable and sustainable ART services to those infected by HIV. In this effort, A National Guideline on the use of ARV drugs was developed and the Antiretroviral Treatment (ART) program was launched in 2003. Subsequently, in 2004, a free ART program was initiated in three government hospitals in Addis Ababa. Since then the geographic distribution and number of centers providing ART services have increased. These efforts have led to marked increase in the number of health facilities and sites providing HIV treatment and care services. While there were 550 facilities providing ART in 2009/10, this reached 743 public and private health facilities in 2010/11 fiscal year.

Furthermore, mainstreaming of HIV/AIDS within many public sectors, non-government and private sector organizations have contributed to improve awareness, increase access and utilization of HCT and ART. As a result of this rapid expansion of service facilities and improved awareness, the number of patients on ART rose sharply over the years to reach 333,453 in 2010/11. However, in some regions 40% of patients who were enrolled to ART dropped out from treatment; a situation that requires serious attention [13].

1.4. MOTIVATION

The necessitated point to this research is that the number of mortality of HIV AIDS patient was declining from time to time. From the different strategies formulated to attain this result the one is implementation of ART treatment which is designed by WHO. Furthermore in this disease the number of 1 million AIDS orphans lift-down; this led to the transmission rate of HIV to be slowdown; as orphan peoples are the one factor for spreading this disease. HAART was the breakthrough in care and treatment of people living with HIV, leading to a reduction in mortality and an improvement in the quality of life. Antiretroviral drugs significantly lowered the rate of HIV transmission from mother to child, and antiretroviral therapy (ART) has become an integral part of the continuum of HIV care [37] [30]. So, if further research is conducted that can integrate ART service with Information Technology; which can visualize the hidden pattern of attributes in multidimensional way, it is possible to identify the specific category of the patient and this is relevant for decision makers to take action based on the identified categories.

1.5. STATEMENT OF PROBLEM

To increase the life time of HIV patients the service given to them by the Federal HAPCO like the ART and Monitoring system in addition to testing and counseling service and the way they are treating themselves are core points.

Patients who are registered in this treatment system takes all benefits given from it. The system uses database that is applicable for storage i.e. registering patient, report generation and searching mechanism. The data stored in this database is used for making decision to hospital administration by reviewing the report; following that line, they are incapable of making business decisions being provided with details. Currently the ART treatment office generate report for the ministry of health monthly that have the following contents: Number of persons on ART and already enrolled in program who transferred during the month, Number of persons who restarted ART during the month, after stopping art for at least 1 month, males greater than

14 years, pregnant and non-pregnant females greater than 14 years, children five up to fourteen year and 12 up to 59 months.

So, the report generated in this way is not sufficient to make full decision in forecasting ART treatment by classifying patients in to different classes and doesn't visualize the hidden attributes causing effectiveness of the treatment for reducing the stopping rate of patients. Hardly the report providing the ability to see large number of summarized, aggregated data, from different perspectives (in a multidimensional way). However the number of patients increased in registering and starting the treatment; the already registered once were leaving the treatment as the 2012 countries report stated like this (However, in some regions 40% of patients who were enrolled to ART dropped out from treatment; a situation that requires serious attention [14]). Data mining is a solution for discovering hidden but important patterns from large volume of data collected over time by viewing in multidimensional way and those are important in improving and forecasting the efficiency of the treatment from different perspectives.

According to [34], data collection for HIV/AIDS related intervention typically involves thousands of data items. From all this data, the health care provider must combine data elements for decision making related to HIV/AIDS and plan for intervention. It is important to assess the clinical and socio-demographic categories of patient's to plan for intervention. ART care adherence is expanding in every nation of the world of which Ethiopia is mentioned nowadays.

Different times at different regions, the factors that affect the adherence of ART treatment are well explored and mechanisms are adjusted in extending the health sectors that gives this service. But there is a gap in exploring their data's to for making decision.

Different researchers apply data mining to visualize these hidden patterns but not came up with a solution that illustrate the causes of a patient that stops or continues taking the treatment after they are registered in the service and don't give due support for them that needs aid for the one who stops taking the treatment because they are not investigating factors that affect the healthy deliver of the service. Therefore in addition to applying data mining techniques to find those factors using knowledge based system; rules and facts, can be used to easily know the cause and give sustainable support for the patients in staying with the treatment.

In the previous related work of [5], apply data mining which analyzes the ART data base for exploring attributes regarding on socio-demographic characteristics of patients but the result is not organized in easy to understand form of decision makers as the aim of the research was to see the applicability of data mining techniques but not how the result be presented to the user in suitable and easy to understand way. So, decision is not taken based on those characteristics. There for, in this research the socio-demographic characteristics of patients who are registering and taking services given from ART office will be identified using data mining machine learning method and the result are converted in to rule based expert system.

This rule based expert system is not mean to replace the human physician but using such system is useful for making effective decision at MOH level and it also assists counseling service givers by categorizing patients in to, clinical and socio-demographic characteristics.

This study explores and finds the following research questions:

1. What are the challenges of classifying patient's data in clinical and socio-demographic category?
2. Is it possible to use rules extracted from data mining to construct knowledge based system?
3. How it is possible to use MVC architectural pattern in knowledge based system?
4. Is it possible to update the knowledge based system?

1.6. OBJECTIVE OF THE STUDY

This section states about general and specific objectives of the study.

1.6.1. GENERAL OBJECTIVE

The main objective of this study is to investigate the applicability MVC design patterns and data mining techniques in the development of rule based expert system for classifying HIV/AIDS patients in ART. The result of this objective is: first the system will be easily maintainable, scalable and interactive; second the current consultation service given to the patients is to be modified means based on the clinical and socio demographic classification result, the consultation is to be given.

1.6.2. SPECIFIC OBJECTIVES

In this research paper the following specific objectives are included having the above general objective

- To review different literatures which have similarity in this study
- To collect and analyze working data
- To analyze and identify appropriate data mining rule based classification algorithms.
- To acquire knowledge from classification algorithms for constructing knowledge base
- To analyze and design the prototype system
- To maintain the knowledge base
- To evaluate the performance of the prototype system
- Concluding and recommending from the result

1.7. SCOPE AND LIMITATIONS OF THE STUDY

Even though there are different design patterns for developing rule based expert system, this study focused only on MVC design pattern. In addition to this, it is limited to design rule based classification system for decision makers at ART voluntary counseling service givers and MoH level in Ethiopia.

The knowledge is acquired automatically from the hidden knowledge of ART database. So the knowledge acquisition method of this study is not relying on only the traditional knowledge acquisition methods like interview; but a collection of both traditional and modern acquisition methods.

So, the research is also limited in analyzing the current ART database and taking this analysis for designing rule based expert system.

1.8. METHODOLOGY

The methodologies that can be used in this research are data sources, data preparation and preprocessing, knowledge acquisition, knowledge representation, implementation tools, testing and evaluation. Even if in the late chapter the proponent deal about them, the highlight of their concept is described as follows.

1.8.1. DATA SOURCES

For this research the proponent uses two data sources. The first one is documentary data source; in which the electronic health records of patients are documented. The thirteen year data is collected from Adama and Assela hospital from 1993 E.C-2006 E.C. Total of 29,439 data is taken from those two hospitals 20000 and 9439 respectively. In addition to this the researcher is used another method which is interview; with the ART monitoring system data clerks and physicians in order to understand the data.

1.8.2. DATA PREPARATION AND PRE-PROCESSING

At this step the following tasks are being included: description of data sources, finding out distinct values, missing values, outliers, noisy data values. Making decisions on how to handle missing values, outliers and noisy data and data transformation/ reduction activities are also undertaken.

1.8.3. KNOWLEDGE ACQUISITION

It deals about the mechanisms of acquiring knowledge from different sources like interview, questioners, documentary sources, and so on. In the development of most knowledge based system knowledge acquisition process is one bottleneck. So, the knowledge acquisition method for this study is automatic i.e. rules extracted from machine language. This method is also reducing the knowledge acquisition time quite drastically.

1.8.4. KNOWLEDGE REPRESENTATION

To represent knowledge in knowledge based system different reasoning mechanisms are there: case based reasoning, rule based reasoning or a combination of two; is most commonly used. The purpose of these reasoning mechanisms is to represent the acquired knowledge in proper format for the selected algorithm at the implementation part. So, for the proposed rule based expert system, the knowledge is represented in the form of if then rules. Then reason behind this selection is that the knowledge which is acquired from the data mining classification algorithm has the form of if…then format. Plus to that the nature of the problem i.e. classifying patients in to different clinical and socio-demographic categories is also best to be represented by rule based reasoning rather than cased based or other techniques of knowledge representation.

1.8.5. SOFTWARE ENGINEERING APPROACHE TOWARDS KNOWLEDGE BASED SYSTEM

1. SOFTWARE ENGINEERING

It is the discipline of software development [44]. It is an engineering branch associated with development of software product using well-defined scientific principles, methods and procedures. The outcome of software engineering is an efficient and reliable software product.

2. SOFTWARE ARCHITECTURE PATTERN

An architectural pattern describes a structure for an entire software system. It defines a set of subsystems, their responsibilities, and the interrelationships between them. There are different types of architectural patterns like Model-View-Controller, layered architecture of enterprise applications, pipe and filter for information processing systems like compiler etc. but for this research the focus is on MVC.

I. MODEL VIEW CONTROLLER MVC) PATTERN

The basic parts of any application are data being manipulated and a user interface through which this manipulation occurs. In MVC architectural pattern, it is common to think of an application as having three main layers: presentation (UI), application logic, and resource management. In MVC, the presentation layer is split into controller and view [29]. As described by [52], pattern enables us to break down the system into three parts, model, view and controller; model handles the application logic layer while view and control is for presentation and resource management layer.

Model: -The model is another name for the application logic layer (sometimes also called the domain layer). It incorporates methods for processing and modifying state.

View: - Renders the model into a form suitable for interaction, typically a user interface element; when model changes, view must be updated.

Controller: - Translates user actions (i.e. interactions with view) into operations on the model. Example user actions: button clicks, menu selections

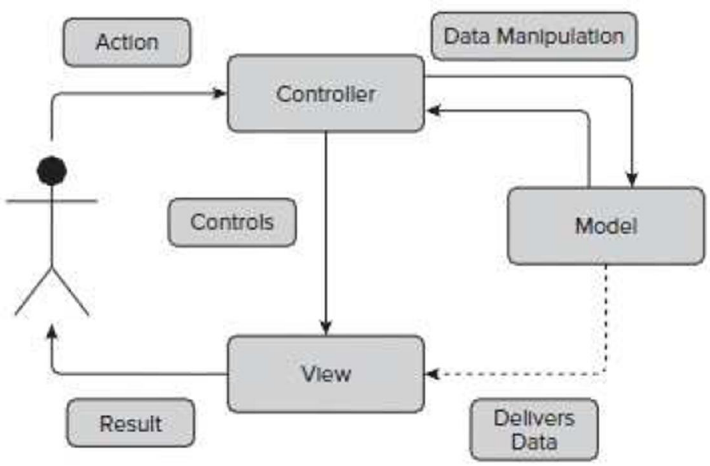


Figure 1-1: Model View Controller Architectural pattern, **[25]**

The user makes an action, which is managed by the controller. Controller handles the required data manipulation with the model and passes it on to the view (resulting in indirect relationship between Model and View), which builds the markup based on the delivered data and shows it to the user.

II. ADVANTAGES OF MVC PATTERN

According to [41], MVC design pattern have the following advantages and disadvantages:-

- The application of MVC patterns allows multiple views over a single model. MVC strictly separates the model from the user interface components and therefore multiple views can be implemented and used with a single model.
- MVC allows the view synchronization. Model change propagation mechanism provides that all views registered with the object of "observer" are notified about changes of application data exactly on time. This allows synchronization of all dependent views and controllers.
- MVC also allows creation of "Pluggable" views and controllers. Conceptual separation of MVC allows to finds a replacement for view and controller object. User interface objects can even be replaced during the time of execution.
- The possibility of "look and feel" change of the user interface characteristics. Since the model is independent of the user interface code, transferring of the MVC application to a new platform does not affect the functional application core.

III. DISADVANTAGES OF MVC PATTERN

- The use of this pattern also creates the potential for an excessive number of updates. If actions of one user result in many updates, the model needs to skip the unnecessary notification of changes.
- The lack of closeness between views and controllers is also a disadvantage because it prevents their specific use.

3. INTEGRATING SOFTWARE ENGINEERING APPROACH WITH KNOWLEDGE ENGINEERING

Software engineering capture some concepts from knowledge engineering and the vice versa it true. Software Engineering can improve the knowledge gathering process and the documentation of the different kinds of knowledge. So software engineering can have a hand in validation and maintenance of large knowledge bases.

A. KNOWLEDGEENGINEERING APPROACH

In knowledge based systems the knowledge base and knowledge processing facilities are basic issues. It uses various knowledge representation methods, which are declarative in nature. In case of RBS these are production rules. The knowledge engineering process in case of RBES involves two main tasks: knowledge base design and inference engine implementation.

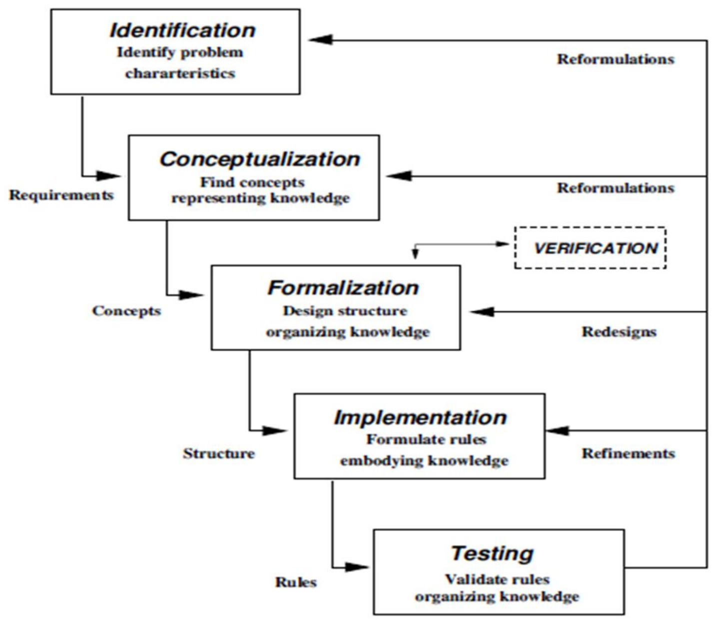


Figure 1-2: Knowledge Engineering Process [34]

As it is illustrated in the above figure five phases are there, the knowledge based system should capture and contain the knowledge about the real world system and the task of the programmer is to develop processing facilities for the knowledge representation. In the KBS there is no single universal modeling method like what is in software engineering-UML.

B. SOFTWARE ENGINEERING APPROACH

In software engineering there are different software development process models. From them one of the most common models is called waterfall model

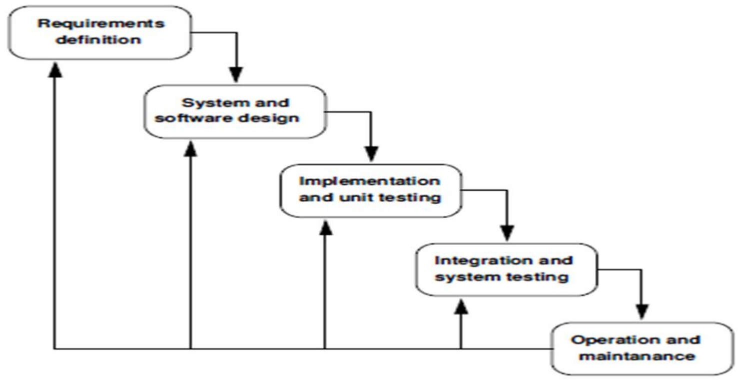


Figure 1-3: Waterfall Model [34]

So by reviewing the above two diagrams, the difference between the knowledge engineering and software engineering is that; knowledge engineering captures and represent what is known about the system while software engineering tries to model how the system works.

Now a day's software engineering becomes knowledge based while knowledge engineering is more about software engineering. Because the actual implementation of KBS has been some of high level programming languages like CLIPS, PROLOG and the modern knowledge engineering tools heavily depend on some common development tools and programming languages, especially when we come to user interfaces.

These leads to in conclusion of software engineering could adopt something from knowledge engineering: advanced conceptual tools like declarative knowledge representation methods, knowledge transformation techniques based on existing inference strategies as well as verification, validation and refinement methods [15].

While knowledge engineering could adopt: programming interfaces to the existing software systems and tools, interfaces in to advanced storage facilities such as databases and data warehouses, modern user interfaces including graphical one.

So, this paper is written from software engineering point of view and this is the reason to compare the knowledge engineering with software engineering.

C. RULE BASED SYSTEM DESIGN ISSUES

Some of the design issues in knowledge representation are expressive adequacy, completeness and flexibility. Different representations address this issue in different ways []. While there are knowledge representation methods, the one is machine learning methods like induction rule.

4. SOFTWARE TOOLS FOR IMPLEMENTATION OF RULE BASED SYSTEM

There are different high level programming tools or software's to implement rule based systems like LISP, CLIPS, PROLOG, JBOSS, JESS and etc.

But in this study jess will be used. It is a software tool to implement rule based systems. Jess is a rule-based system shell entirely written in Java [18]. It is developed at the Sandia National Laboratories in Livermore, California for developing rule based expert systems. Jess was inspired by the open-source expert system shell C Language Integrated Production System (CLIPS). Jess itself is not open-source, though the source code can be acquired by obtaining a license. Academic institutions can apply for a license for free and the binaries are available for download at no cost for 30 days usage [18]. Jess programs are written in the Jess rule language (a List Processing (LISP) like syntax) or in XML and are interpreted by Jess rule-based system [17].

Jess supports both forward chaining and backward chaining rules and it uses special algorithm called Rete to much the rules to the facts. The researcher chooses this tool because Jess is popular language/shell, active user community, implemented in java and tight integration means that it supports changes to representation models when needed plus to the unified user interface

1.8.6. IMPLEMENTATION TOOLS

In order to discover the hidden knowledge from the database the WEKA 6.8.1 is used. It is desire since using that software is powerful of data mining as stated by many. After the data preprocessing and preparation step is completed, using rule induction classification algorithms i.e. PRISM, PART, RIPPER and OneR the researcher tries to acquire knowledge. Total of two experimental settings are conducted and the detail will be left on the third chapter. PRISM and OneR induction algorithms are discarded as they don't support instances which have missing value. Generally the WEKA is used by the researcher to get input result which enables to develop the acquired rule based expert system.

To represent rules which are constructed from the data mining rule based classifier algorithms into rule based expert system Jess is used. For designing the user interface of the proposed system java programming language is used by.

1.8.7. TESTING AND EVALUATION

After the prototype of the system has been accomplished, as the discipline of software engineering [28], system testing and user acceptance testing should be considered. The major aim of system testing is to assess how many of the rules are efficient in classifying patients in to their really category while user acceptance testing is to assess the system readiness for confirming that the system work as expected.

So, for user acceptance testing questioner be prepared and delivered to the system users i.e. counseling service givers and office workers at MOH level in HIV/AIDS case. Then the result be analyzed and discussed by comparing with what was done by other related works of this research.

To apply system testing, different scenarios will be selected to test the system. For example, in the developing rule based expert system rules are categorized in to different modules. So, here the consistence of rules in each module be tested and also the same is true which might occur as syntax errors within the module. Totally, for system testing four scenarios is taken by the researcher and the detailed discussion is illustrated at testing, analysis and interpretation chapter.

1.9. CONTRIBUTION OF THE WORK

In this research a better way of eliciting patients in their proper category is developed by integration of data mining and knowledge based system. Decision makers like MOH and voluntary counseling service givers can use this classification for giving better service to HIV patients taking the ART treatment.

The researcher analyzes the different methods and techniques which are used by the previous researchers for developing knowledge based systems. So, by extending their future work, rule based expert system for classifying HIV patients be developed by deploying basic components of the system in to MVC architectural pattern.

Most of the time knowledge based systems are recommendation based. So the other contribution of this research is that development of easily interactive, scalable and updatable classification based expert system.

1.10. THESIS ORGANIZATION

This section highlights the different chapters restricted for the building of this thesis. Chapter 2 deals about literature review and related work. In the literature review section important terminologies that have relation in this thesis work will be reviewed from different published literatures like article, report, conference proceeding, journal, books, previous master's thesis, etc. While chapter 3 discusses data preparation and knowledge acquisition methods, the fourth chapter illustrates the design of the proposed system by identifying and defining the key components of the system with their relationship. Chapter five discusses the implementation of the proposed system using software development tools while chapter six states about the testing part by making clear criteria of testing and evaluation and the analysis and interpretation of what is tested and evaluated. At the last chapter what is done in this research will be summarized as conclusion and specifically state about the recommendation of the research.

CHAPTER 2

LITERATURE REVIEW

Rule based system is one of the knowledge based system and its notion is derived from the field of AI. In developing such rule based expert system for classification, software design patterns are also important aspects from that of software engineering discipline perspective.

2.1. ARTIFICIAL INTELLIGENCE

Artificial intelligence (AI) is a broad interdisciplinary field which encompasses working in robotics, computer vision and natural language processing, as well as expert systems. AI got its name starting from year 1956 at Dartmouth conference. Definitions of AI vary from people to people that means there is no universally accepted definition. This means that different people working in this area for many years have proposed different definitions; the reason behind is Artificial Intelligence (AI) technology provides techniques for developing computer programs for carrying out a variety of tasks, simulating the intelligent way of problem solving by humans. The problems that humans solve in their day-to-day life are of a wide variety in different domains. Though the domains are different and also the methods, AI technology provide a set of formalisms to represent the problems and also the techniques for solving them. Therefore, definitions of artificial intelligence vary from the 'study of mental facilities through the use of computational models' to the 'science of making machines do things that would require intelligence if done by men', [26]. According to Rich, AI is the study of how to make computers do things at which, at the moment, people are better, [43]. It is the automation of activities we associate with human thinking such as perception, reasoning, decision making, problem solving and learning." [9].

Intelligent program (AI) has interdisciplinary approach of various disciplines like computer science, cognitive science, hardware field, etc. In the medical applications areas such as diagnostic techniques in ultrasonography, x-ray, computerized tomography scans, nuclear magnetic resonance imaging etc. The other areas can be clinical laboratories, pathological investigations and computer assisted decision-making. Therefore, the field is potentially relevant to any sphere of human intellectual activity. AI started with a goal to replicate human level of intelligence in machine learning.

So, when we come to the services given by medical sectors like ART, artificial intelligence enables concerned parties at MOH level and counseling service givers to take computer assisted decision making by capturing knowledge using machine learning tools from the diagnostic record of patents who are on ART.

2.2. KNOWLEDGE BASED SYSTEM

Expert system is one of the areas of artificial intelligence. An expert system also known as knowledge based system is a computer program that contains the knowledge and analytical skills of one or more human experts in a specific problem domain [48] [42]. So, a knowledge-based system (KBS) is a software system that contains a significant amount of knowledge in an explicit, declarative form. As [48] stated, the goal 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. The Knowledge-based systems can be used for interpretation, prediction, diagnosis, design, planning, monitoring, debugging, repair, instruction, and control [45], in improving the service delivery of different fields etc. from those one is health improvement. According to [45], it gives the following improvements in health sectors; Government schemes information system, Diet planning system, Disease diagnostic system, Disease diagnostic system for cattle and live-stock, Patient monitoring system, Medical insurance, KBS monitoring in surgical process, KBS for guided neuro-surgery, etc.

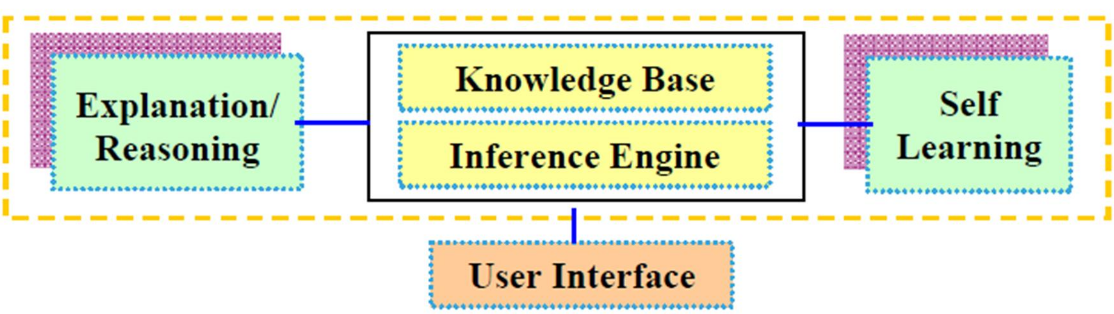


Figure 2-1: Architecture of knowledge based system [45]

2.2.1. ARCHITECTURE OF EXPERT SYSTEM

An expert system tool, or shell, is a software development environment containing the basic components of expert systems. It can be developed for any problem that involves a selection from among a definable group of choices where the decision is based on logical steps. Every knowledge base systems will have basic components of knowledge base, inference engine and user interface [20]. So the following diagram illustrates the core components of expert system.

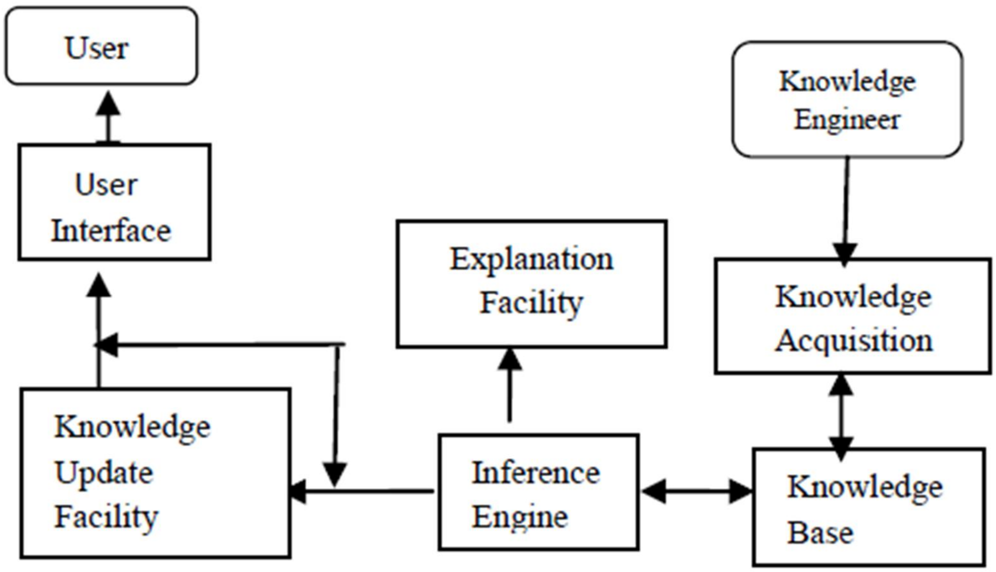


Figure 2-2: Architecture of Expert System [48]

Knowledge based: The heart of the expert system is the knowledge base, which contains the problem-solving knowledge of the particular application. It is constructed from the domain knowledge that is normally provided by the human experts for understanding, formulating and solving particular problem domain [26] [11].

Inference Engine: - is a brain of expert system. It is responsible for controlling the way knowledge in the knowledge base accessed and communicates the result via user interface to the user. It is the knowledge processor, modeled after the expert's reasoning that works with available information on a given problem, coupled with the knowledge stored in the knowledge base, to draw conclusions or recommendations [28]. It is used to perform the task of matching antecedents from the responses given by the users and firing rules. The major task of inference engine is to trace its way through a forest of rules to arrive at a conclusion. Here two approaches are used i.e. forward chaining and backward chaining [48].

Knowledge Acquisition: - Knowledge acquisition is the process of extracting, structuring, and organizing knowledge from several knowledge sources, usually human experts so that the problem-solving expertise can be captured and transformed into a computer-readable form. Knowledge is the most important component of expert systems [28]. For knowledge acquisition, techniques used are protocol analysis, structured or focused interview, introspection, protocol analysis, dialogs, review, rating grids, hierarchical clustering, and machine learning methods [48] [24].

Explanation Facility: - enable the user to ask the expert system how a particular conclusion is reached and why a specific fact is needed. An expert system must be able to explain its reasoning and justify its advice, analysis or conclusion [36].

User Interface: - The user interface allows the user to communicate with the expert system and also provides the user with an insight into the problem-solving process carried out by the inference engine. There are several ways to communicate with the expert system such as questions and answers, menu choices, statements, and others. Responsibility of user interface is to convert the rules from its internal representation (which user may not understand) to the user understandable form [48].

Knowledge Engineering: - it means building an expert system. Knowledge engineer is involved in the development of the inference engine, structure of the knowledge base and user interface. The expert and knowledge engineer should anticipate user's need while designing an expert system [48]. The major task of knowledge engineer is to represent the elicited knowledge. According to [8], Knowledge can be represented in many ways, but the most common ways are Rule-based system, Case-based system, Logic-based system, Frame-based system, Object-based system.

2.2.2. RULE BASED SYSTEMS

It is one of the most common forms of knowledge representation used in expert systems to represent knowledge. Knowledge is represented in term of rules. If some condition, then some action this means rule based system identifies the pattern and draws conclusions, advices what should be done about it or takes appropriate action. The term rule in AI, which is the most commonly used type of knowledge representation, consists of two parts: the IF part, called antecedent (premise or condition) and the THEN part called consequent (conclusion or action) [21].

IF <antecedent>

THEN <consequent>

A rule can have multiple antecedents joined by the keyword AND (conjunction), OR (disjunction) or a combination of both.

IF <antecedent 1>

AND <antecedent 2>

....

THEN <consequent>

The antecedent or if part of the rule is often referred to as LHS (left hand side) while the consequent or then part is RHS (right hand side).

When rules are examined by the inference engine, actions are executed if the information supplied by the user satisfies the conditions in the rules [22]. If the condition part of the rule is satisfied, then the rule is said to fire and the action part is executed.

Rule based expert systems have a significant role in many different domain areas such as medical diagnosis, electronic troubleshooting and data interpretations.

2.2.3. RULE BASED REASONING TECHNIQUE

The two common methods of reasoning, when using inference rules, are forward and backward chaining.

FORWARD CHAINING

Rules can be forward-chaining, also known as data-driven reasoning, because they start with data or facts and look for rules which apply to the facts until a goal is reached. It is a top-down method which takes facts as they become available and attempts to draw conclusions (from satisfied conditions in rules) which lead to actions being executed [22].It used a Rete based algorithm called ReteOO.

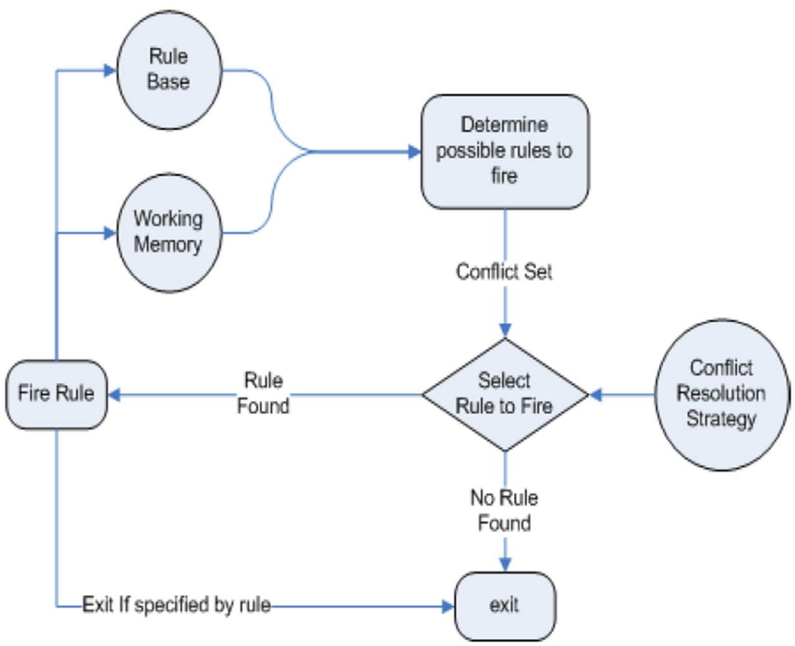


Figure 2-3: forward chaining system [4]

As illustrated in Fig-5, when a new fact is inserted into the working memory, the inference engine matches the fact against rules in the rule base and through the conflict resolution strategy a rule or multiple rules will be fired.

Rules can also be backward-chaining, also known as goal-driven reasoning, because they start with a goal and look for rules which apply to that goal until a conclusion is reached. It is a bottom-up procedure which starts with goals (or actions) and queries the user about information which may satisfy the conditions contained in the rules [22].

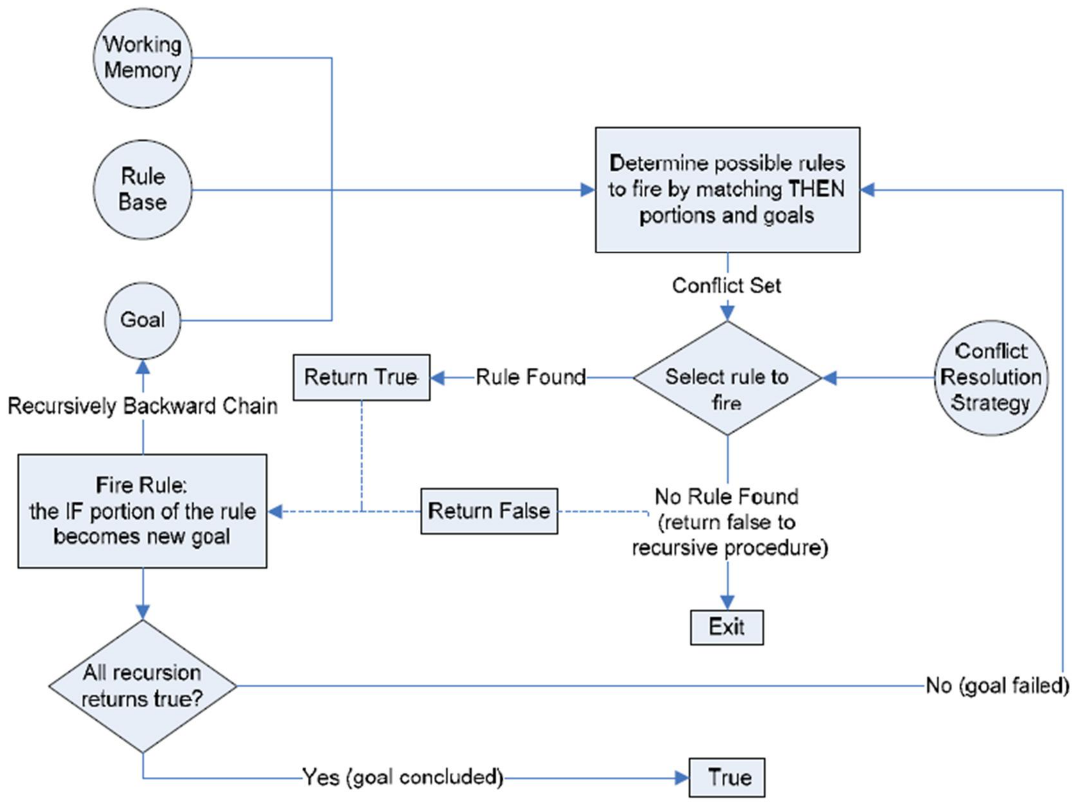


Figure 2-4: backward chaining system [4]

As illustrated in Fig-6, the backward chaining system uses depth-first algorithm that searches a data in tree as deep as possible before backtracking. When a new fact is inserted in to the working memory, the inference process begins from that fact and it is called as a goal. The inference engine matches the fact against the rule in the rule base and the conflict resolution selects the rule to be fired and if the IF portion of the rule becomes new goal then the inference process moves backward to other facts behind the goal.

2.2.4. MODEL OF RULE BASED LANGUAGES

There are different types of modeling languages that are used to depict the system. According to [51], the UML representation of rule based languages is illustrated as follows:-

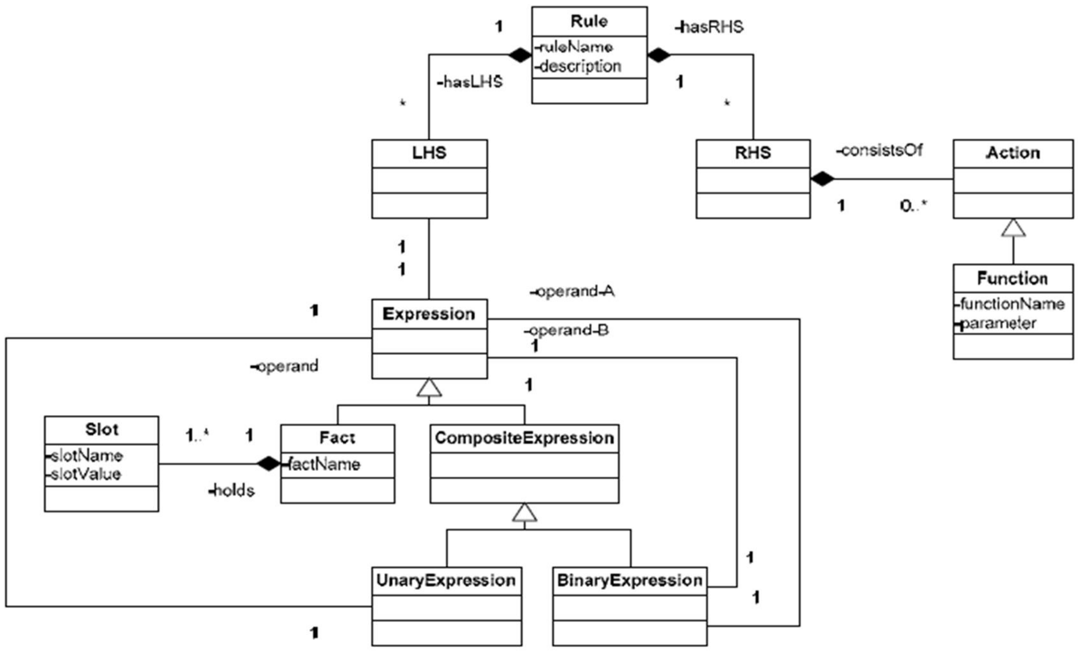


Figure 2-5: UML diagram for rule based languages

The above diagram is composed of eleven classes and the description of each of them is as follows:

Class rule has attribute ruleName and description. While the ruleName tells about the name of the rule the description is also handles the description or function of that rule. Rule has two parts [21]; the LHS and RHS. If the LHS rules are satisfied then the action on the RHS is be performed this means: the LHS rules are expressions but the RHS are actions. An action is a function characterized by function Name and parameter. But when we come to expressions, it consists of several expressions that are connected by conditional operators like AND, OR, NOT respectively for binary and unary expressions. Class fact is also defined with one attribute which is factName. And this class has connection with class slot. The slot class is composed of slotName and slotValue for representing the name and its value of the slot.

2.3. DATA MINING

Data mining as an interdisciplinary field which gets together the techniques from machine learning, pattern recognition, statistics, databases, and visualization [38]. It is used to analyze large observational data sets, find unsuspected relationships [14], and discover patterns and trends [27]. Data mining is a multidisciplinary field consisting of many contributing scientific domains related to computing mainly Artificial Intelligence (AI), databases, and mathematics (statistics and probability) (Witten and Frank, 2002). There are many definitions for data mining, for example [47], defined data mining as the process of producing new patterns from large data sets utilizing intelligent methods. We have defined data mining as a science that is concerned about revealing unseen information in a user preferred format from data for specific use.

Due to the wide availability of huge amounts of data and the imminent need for turning such data into useful information and knowledge [23], data mining has attracted a great deal of attention in various domains. In spite of various utilizations of data mining techniques in previous investigations, there are many domains which still have not used data mining methods. The most common tasks, which data mining can accomplish, are as follows: description, estimation, prediction, classification, clustering, and association [38]. Depending on the target of investigation which contains one or a compound of the named tasks, data mining has several defined algorithms and methods.

2.3.1. CLASSIFICATION METHODS

Classification is one of the most frequently studied problems by DM and machine learning (ML) researchers'. It is a data mining technique used to forecast group relationship for data instances. The hidden knowledge of vast amount of data is enhanced by classifying the organizational data in given class. There are two types of classification methods in data mining. These are:

1. Direct method: - in this method rules are extracted directly from data. Weka supports algorithms that work in this method. Example RIPPER and OneR classification algorithms. This method works as follow:

Let E be the training set, for each class C

- Initialize set S with E
- While S contains instances in class C
  - o Learn one rule R for class C
  - o Remove training records covered by the rule R

This means that to create rules that cover many examples of a class C and none (or very few) of other classes.

Indirect method: - here rules are extracted from other classification models like decision trees, PART, C4.5 algorithms

Even though many classification models are used to classify new objects as [29], Decision Tree, Naïve Bayes, Instance Based Learning, Rule-based Methods, Neural Networks, Bayesian Belief Networks and Support Vector Machines; we discussed the rule based classification method that uses the above two methods of rule based classification. These are: coverage classification, rule induction, PART algorithm and OneR.

COVERING CLASSIFICATION

In covering classification like PRISM algorithm, rules are derived in greedy way in which PRISM splits the training data set into subsets with respects to class values [10]. Then for each subset, the algorithm forms an empty rule and searches for the attribute value that has the highest expected accuracy and appends it into the rule body, and continues finding attribute values until the current candidate rule becomes with maximum expected accuracy (often 100%). Once this happens, the algorithm generates the rule and removes all of its positive instances (data in the subset that belong to the rule). The same process is repeated to produce the rest of the rules from the remaining uncovered data in the subset until the subset becomes empty or no rule with acceptable expected accuracy can be derived. At that point the algorithm moves on to the next class subset and repeats the same process until all rules in all class data subsets are generated and merged to form the classifier. One notable problem about this classification approach is that rules are found from subsets of the training data and not from the whole data set. This makes local classifiers rather globally ones. Further, the effort required to find the best attribute value to append into a rule at any stage of the learning phase is exhaustive when we have high dimensional training data sets.

RULE INDUCTION

Rule induction approach in classification works similar to covering approach but with more intensive pruning and optimization of the rules set. Often, a rule induction algorithm like RIPPER divides the training data set in to growing and pruning with respect to class labels [12]. Then starting with the least frequent class set, it builds a rule by adding items (attribute values) to its body until the rule is perfect (the number of negative examples covered by the rule is zero). For each candidate empty rule, the algorithm looks for the best attribute value in the data set using IG and appends it to the rule's body. It keeps adding attribute values until the rule becomes perfect at that point the rule gets generated. This phase is called rule growing. At the same time while rules are built, RIPPER uses extensive pruning using both the positive and negative examples associated with the candidate rules to reduce rules redundancy and eliminates unnecessary attribute values. The algorithm stops building the rules when any rule found has 50% error or in a new implementation of RIPPER when the minimum description length (MDL) of the rules set after adding a candidate rule is larger than the one obtained before adding the candidate rule.

Another pruning occurs on the candidate rules set to devise the final classifier. So for each candidate rule generated, two substitute rules are made: its replacement and its revision. The first one is made by growing an empty rule and filtering it to minimize the error on of the overall rules set. The revision the rule is built in similar fashion except the algorithm just inserts an additional item to the rule's body, and examines the original and the revised rule against the data to choose the rule with the least error rate. These extensive pruning in RIPPER explains the small size classifiers generated by this type of algorithms.

PART APPROACH

A hybrid classification algorithm that uses decision trees and rule induction approaches together to produce classifiers in one phase rather than two phases called PART was proposed According to [49], PART employs rule induction to generate the candidate rules set and then filters this set out using pruning methods adopted from decision trees. PART builds a rule as rule induction algorithms but rather constructing the rule directly from the data, it derives a sub-tree (partial decision tree) from the data and then PART converts the path leading to the leaf with the largest coverage into a rule and the sub-tree gets discarded along with its positive instances from the data set. The same process is repeated until all instances in the data set are removed.

ONERULE

One Rule is a simple rule based algorithm that was proposed by [49]. Only one attribute (that is why it is called 1R) is assumed sufficient to classify the new incoming records. The basic idea in this approach is to find the best attribute to perform the classification with the lowest errors as based on the training data. The `best' is defined by counting the number of errors for each attribute. In other words, the one with the least total error of all of the values will be the `best' attribute for the OneR algorithm. Therefore, OneR selects the rule with the lowest error rate.

This algorithm makes a one-level tree and produces rules that are connected with the most frequent class in the training data set (having the largest data coverage). For all attribute values in the training data set, OneRule iterates over the training data examples and computes the frequency of each attribute value with respect to available class labels. The algorithm selects the most frequent attribute and class and generates them as rule if they pass an error rate check. Finally, the algorithm repeats the same step to generate the subsequent rules until it finds a rule with unacceptable error at that stage the rule discovery process terminates.

2.4. RELATED WORKS

In this part the related works done on knowledge based expert system in different sectors and in health especially are reviewed and discussed as follows.

The first one is [6], KNOWLEDGE BASED SYSTEM FOR PRE-MEDICALTRIAGE TREATMENT AT ADAMA UNIVERSITY ASELLAHOSPITAL. The motivation of this study was in the triage treatment, there are dependent systems like effective patient care depends on the interaction of emergency physicians, triage nurse emergency nurses, laboratory and diagnostic imaging services, and inpatient units. As healthcare services are highly complex and interdependent, if any one of these interdependent components was performed poorly or overwhelmed, delivery of care in the Emergency Department will suffer. As he discussed using knowledge based system improves quality of decision made by general practitioners, to provide effective and efficient services to the patients, and to improve shortage of human expert in triage service.

To acquire knowledge of the expert conventional knowledge acquisition methods like questioner, interview and hard copy document observation and analysis was used. The rule based and case based types of knowledge based types are also exercised. The researcher uses decision tree to illustrate rules diagrammatically and prolog language to implement the prototype of the proposed system.

[31], proposes the integration of data mining with knowledge based system in case of network intrusion detection. He applies data mining methods for acquiring knowledge to build the knowledge base. The JRIP algorithm was used to generate rules automatically from the pre-processed data. His expert system is recommendation based expert system.

The implementation is conducted in prolog language but the system lacks the user interface. He recommends that first; integration of data mining with knowledge based system in other domain areas than intrusion detection is essentially especially where there is shortage of domain expert. Secondly, building KBS with graphical user interface which is simple to use and attractive to users.

TEKLU URGESSA [5], has conducted data mining research in health care with the above title in case of Adama and Asella hospital. The aim of his research was to see the applicability of data mining techniques on ART data by taking Adama and Asella Hospitals as cases to come up patterns that affect continuity/termination behavior of patients once registered for the service. The researcher used total of 18,740 records of the ART databases of both hospitals. The dataset consists of a total of 17 attributes. These attributes are: 'Item Number', 'Card Number', 'Registration date', Eligible date', Eligible and Ready date', 'On ART Date', 'Sex', 'Age', Functional Status ,'Termination', 'Religion', 'Educational Level', 'ART Stage', 'Occupation', 'Marital Status', 'OAWHO' (WHO Stage) and 'OACD4' (CD4 count).

The researcher was tried to find the relationship among the various attributes and their values to determine the termination/continuity status of a patient. He applied binary decision tree and association rule mining of data mining techniques in his research. The researcher tries to mine important (interesting) rules at different minimum support and confidence thresholds using weka as a tool. One of the most commonly used association rule modeling technique is apriori algorithms.

The first paper address how the knowledge based system is applied for hospital triage service but our research proposes firstly how the software engineering architectural pattern which is MVC will be used in developing rule based expert system because it has mainly three layers. This three-tier architecture allows programmer to add further modules in the software without disturbing previous one; and secondly the way of knowledge organization is based on the concept of data mining knowledge discovery process.

The second paper focuses on integration of data mining with knowledge based system. The main target of this paper was how to gather historical data and information on the disease since those are distributed among different actors in the agriculture production system. When we compare this paper from the first one it is better because it uses Weka software library which provides Machine Learning and Data Mining Algorithm, applied to the analysis of the data collected byKDDcup'99. As it is in the recommendation part of this thesis, it lucks well designed user interface.

The last one is the paper done by Teklu Urgesa. The aim of his research was to see the applicability of data mining techniques on ART data by taking Adama and Asella Hospitals as cases to come up patterns that affect continuity/termination behavior of patients once registered for the service. He applies association rule mining by using apriori algorithm. The limitation of this research is that these patterns are not organized in suitable format for making decision for the concerned parties or stake holders.

So in our paper we propose “rule based expert system for classifying HIV patients taking ART treatment” which will focus on MVC architectural pattern, PART algorithm of data mining and rule based expert system.

Author	Purpose	Method and Techniques	Key Findings
Teklu Urgesa (2010)	**APPLICATION OF DATA MINING TECHNIQUES ON ANTIRETROVIRALTHERAPY (ART) DATA: THE CASE OF ADAMA AND ASELLA HOSPITALS**	**Data mining methods like Classification model i.e. Decision tree and association model.**	**Classification analysis of patients who are stopping/terminating the treatment.**
Abdulkerim Mohammed (2013)	**Towards Integrating Data Mining with Knowledge Based System: The Case of Network Intrusion detection**	**- Data mining methods** **- JRIP algorithm** **- prolog language** **- uses system performance and user acceptance testing.**	- **Automatic knowledge acquisition for constructing knowledge base.** - **Consultation based knowledge based system developed based on the type of NW intrusion attacks.** - **Reading data mining rules from a file. So, that it facilitates automatic updating mechanism.**
Tagle Aboneh(2013)	**KNOWLEDGE BASED SYSTEM FOR PRE-MEDCAL TRIAGE TREATMENT AT ADAMA UNIVERSITY ASELLA HOSPITAL**	**- conventional knowledge acquisition methods** **- uses decision tree to illustrate rules diagrammatically** **- Prolog language to implement the prototype of the proposed system.** **- uses user acceptance and test case validation techniques to measure the performance of the system.**	- **Applicability of knowledge based system for triage patient treatment.** - **The system provides advices on the bases of patient disease during physical and laboratory examinations by human expert.**

CHAPTER 3

KNOWLEDGE ACQUISTATION USING DATA MINING

It is complex and time consuming step in the development of knowledge based systems. Implicit and explicit ways of knowledge acquisition methods are used to extract knowledge using conventional and modern techniques of acquiring that knowledge. Due to some limitations of traditional techniques; like it is time consuming, labor intensive, the information may be biased, and also they are incapable of capturing the knowledge of many years vast amount of data, experts may not explore knowledge in suitable form for decision makers; so for this research the modern technique is exercised to discover the hidden knowledge from data set. Data mining is a tool to discover this knowledge from huge amount of data set by using different classification algorithms. The aim of this research is towards on the development prototype and rule based expert system for classifying HIV patients in to different clinical and socio-demographic categories using the MVC architectural pattern. So, the rules are extracted from the patients' data by applying classification based rule induction algorithms.

3.1. DATA SELECTION AND PREPARATION

The data for this thesis work are patient records collected from the ART database of two hospitals; Asella and Adama. A total of 29,437 records of the ART databases of both hospitals are used. The dataset consists of a total of 81 attributes. These attributes are: ID, BioDataID, TBCardNo, PMTCTCode, ARTStatus, Status, BirthDate, BirthDateET, Sex, AgeInMonths, AgeInYears, UANo, HIVDate, HIVDateET, HIVTestTypeID, MaritalStatusID, EducationalLevelID, ReligionID, LocationID, Kebele, HouseNo, TelephoneNumber, WorkTelephone, MobileNo, Occupation, HealthFacilityID, ReferralID, OtherReferralReason, ReferringFacility, RegistrationDate, TransferInYN, RestartYN, ReadmitYN, FollowupDate, FollowupDateET, NextFollowupDate, NextFollowupDateET, WHOStage1, WHOStage2, WHOStage3, WHOStage4, ELDate, EligibleReasonID, ERDate, ARTStartDate, TerminatedYN, FamilyPlanningYN, PregnancyID, OACD4, IsCD4CountYN, OATLC, CohortID, OAStatus, OAWeight, OAHeight, OAWHO, M0, OriginalRegimen, CurrentRegimen, PayingYN, StoppedYN, PMTCTYN, PastARVTreatmentYN, WithParentsYN, ChildrenYN, HasFollowup, HasFamily, ARTStartETYearMonth, TODeadType, TODeadDate, PregnancyIDOnARTStartDate, PregnancyIDAtIntake, SMSContact, BMIID.

From the initial column of 81 attributes 14 of them are necessary or essential for this study by dealing with the physicians. These attributes are selected: ARTStatus, AgeInYears, Status, Sex, MaritalStatusID, EducationalLevelID, ReligionID, Occupation, TerminatedYN, FamilyPlanningYN, OACD4, OAWHO, StoppedYN, and HasFamily. The following table describes the instance of each attribute with their data types as follows:-

Table 3-1: Descriptive Summary of Attributes

No	Attribute	Meaning	Value	Data type
1	ART status	Stage of ART	EL- Eligible to be on ARV Drug ER- Eligible and Ready to start ARV Drug OA –On taking ARV drug IN- In Care for other disease	Nominal
2	Age	Age of the patient	Stage-I <5 Stage-II<10 Stage-III<19 Stage-IV>19 Stage-V<0	Numeric
3	Status	Status of the patient while starting the treatment	W- W- Able to perform usual work in or out of the house. A-Ambulatory (able to perform activities of daily living) B-Bedridden(Not able to perform activities of daily live) D- Delay(failure to attain millstone for age) R- Regression(Loss of what has been attained for age) P-Appropriate	Nominal
4	Sex	Sex of the patient	F- female M- male	Nominal
5	Marital status	Marital status of the patient	Never married Married Separated Divorced Widower Others	Numeric
6	Educational	Education level of	No education Primary	Numeric

	level	the patient	Secondary Tertiary Other	
7	Religion	Religion of the patient	Muslim Orthodox Protestant Catholic Other	Numeric
8	Occupation	Occupation of the patient	Employee Gov't employee Private employee NGO Self-employee Farmer Unemployed Other	Numeric
9	Terminate	Termination status of patients on ART	False-On ART True- Not on ART	Nominal
10	Family planning		False- has no family planning True- has family planning	Nominal
11	OAWHO	WHO Stage at which the patient is on ART	Stage 1 Stage 2 Stage 3 Stage 4	Numeric
12	OACD4	CD4 count values of the patient used as a base line to on ART	Class-I>500 Class-II>350 Class-III<500 Class-IV=0	Numeric
13	Hasfamily	True False		Nominal
14	Stopped	True False		Nominal

3.2. DESCRIPTIVE STATISTICAL SUMMARY OF ATTRIBUTES

It is used to describe each attributes based on the instances that have and how many of records are also categorized in each of those instances. The missing values are also identified filled by different missing value handling techniques if the numbers of missed records are few but in attributes where around 8746 records are missed it is difficult to replace them. As per the rule of handling missing values, these records must be removed from the data base. But removing these records means that the number of tuples be reduced in turn the expected output of knowledge discovery will not be accurate. So, researcher categorizes them as unknown and the recommendation is given to ART monitoring system workers to record the patients' data correctly. Having this limitation the descriptive summary of the selected 14 attributes are discussed as follows:-

1. ART status

As it is described in the above table, this attribute stands for the ART status stage and it has four attribute values. The frequency table for this attribute looks like the following.

Table 3-2: ART Status Frequency Table

	ART status
ART status	Frequency
EL	467
ER	47
IN	9632
OA	19290
Total	29436

As we can see from the above table, the modal value of patients ART status is failed under the category of OA. Plus to this missing attributes are not occur i.e. there is no missing value in this field.

2. Marital status

Attribute marital status has nominal value as it is mentioned in the above table. The following table illustrates the general information of each marital status values.

Table 3-3: Marital-Status Frequency Table

Marital status	
Never married	4248
Married	12107
Separated	1208
Divorced	3629
Widower	3965
Missing	4
Blank/unknown	4270
Total	29432

The missing values of the attribute are four. Since the modal value of this attribute is married and the age of patients whose marital status value missed are greater than 25, logical it is inferred that their marital status is to be married.

3. Educational level

Educational level is an attribute having six distinct values. The summary of these values will be demonstrated like this:

Table 3-4: Education Frequency Table

Educational level	
No education	6540
Primary	9801
Secondary	6952
Tertiary	1790
Others	18
Blank/ unknown	4335
Total	29436

The most frequent value for educational level is primary followed by secondary and no education respectively. From the total record 4334 patients doesn't have specific educational level value as it is summarized in the above table.

4. Religion

Table 3-5: Religion Frequency Table

Religion	
Muslim	3685
Orthodox	19120
Protestant	2158
Catholic	41
Other	93
Blank/ unknown	4339
Total	29436

As you can see there are no missing values for religion attribute. As you compare the religion attribute values the frequency of orthodox is the high (it can be the modal value of attribute religion) followed by blank and Muslim while other and catholic have small frequency.

5. Occupation

This attribute is composed of eight values (1(employee), 2(gov't employee), 3(private employee), 4(NGO), 5(self-employee), 6(farmer), 7(unemployed), 8(other), 0(blank)). No missing value is defined under this attributer. The statistical summary for the value of occupation attribute is:

Table 3-6: Occupation Marital-Status Frequency Table

Occupation	
Employee	6251
Gov't employee	935
Private employee	839
NGO	17
Self -employee	594
Farmer	2924
Unemployed	6568
Other	2871
Blank/ unknown	8437
Total	29436

Some attributes that use numeric values may actually be nominal values. For example attribute occupation has numeric values like this; (1(employee), 2(gov't employee), 3(private employee), 4(NGO), 5(self-employee), 6(farmer), 7(unemployed), 8(other), 0(blank)). But in weka environment there is a mechanism to convert such numerical fields in to nominal or categorical value using some sort of Discritization. There is an option in Weka to treat an attribute with numeric values as nominal, using the NumericToNominal option in the filters/supervised/attribute folder.

6. Status

This attribute describes the functional status of the patient whether they are ambulatory, bedridden, delay, regression or working. From the total record the value of 3671 of them are missed. The modal value for this attribute is ‘W’ or working. The researcher decided to fill the value of those missed records as working. Total of 3655 records are missed and replaced by the most frequent value which is W (working)

Table 3-7: Status Frequency Table

Functional status	
A	3544
B	1218
D	108
P	1474
R	23
W	19398
Missed	3671
Total	29436

7. OAWHO

The modal value of OAWHO attribute is OAWHO stage 3, because even though the frequency of ‘blank’ is greater than the frequency of ‘3’, it is not recommended to take consideration of that blank value as modal value.

Table 3-8: WHO Frequency Table

OAWHO	
1	1446
2	2252
3	7554
4	2076
Blank/ unknown	8454
Missing value	7654
TOTAL	29436

8. OACD4

This attribute has numeric number value starting from zero up to 2000. When you look the descriptive summary of the data there is one record which has value 3870 and the researcher takes this as a noisy data.

Table 3-9: CD4 Frequency Table

OACD4 attribute	
Missing value	11,391
Min	107
Max	3870

The value of this attribute was numeric means that 0 up to 3870. But in weka environment there is a mechanism to convert such numerical fields in to nominal or categorical value using some sort of Discritization. It is done by dividing the range of possible values into sub ranges called buckets or bins. The range is stated in the table of descriptive summary: it is ranged from stage-I up to stage-V as follows: OACD4>500; stage-IV, OACD4>350; stage-III, OACD4>200; stage-II and OACD4, stage-I.

9. Sex

It has two values i.e. male and female. There is no missing value for this attribute. When we look the descriptive summary the modal value is female as compared to males.

Table 3-10: Sex Frequency Table

Sex attribute	
Male	12464
Female	16972
Total	29436

10. Age

This attribute holds the age value of patients in numerical format. It has distinct values from 0 to 144. The mean value of this attribute is 31 while 114 is recognized as an outlier. The following table describes the statistical summary of age as follows

Table 3-11: Age Frequency Table

Age attribute	
Minimum	0
Maximum	114
Mean	31,354
Std.deviation	12, 315
Total	

In additions to this like that of OACD4 attribute, this one is also discretized from numeric to nominal value as it is categorized form descriptive table by using the following range specification:

Age<5, stage-I ;< 10, stage-II <19, stage-III ;< 24, >19, stage-IV; other, stage-V.

11. Terminated

It refers the termination status of patients to check whether they are terminated taking the treatment or not. The value for this attribute is nominal mean true or false. The true value represents the patient continues in taking the treatment and the reverse is true for the false value. In this attribute there are no missing and outlier records.

Table 3-12: Terminated Frequency Table

Terminated	
True	12,738
False	16,698
Total	29,436

12. Family planning

Family planning attribute is represented by nominal value. It has distinct values i.e. true or false. As you can see from the above table there are no missing values. The following table shows the summary information of attribute family planning.

Table 3-13: Family Planning Frequency Table

Family planning	
True	12,516
False	16,920
Total	29,436

13. Has family

The function of this attribute is to check whether the patient has family planning or not. It takes nominal value.

Table 3-14: Has-Family Frequency Table

Has family	
True	15,280
False	14,156
Total	29,436

14. Stopped

Tells whether stops taking the treatment based on the doctor decision or not. There are certain situations in which the doctor decides the patient to stop taking that treatment. As another attributes this one takes nominal value and it has true or false values. The summary information is depicted as follows

Table 3-15: Stopped Frequency Table

Stopped	
True	12,735
False	16,701
Total	29, 436

3.3. EXPERIMENTATION

After the mining process is completed, the next part is to make experimentation for deriving rules using rule based classification algorithm of WEKA tool. Two experimentations are directed by selecting two classification algorithms for the ART treatment data set. The algorithms are Ripper and PART. The sample experimentation is undertaken at the sample data set containing 29434 records and 14 attributes.

3.3.1. EXPERIMENT-1

This experiment is exercised using RIPPER algorithm involving its default value of parameters as it is indicated in the following weka classifier environment.

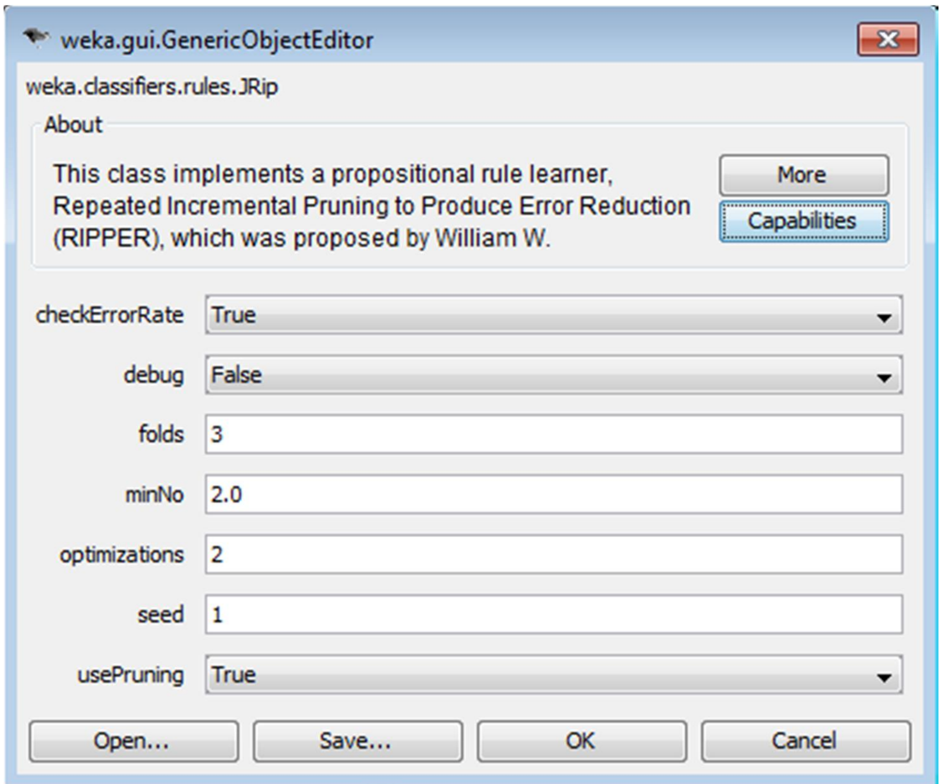


Figure 3-1: default parameter and values of RIPPER algorithm

As test option 10-fold-cross validation is selected. This algorithm correctly classified 27291, 19970, 28096, 14854, 23203, 22574, 17284, 26826, 14324, 7895, 23368, 20672, 29433 and 29433 instances out of 29436 respectively for attribute age, sex, ARTStatus, education, FamilyPlanning, HasFamily, marital Status, OACD4, OAWHO, occupation, religion, status, stopped and terminated. The number of generated rules and the time taken to generate these rules is illustrated at table 3-16.

3.3.2. EXPERIMENT-2

In the second experiment PART rule based classifier algorithm is involved. Like that of RIPPER classifier 10-fold-cross validation test option/mode is used to classify instances. The default parameter values of this algorithm under weka environment are illustrated as follows:

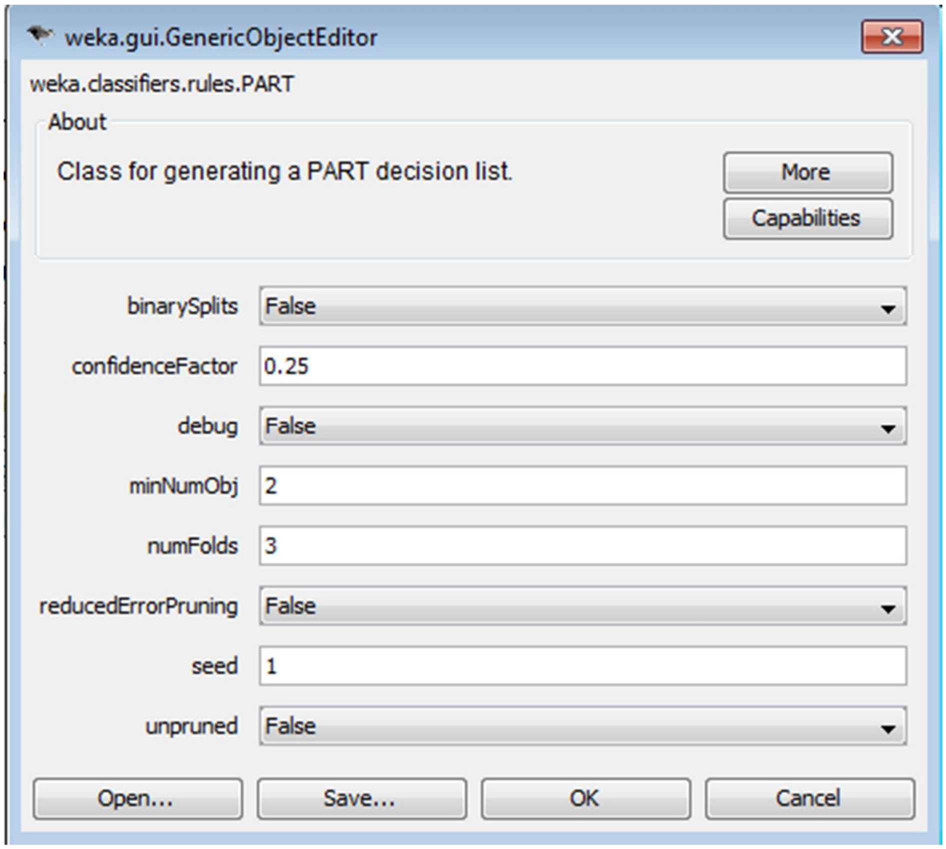


Figure 3-2: default parameter and values of PART algorithm

27504, 19854, 28472, 14807, 23013, 17986, 26843, 14575, 8497, 23244, 20454, 22324, 29433 and 29433 instances are correctly classified with respect to attribute age, sex, ARTStatus, education, FamilyPlanning, marital status, oacd4, who, occupation, religion, status, HasFamily, stopped respectively. The number of rules gained from this algorithm and the time taken to produce those rules are also illustrated at table 3-16.

3.4. PERFORMANCE RULE BASED CLASSIFIERS

The following table shows the performance to the above two experimentations based on comparisons of the number of rules they generate, the time each algorithm takes for creating rules and confusion matrix.

Table 3-16: Algorithms Comparison

Dataset attributes	Time		No of generated rules		Algorithms							
					Ripper				PART			
	Ripper	PART	Ripper	PART	Precision	Recall	F-measure	TP Rate	Precision	Recall	F-measure	TP Rate
ARTstatus	11	1.98	9	63	93	95	94	65	95	96	95	96
Status	27.16	10.04	17	400	71	79	72	79	73	95	81	79
OAWHO	10.59	7.15	14	569	60	69	60	69	64	69	64	69
OACD4	4.2	5.27	7	88	87	91	89	91	88	91	89	91
MaritalstatusID	46.93	15.27	23	846	52	58	48	58	55	60	56	60
ReligionID	4.04	8.16	7	380	70	79	70	79	68	78	71	78
EducationlevelID	9.07	39.39	15	1149	55	50	42	50	49	50	48	50
Occupation	19.52	22.04	20	1054	38	37	28	37	37	40	37	40
Age	11.8	4.29	12	114	90	92	91	92	92	93	92	93
Sex	48.58	22.42	14	709	67	67	67	67	67	67	67	67
Stopped	0.49	0.18	2	2	1	1	1	1	1	1	1	1
Terminated	0.65	0.06	2	2	1	1	1	1	1	1	1	1
Hasfamily	14.47	8.19	8	426	77	76	76	76	76	76	76	76
Familyplanning	58.65	9.74	22	471	79	78	78	78	78	78	78	78

While time is an effective performance measure that estimates the time taken by each algorithm for generating the classification; the number of generated rules counts the total number of rules generated by each algorithm.

As it is illustrated in the above table; the JRIP algorithm takes average time 19.08 second while PART takes 11.01 second's to classify attributes.

The performances of those algorithms are also assessed by confusion matrix result. Within little difference the PART algorithm has better performance from that of JRIP.

CHAPTER 4

DESIGN OF THE PROPOSED SYSTEM

In this chapter we will discuss the possible architectural design of the RBES, the main components and their interactions.

4.1. GENERAL STRUCTURE OF RULE BASED SYSTEM

Every knowledge base decision support systems have basic components of knowledge base, inference engine and user interface [19]. The knowledge in the knowledge base may include implicit and explicit knowledge from an expert, formal knowledge such as that found in texts or legislation, heuristic knowledge, procedural (problem solving knowledge) and declarative (factual) knowledge. There are many different ways of representing this knowledge including logic, production rules, semantic nets and frames. The inference engine acquires necessary information from the expert system user through the user interface, controls the way the knowledge in the knowledge base is accessed and communicates results to the user, again through the user interface. The user interface can be thought of as simulating the interaction which would occur when there is a consultation between a human expert and a seeker after advice. This simulation is virtual rather than substantive, as though some systems provide forms of natural language processing, most systems use interface methods similar to those in other programs: menu, question and answer, form fill-in and graphical interfaces.

In this study, the fundamental components of the system are identified in to three layers: model/application logic layer, view and controller/presentation layer. Figure 5.1 shows the general architecture of RBES of the proposed system.

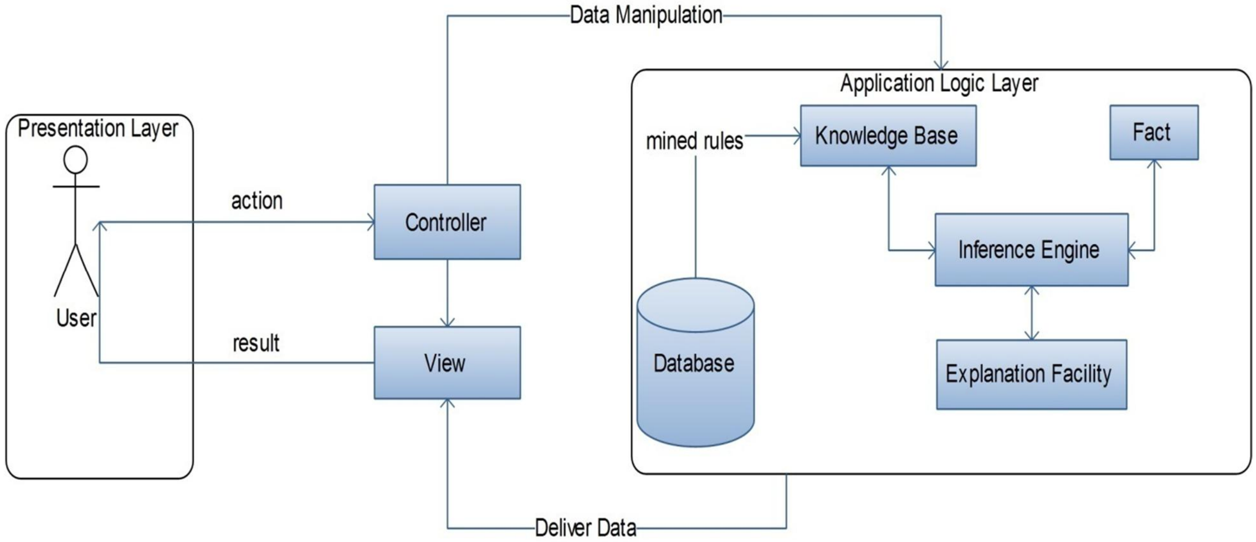


Figure 4-1: Proposed System Architecture extended from [25] [35]

4.2. APPLICATION LOGIC LAYER

This is the domain layer that encapsulates basic parts of the application of this research. This layer has direct communication with controller and indirect communication with view or presentation layer. The database, knowledge base, inference engine, fact and explanation facility are components laid at this layer. The detailed discussions on these components are below:

4.2.1. DATABASE

The database component contains ART data and the concept of data mining is required to mine and generate rules based on the filtered data. In this research huge volume of patient's data are collected from ART database starting from 1993 E.C and around 29,436 patients are starting antiretroviral therapy treatment. Different data mining tools, techniques and algorithms are used to mine the data and generate the required rules.

4.2.2. KNOWLEDGE BASE

It is the basic component in the development of knowledge based system. In addition, it communicates with the inference engine. In chapter two when we talk about architecture of expert system we have said that knowledge is the most important component of expert systems. So to have these knowledge different types of knowledge acquisition methods are there as it is discussed in previous chapter. So, knowledge is constructed based on the rules gained from that rule based algorithm in addition to certain clarifications gained from the expert. These rules have the following sequences and expression form:

If <conditions>, then <actions>

If the conditions are true then, the actions are executed. When rules are examined by the inference engine, actions are executed if the information supplied by the user satisfies the conditions in the rule. Conditions are expressions involving attribute and logical connective 'and

1. CLASSIFICATION BY THE LEVEL OF EDUCATION OF PATIENTS

The PART algorithm derives 1149 rules based on the classification by education by relating education with other instances as it is visible in the below table. Sample rules are taken from the generated rules for developing the prototype part. The following are sample rules incorporated in the knowledge base by 'FamilyPlanning' classification:

No	Rule	Conclusion
1	(status W) (has-family true) (art-status OA) (stopped false) (age stage-IV) (sex M) (marital-status married)	True
2	(art-status IN) (who others) (oacd4 class-IV) (terminated true) (marital-status married) (occupation employee)	False
3	(art-status IN) (who others) (oacd4 class-IV) (has-family false)	False

	(education primary)	
4	(status W) (has-family true) (art-status OA) (stopped false) (age stage-IV) (who others) (oacd4 class-III)	True

2. CLASSIFICATION RULE BY STATUS OF PATIENTS

400 rules are generated by PART algorithm by classifying the data based on status of patients. From them, in this document 4 of them are included. The following table shows that:-

No	Rule	Conclusion
1	age stage-IV religion orthodox who stage-3 art-status OA education primary occupation farmer sex M has-family false	A
2	age stage-IV who stage-4 religion others terminated true education others	B
3	AgeInYears = stage-IV AND WHO = others AND ARTStatus = IN AND TerminatedYN = FALSE AND Occupation = Employee	W
4	AgeInYears = stage-IV AND WHO = others AND ARTStatus = IN AND TerminatedYN = FALSE AND Occupation = Employee:	W

4.2.3. INFERENCE ENGINE

Inference engine is responsible for controlling the way knowledge in the knowledge base accessed and communicates the result via user interface to the user. Based on the user request it matches rules' using forward chaining mechanism and it works by Rete algorithm and the matched rule is fired. The following diagram simulates it.

4.2.4. EXPLANATION FACILITY

This component creates better communication between the user and system. When there are certain ubiquities' of the user with the system, using what and why questions, it tries to make clear. For example, the system gives command to user to insert the "religion" of the patient. Based on that particular rule is executed and the query result is displayed. But the user doesn't know the paths to reach the final conclusion. Consequently the users ask question "why" this result is reached. So, the system tries to clarify the steps passed to reach that conclusion.

4.2.5. FACT

Fact is a user query. When the user inputs query using the user interface query dialog, at that time the fact is asserted. It can be stored in the working memory in three forms: ordered unordered and shadow facts. In this proposed system unordered facts are used. The inserted facts are matched with the knowledge based rules. The matched rules are stored in the activation list.

4.3. PRESENTATION LAYER

This layer presents the model in a way that is suitable for the user via user interface. The user input their query using user interface elements and the output is also delivered at this layer. When the user inputs the query via view, the data manipulation be performed by the application layer but for the data to be manipulated the controller takes action. So, there is communication between presentation layer and application logic layer via controller for the purpose of data manipulation and also that communication is enhanced via view for delivering the output.

4.4. PROPOSED SYSTEM MODELING

The proposed system is modeled to depict functional requirements of the system in to pictorial representation. And it is clear that the requirement of this research work is elicited from the gap of previously conducted related research works. UML is usually used for modeling object oriented systems. It can also be used for modeling other types of systems such as frame based and rule based systems [21]. In modeling this system with the help of UML, use case and sequence diagrams are used for this research.

4.4.1. USE CASE MODELING

It is used to show the interaction of the user with the system. It provides a way to users and developers to communicate easily. It contains the use cases, actors and their relationship. In this research six uses cases are identified. The following diagrams represent each of them:-

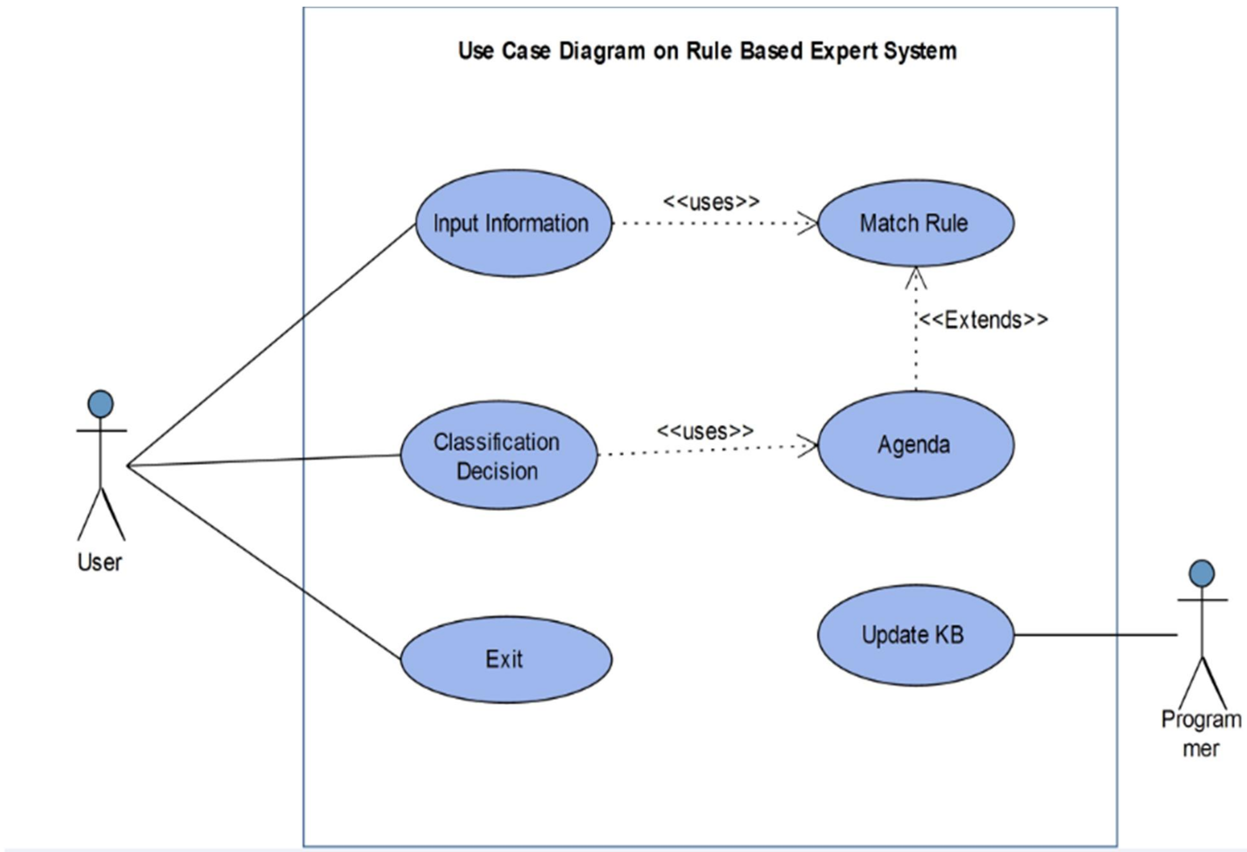


Figure 4-2: use case diagram of rule based expert system

Use case ID	UC-01
Use case name	Input information
Actor	User
Description	Describes the information or query that the user feeds
Precondition	None
Post condition	User query is captured
Basic course of action	1. User wants to input query

	2. System displays query dialog 3. User inputs query 4. User submit the query 5. Use cased ends

Use case ID	UC-02
Use case name	Match rules
Actor	User
Description	It is the use case used by input information use case. Here rules are stored to be checked based on the input information
Precondition	None
Post condition	Matched rule is activated
Basic course of action	1. The system retrieves rules from memory 2. System store match rules in to agenda 3. Use case ends

Use case ID	UC-03
Use case name	Agenda
Actor	User
Description	This describes activated rules which are stored in agenda can be extended and used by the classification use case.
Precondition	None
Post condition	The activated rule is fired
Basic course of action	1. The system rank activated rules in last in first activated order 2. Use case ends

Use case ID	UC-04
Use case name	View classification decision
Actor	User
Description	It enables the user to view the response of input information.
Precondition	None
Post condition	View classification result

Basic course of action	1. The user wants to view the classification result. 2. The system fires activated rules based on the ranking order 3. Use case ends

Use case ID	UC-05
Use case name	Exit
Actor	User
Description	The system user decides when to exit from the system
Precondition	None
Post condition	Exited from the system
Basic course of action	1. Use case ends

Use case ID	UC-06
Use case name	Update KB
Actor	Programmer
Description	This describes the scenario where the programmer updates knowledge base
Precondition	None
Post condition	The knowledge based is updated
Basic course of action	1. The user wants to modify the stored knowledge 2. The programmer build the new knowledge 3. The programmer update the knowledge base 4. Use case ends.

4.4.2. SEQUENCE DIAGRAM

The UML sequence diagram is used to describe the flow of messages, events and actions between objects as it is depicted in the above figure.

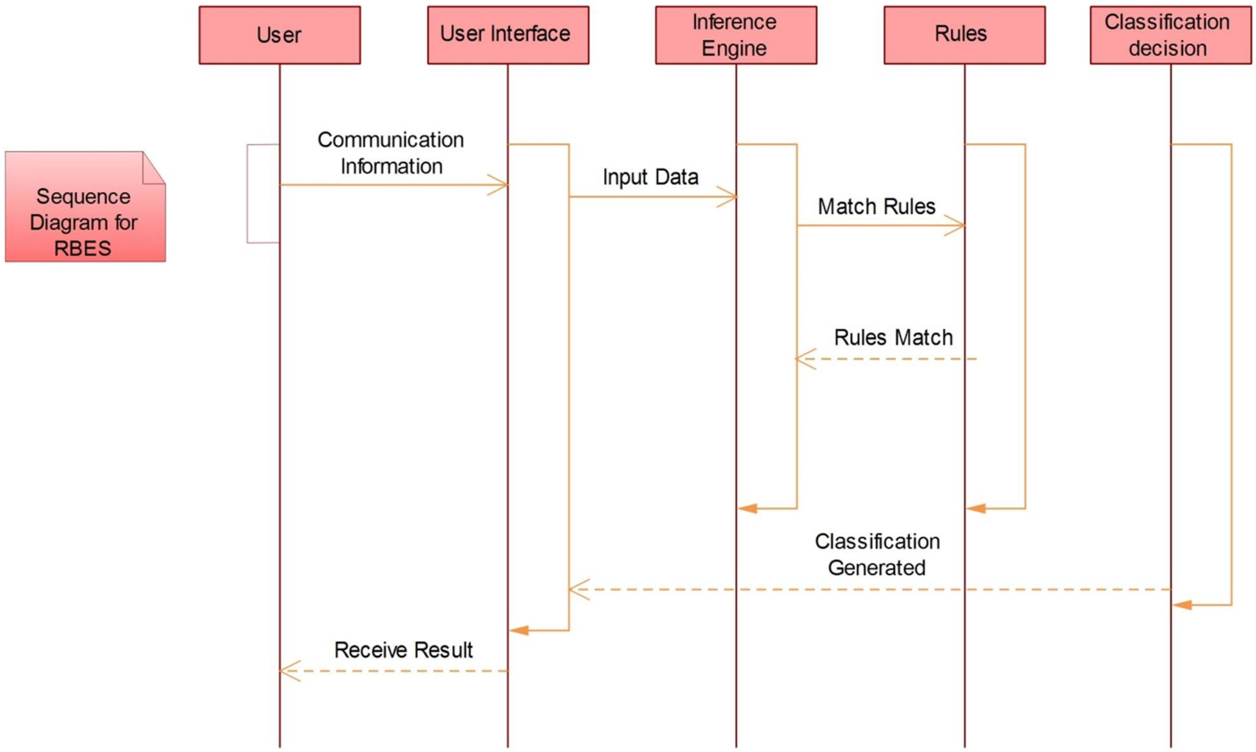


Figure 4-3: Sequence Diagram

The user communicates with the user interface module by in putting the query data. The inference engine module match query data with the rules stored in the working memory. The classification module sends the generated classification result to the user via user interface; so that the user receives the result.

CHAPTER 5

IMPLEMENTATION OF THE PROPOSED SYSTEM

The goal of this chapter is to demonstrate the implementation of the proposed system plus to that it briefly describes the software tool used to implementation purpose with the basic reason of why this software tool is selected.

Even though, there are different languages to implement rule based expert system as it is mentioned in the literature review part, the focus of this thesis implementation is on using JESS. The detailed discussion of this part is as follow:

5.1. JESS

As compared CLIPS with JESS, jess is more popular in the application which can be put in the applet-form. The future of JESS depends highly on the evolution of the web, the JAVA programming language and its own future stability. These three conditions make that there is a great possibility that JESS will become more popular and more frequently used. Especially the object-oriented possibilities and the easy integration into JAVA code makes JESS' future very promising. It also provides support for modular development of the rules. Lastly it can fire more than 80,000 rules per second and it can perform almost 600,000 pattern matching operations per second and also can insert more than 100,000 facts per second [22]. So it is fast and has the advantage of asserting large amount of rules and facts at a time. There for this thesis Jess is selected to implement the prototype of the proposed system as per its objective.

5.2. ARCHITECTURE OF JESS

We have discussed the components of rule based expert system in the previous chapter at section 5.1, based on that the Jess architecture is composed of those components. The following diagram illustrates the architecture of Jess and how it works.

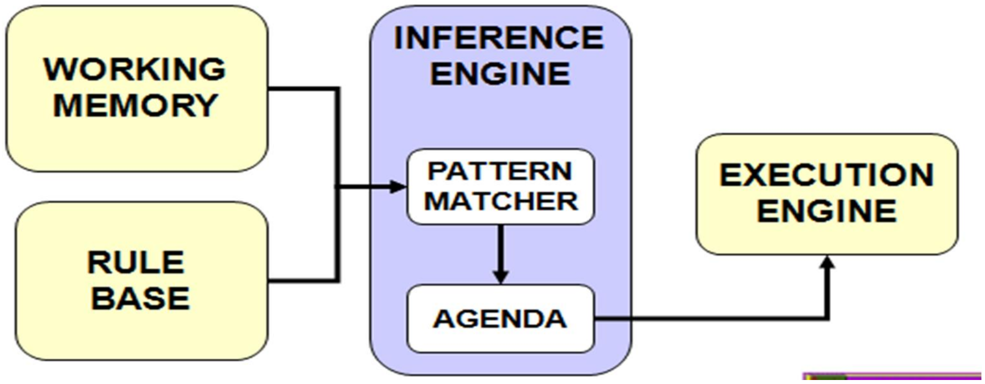


Figure 5-1: The JESS Architecture [33]

Working Memory: - the working memory contains a collection of facts; for this reason sometimes called fact base. Facts are information that the rule base system is working with and they can be used as premises (LHS) and conclusions (RHS) of the rules.

Rule Base: - contains all rules that the engine knows. In rule-based systems, rules are sometimes stored as strings of text, but Jess holds a rule compiler for processing them into some format that the inference engine can manage more efficiently. Particularly, the Jess ' s rule compiler builds a Rete Network, which is a complex and indexed data structure for speeding up rule processing.

Inference Engine: - it consists of pattern matcher and the agenda. It decides what rules to fire when.

The pattern matcher decides which rules to activate based on the current contents of the working memory. A rule is activated when the pattern matcher finds facts that can satisfy the RHS of this rule. This is not a trivial task if we take into account that the working memory may contain thousands of facts and the rule base contains complicated rules with several premises and conclusions. In these cases, the pattern matcher might need to search through millions of combination of facts to find those combinations that satisfy rules. Fortunately, efficient ways exist to solve the problem, since a lot of research has focused on this area. The Jess uses Rete algorithm to match patterns.

The agenda stores the list of rules that could be potentially fired; schedules the order in which activated rules will fire. The agenda consists of an ordered list of rules, whose RHS can be executed. The agenda has to decide which rules have the highest priority and should be fired first.

Execution Engine: - is responsible for firing rules. It is the component that executes the RHS of the rule once the inference engine decided what rule to fire. Jess offers a complete programming language to define what happens when a given rule fires, whereas other rule-based systems just offer the possibility of adding, removing, and modifying facts in the working memory. To execute rules based on facts for this research backward chaining is used.

The Jess rule engine works in discrete cycles consisting of three steps, as we can see from the above diagram;

1. All the rules contained into the rule base are compared to the working memory in order to decide which ones should be activated during this cycle. The list of these activated rules, together with any other rules activated in previous cycles, is called the conflict set.
2. The conflict set is ordered to form the agenda. This process is called conflict resolution.
3. To complete the cycle, the first rule on agenda is fired and the entire process is repeated again.

5.3. IMPLEMENTING THE SYSTEM USING JESS

To develop the prototype of rule based expert system for classifying HIV patients with Jess, Eclipse IDE was used. As the MVC architecture is used to implement the system, in the application logic layer the knowledge is represented by Jess language while the view and control are by java.

5.3.1. REPRESENTING THE KNOWLEDGE BASE

Based on the structure of knowledge based expert system discussed in the previous chapter, the knowledge base does have rules, facts, inference engine: pattern matcher, agenda and explanation facility. 182 rules are selected from the total rules extracted from the PART algorithm from 14 attributes to implement the prototype of this research. These rules are defined by the defrule keyword or the language and the assert function is used to assert the new attribute value if LHS variables are satisfied. In the design part 4 rules are taken as a sample from attribute FamilyPlanning and status. The following example illustrates the equivalent Jess code of those rules.

```
(defrule famprule5

    (status W)
    (has-family true)
    (art-status OA)
    (stopped false)
    (age stage-IV)
    (who others)
    (oacd4 class-III)
    =>
    (assert (family-planning true))
    (bind ?*trueCount* (+ ?*trueCount* 1))
    (bind ?*trueInstance* (+ ?*trueInstance* 1697))


  )
```

```
(defrule statrule6

    (religion orthodox)
    (art-status OA)
    (who stage-3)
    (marital-status married)
    =>
    (assert (status W))
    (bind ?*BBBCount* (+ ?*BBBCount* 1))
    (bind ?*BBBInstance* (+ ?*BBBInstance* 17))
    )
```

```
(defrule famprule9

    (art-status IN)
    (who others)
    (oacd4 class-IV)
    (terminated true)
    (marital-status married)
    (occupation employee)
    =>
    (assert (family-planning false))
    (bind ?*falseCount* (+ ?*falseCount* 1))
    (bind ?*falseInstance* (+ ?*falseInstance* 285))

  )
```

```
(defrule statrule8

    (age stage-IV)
    (religion orthodox)
    (who stage-3)
    (art-status OA)
    (education primary)
    (occupation farmer)
    (sex M)
    (marital-status married)
    =>
    (assert (status A))
    (bind ?*AAACount* (+ ?*AAACount* 1))
    (bind ?*AAAInstance* (+ ?*AAAInstance* 31))

    )
```

The bind keyword is used to count and add the number of instances in which the fact is exactly matched from the total number of records.

5.3.2. THE USER INTERFACE OF THE SYSTEM

The acceptability of the knowledge based system depends on the quality of the user interface. It is built from the clinical and socio-demographic category of patients. The javax.swing library is used to develop this interface. It has three steps. So, the user interacts with the system through button click by selecting the clinical or socio-demographic values of each attribute

STEP-1-: CHOOSE AN ATTRIBUTE TO CLASSIFY WITH

At step1, the user chooses an attribute in which classification is built. At a time one attributed could be selected. For example to classify patients by their FamilyPlanning, the FamilyPlanning radio button must be selected. Then to input known facts the user must click the "submit" button". In our previous example, the FamilyPlanning attribute is selected to classify...

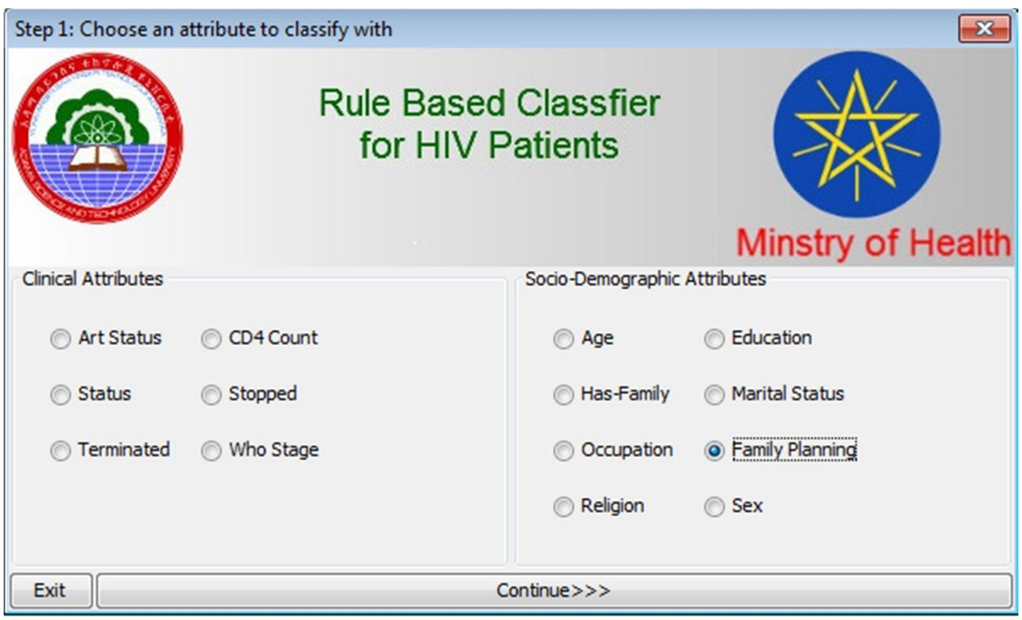


Figure 5-2: Attribute Selection

STEP-2: INPUT KNOWN FACTS

Step2- this is the window in which the user input facts. This is similar with the query dialog. After the user inputs necessary attributes, can continue to the last step. For example, we input the facts which are available in rule-1 as follows:

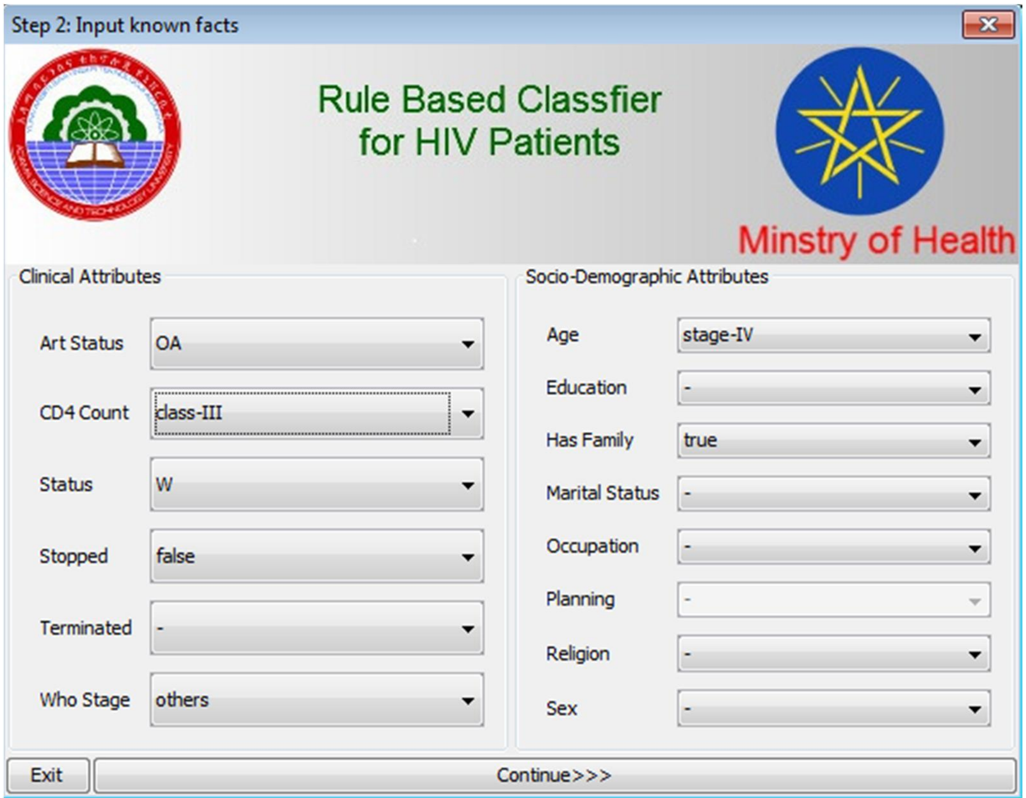


Figure 5-3: Facts Entry

STEP -3: SHOWING CLASSIFICATION RESULT

Next to fulfilling the input values at the second step, the user can view the classification result by clicking on the "continue" button. In this window: attribute values, number of rules and number of instances are stated.

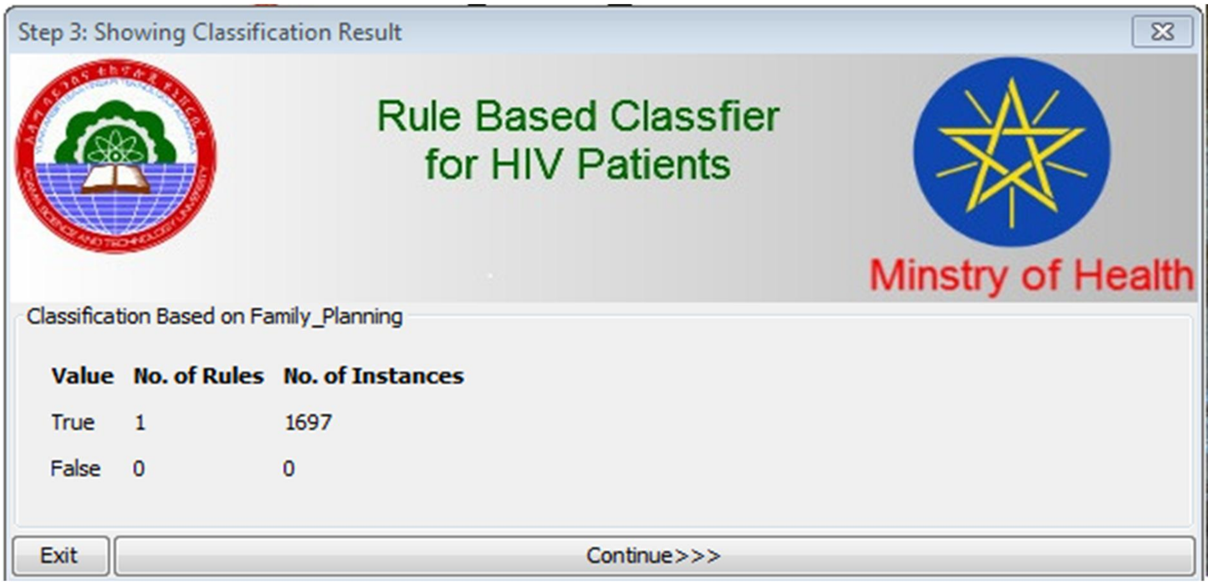


Figure 5-4: Classification Result

As it is illustrated from classification result window, the selected attribute for classification is "FamilyPlanning". It has two attributes having value 'True/False'. The fact inputted by the user is matched with the total of 1697 rows/instances but this rule is exactly matched with only one rule and the result is classified as "True".

CHAPTER 6

TESTING, ANALYSIS AND INTERPRETATION

After implementing the knowledge based system, it must be tested to measure the performance of the system whether is achieve the objective or not. It is the final step of the research. System and user acceptance testing is carried on the basis of interaction of the user with the system.

6.1. SYSTEM TESTING

As mentioned in the design part, the proposed system is developed by integrating java expert system shell and java. Therefore, it contains half java code and half JESS scripts. The rules of the system are written in JESS. The JESS working memory gets facts from an input on a java user interface. Once facts are in the JESS working memory then the inference engine activates rules to the agenda. Then based on last activate first fire rule, it fires activated rules. Those fired rules may create additional facts to the working memory. Finally the required data from the working memory is sent back to Java.

In system testing the following steps were tested.

STEP 1: TESTING RULES SYNTAX AND CORRECTNESS

The classification rules are generated by Weka; those rules were required to be input into the rule based classifier. So rules were selected and programmed using JESS syntax. Totally, there are 6 clinical attributes and 8 socio-demographic attributes. The rules for each of the attributes are kept in a separated ".clp" file.

Basically, two things were tested in this step. One is freeing up the rules from any syntax errors and the other one is the consistency of the templates used to assert facts. For example, "occupation" is used to identify occupation facts, "who" is used to identify who stage facts. Facts like this: (occupation govt-employee) is used to represent government employees, (sex F) is used for represent female patients. Such kinds of formats are needed to be used consistently throughout all the rules. To test the above issues all rules were revised again and again plus the JESS compiler is used to correct syntax errors.

STEP 2: SELECTING CLASSIFICATION ATTRIBUTE

The system classifies given data on a single attribute at a time, due to this the user is requested to input an attribute which is going to be used for classification. A Java Swing GUI which has a set of radio buttons appears on the screen. Therefore the user is allowed to choose one attribute only.

Here, the first thing tested is, whether all attributes are correctly available in the list of choices. The second thing is that based on the selected attribute, suitable JESS script (contains rules) should be executed. For instance, if art-status is selected then the JESS script which contains art-status classification rules should be executed. The following things were implemented as shown in the code below.

```
if (Art_Status.isSelected())
{
             engine.executeCommand("(batch Scripts/art.clp)");
             selected="Art_Status";
}
elseif (CD4.isSelected())
{
             engine.executeCommand("(batch Scripts/cd4.clp)");
             selected="CD4_Count";

}
```

The system starts from the main function, resets the working memory, clears all facts and invokes the (watch all) function which helps to visualize everything that happens in rule based classifier. Then it creates the first dialog for attribute selection.

As shown in the figure below, all the 14 attributes are available for choice. As a test case, religion is selected.

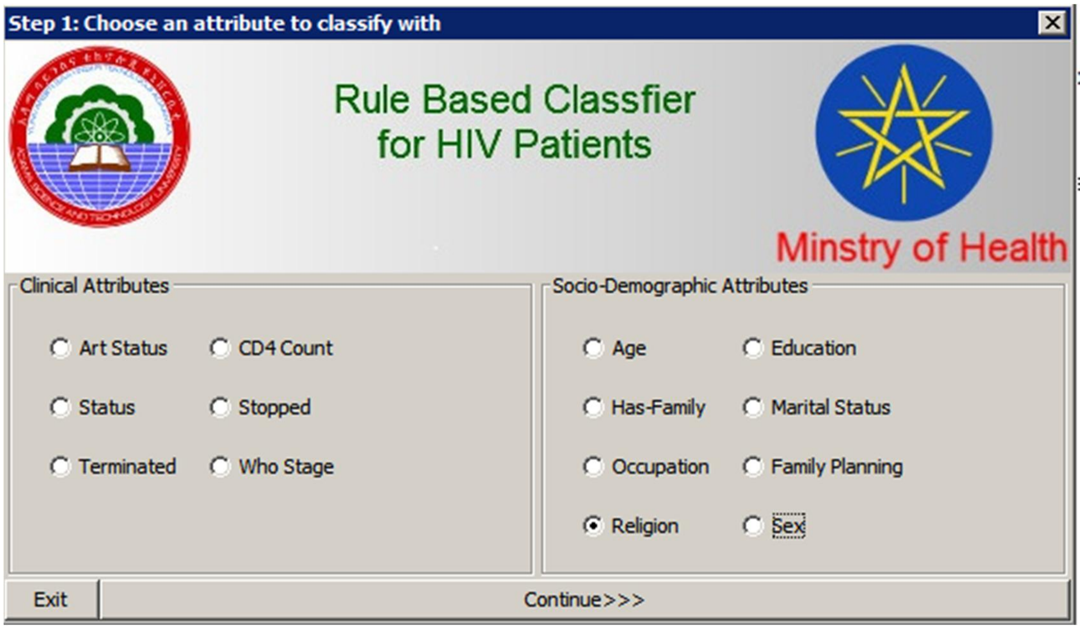
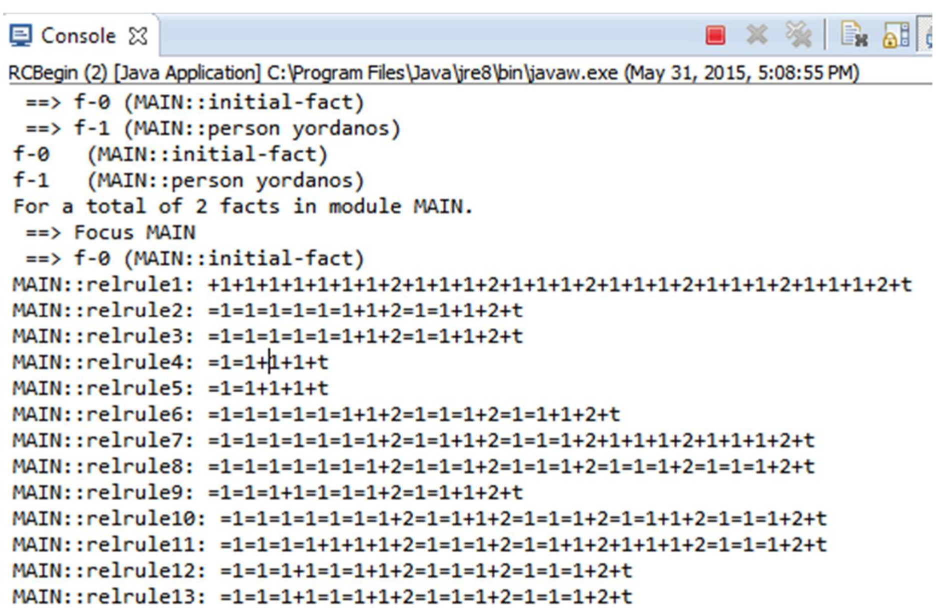


Figure 6-1: Attributes Selection Test Case

STEP 3: CREATE FACTS FROM JAVA TO JESS WORKING MEMORY

Once the religion attribute is selected in the previous step, the requested .clp file is loaded and the rules inside it are executed. This can be confirmed from the console window. The console window shows the internal representation of the defined rules.



```
Console
RCBegin (2) [Java Application] C:\Program Files\Java\jre8\bin\javaw.exe (May 31, 2015, 5:08:55 PM)
 ==> f-0 (MAIN::initial-fact)
 ==> f-1 (MAIN::person yordanos)
f-0   (MAIN::initial-fact)
f-1   (MAIN::person yordanos)
For a total of 2 facts in module MAIN.
 ==> Focus MAIN
 ==> f-0 (MAIN::initial-fact)
MAIN::relrule1: +1+1+1+1+1+1+1+2+1+1+1+2+1+1+1+2+1+1+1+2+1+1+1+2+1+1+1+2+t
MAIN::relrule2: =1=1=1=1=1=1+1+2=1=1+1+2+t
MAIN::relrule3: =1=1=1=1=1=1+1+2=1=1+1+2+t
MAIN::relrule4: =1=1+1+1+t
MAIN::relrule5: =1=1+1+1+t
MAIN::relrule6: =1=1=1=1=1=1+1+2=1=1=1+2=1=1+1+2+t
MAIN::relrule7: =1=1=1=1=1=1+2=1=1+1+2=1=1=1+2+1+1+1+2+1+1+1+2+t
MAIN::relrule8: =1=1=1=1=1=1+2=1=1=1+2=1=1=1+2=1=1=1+2=1=1=1+2+t
MAIN::relrule9: =1=1+1=1=1=1+2=1=1+1+2+t
MAIN::relrule10: =1=1=1=1=1=1+2=1=1+1+2=1=1=1+2=1=1+1+2=1=1=1+2+t
MAIN::relrule11: =1=1=1+1+1+1+2=1=1=1+2=1=1+1+2+1+1+1+2=1=1=1+2+t
MAIN::relrule12: =1=1=1+1=1=1+1+2=1=1=1+2=1=1=1+2+t
MAIN::relrule13: =1=1=1+1=1=1+1+2=1=1=1+2=1=1=1+2+t
```

Figure 6-2: Fact Creation Window

The above rules are neither activated nor fired, there are just defined or introduced.

The facts entry dialog is shown in this step. In this dialog all the 14 attributes except the one selected as the classification attribute are displayed and enabled. One crucial issue which was tested here was matching the templates and formats of facts in the rule definition with the choices given in the JCombo boxes. This had to be checked because a user chooses an item for each attributes which in turn is used for creation of facts in the working memory. And those facts are compared with the left hand side of the rules. As a test case some attributes are given value on the facts entry dialog shown below.

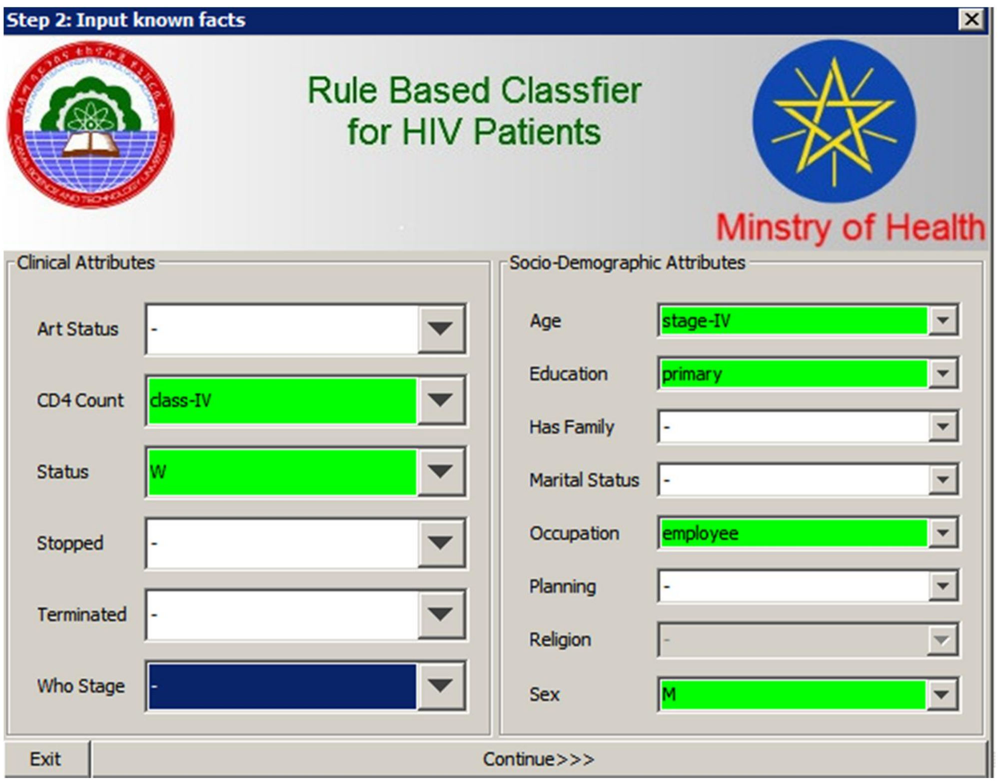


Figure 6-3: Facts Entry Test Case

Then when the continue button is clicked the facts should be created. According to the RETE algorithm, rules are activated if all conditions on the left side are matched. So, creation of facts often initiates the activation and firing of rules. In this test case, we can confirm that all facts are correctly created; matched rules are activated and fired. It is shown in the console output below.

```
 ==> f-1 (MAIN::oacd4 class-IV)
 ==> f-2 (MAIN::status W)
 ==> f-3 (MAIN::age stage-IV)
 ==> f-4 (MAIN::education primary)
 ==> f-5 (MAIN::occupation employee)
 ==> f-6 (MAIN::sex M)
==> Activation: MAIN::relrule7 :  f-3, f-4, f-6, f-2, f-5, f-1
FIRE 1 MAIN::relrule7 f-3, f-4, f-6, f-2, f-5, f-1
 ==> f-7 (MAIN::religion orthodox)
 <== Focus MAIN
f-0   (MAIN::initial-fact)
f-1   (MAIN::oacd4 class-IV)
f-2   (MAIN::status W)
f-3   (MAIN::age stage-IV)
f-4   (MAIN::education primary)
f-5   (MAIN::occupation employee)
f-6   (MAIN::sex M)
f-7   (MAIN::religion orthodox)
For a total of 8 facts in module MAIN.
```

Figure 6-4: Created Facts for Test Case

As it can be seen, five new facts were created first. These facts were all part of the left hand side of relrule7. The rule is activated and fired. When the rule is fired it creates a new fact which is f-7. So the patient is classified as orthodox due to one matched rule.

STEP 4: TRANSFERRING JESS RULES DATA TO JAVA

In this step, the data about the fired rules is computed in JESS and transferred to Java to display the summary. Therefore, for the test case the following rule was fired.

```
(defrule relrule7
    (age stage-IV)
    (education primary)
    (sex M)
    (status W)
    (occupation employee)
    (oacd4 class-IV)
    =>
    (assert (religion orthodox))
    (bind ?*ORTHCount* (+ ?*ORTHCount* 1))
    (bind ?*ORTHInstance* (+ ?*ORTHInstance* 500))
  )
```

In the right hand side of this rule a two counter variables are available. The first one increases the number patients classified under orthodox religion (based on the given facts) by one. The second one increments the number of instances/rows matched by the LHS parameters of matched rules in the full dataset. Therefore the number 500 is taken for the Weka classifier data. For this test case only one rule is satisfied, therefore the first count is 1. The total amount of instances is also only 500 that is taken from the single rule. If they were other rules which classify the patient to the same category, then number of rules and instances would have been increased.

So those JESS variables are transferred to Java using code like the shown below.

```
elseif (ClassfDialog.selected.equals("Religion")) {
       engine.executeCommand("(store musCount ?*MUSCount*)");
       engine.executeCommand("(store orthCount ?*ORTHCount*)");
       engine.executeCommand("(store proCount ?*PROCount*)");
       engine.executeCommand("(store catCount ?*CATCount*)");
              engine.executeCommand("(store othsCount ?*OTHSCount*)");

       engine.executeCommand("(store musInstance ?*MUSInstance*)");
       engine.executeCommand("(store orthInstance ?*ORTHInstance*)");
       engine.executeCommand("(store proInstance ?*PROInstance*)");
       engine.executeCommand("(store catInstance ?*CATInstance*)");
       engine.executeCommand("(store othsInstance ?*OTHSInstance*)");

elseif (ClassfDialog.selected.equals( "Religion")) {
       rows=5;
       resultPanel.add(newJLabel("Muslim"));
       resultPanel.add(newJLabel(engine.fetch("musCount").toString()));
       resultPanel.add(newJLabel(engine.fetch("musInstance").toString()));

       resultPanel.add(newJLabel("Orthodox"));
       resultPanel.add(newJLabel(engine.fetch("orthCount").toString()));
       resultPanel.add(newJLabel(engine.fetch("orthInstance").toString()));

       resultPanel.add(newJLabel("Protestant"));
       resultPanel.add(newJLabel(engine.fetch("proCount").toString()));
       resultPanel.add(newJLabel(engine.fetch("proInstance").toString()));

       resultPanel.add(newJLabel("Catholic"));
       resultPanel.add(newJLabel(engine.fetch("catCount").toString()));
       resultPanel.add(newJLabel(engine.fetch("catInstance").toString()));

       resultPanel.add(newJLabel("Others"));
       resultPanel.add(newJLabel(engine.fetch("othsCount").toString()));
       resultPanel.add(newJLabel(engine.fetch("othsInstance").toString()));
}
```

The results are correctly shown in the results dialog as it can be seen in the next figure

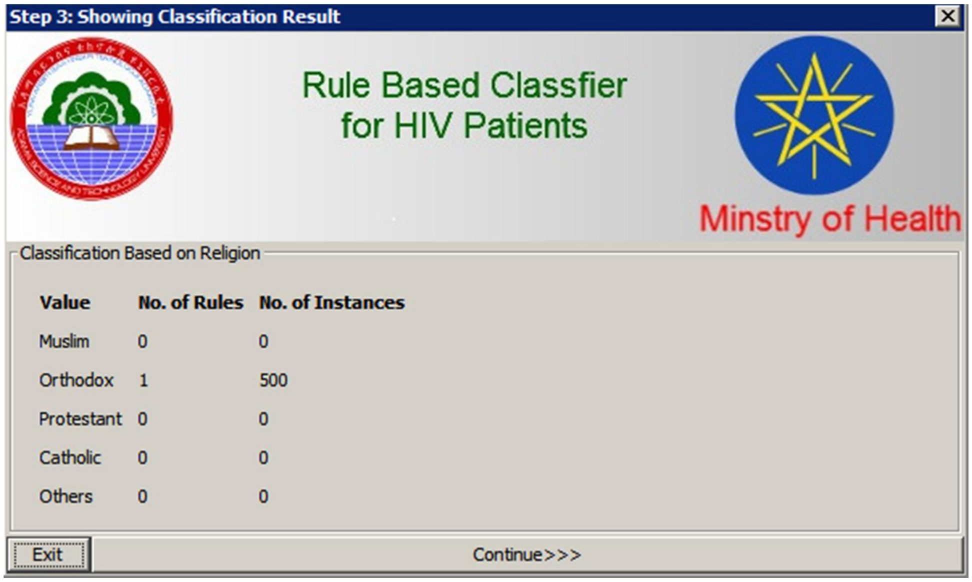


Figure 6-5: Result Dialog for Test Case

6.2. USER ACCEPTANCE TESTING

It is conducted to assess whether the system can support decision makers in their decision making process and also it ensures the accuracy of the system. The questionnaire which has five open ended question is prepared and conducted for experts and the result is analyzed as follows:

No	Question	Poor (1) 2	Fair (2)	Good (3)	V.good (4)	Excellent (5)	Average	Total Average performance
1	Is the user interface attractive?	0	0	1	2	2	4.2	84
2	Is the system efficient in time?	0	0	1	1	3	4.4	88
3	How much accurate the system is accurate in categorizing the patient in to correct category	0	0	1	1	3	4.4	88
4	Does the system incorporated sufficient knowledge to make in decision in a given category	0	0	1	2	2	4.2	84
5	Does the system have contribution in the domain area	0	0	1	2	2	4.2	84
6	Simplicity to use and interact with the System	0	0	1	2	2	4.2	84
							4.26	85.33%

As the users of the system evaluator's response, 60% of them are also weighted as V.good and the same is true for Excellent. In the same way for the second question 40% of the respondents respond as good while 60% of them are rated as V.good. For the third question, 40% of them are said the system is accurate in categorizing patients in to their correct classification group. The remaining 40% and 20% are rated as V.good and excellent respectively. At the same time for criteria "Does the system incorporated sufficient knowledge to make in decision in a given category ", 20% respondents said good while 40% and 40% of them weighted as V.good and excellent. As we look t fifth question, 100% of the respondents respond that the system has contribution in the domain area. The last one is to check the simplicity of the user interface when the user interacts with the system and it has got the vote of 100% as excellent.

To calculate the average performance (AVP),

$$\text{Scale value a} * \frac{\textit{number of respondents at scale a}}{\text{total number of respondents}} + \text{Scale value b} * \frac{\textit{number of respondents at scale b}}{\text{total number of respondents}} \ldots$$

Where, scale value is the numerical representation value of poor, fair, good, V.good and excellent.

So, based on the response of respondents the average performance is obtained 4.2 of the five scales while it is 85.33% as it is described by average percentage performance. This means that the performance of the proposed RBES is accepted by users at the rate of 85.33%.

Generally, even though; the combined result of system and user acceptance testing is good, the system testing is enhanced by the programmer since the users of the system doesn't have know-how about the programming environments. But simply by training them on how they can use the system and the functionalities of each buttons which are found in the user interface part, the corresponding questionnaire is delivered for them to evaluate performance of the system. So, the system testing is conducted by the researcher by putting different testing scenarios as it is discussed in the system testing section of testing, analysis and interpretation chapter.

CHAPTER 7

CONCLUSION AND RECOMMENDATION

7.1. CONCLUSION

To reduce the mortality of HIV/AIDS patients different strategies are introduced in the world with the help of WHO and UNAIDS. From those the one is ART drug. In adherence of ART service especially in African countries so many studies are conducted to scale up it and the same is true in our countries. But beyond this there is a gap in identifying patients in to different classes like based on their age, sex, educational level, occupation, religion, ARTstatus, CD4 count, marital status, stop, terminate and so on and the relationship between those. It creates information gap for decision makers.

So to address the above problem or gap the RBES for classifying patents in to clinical and socio-demographic categories is proposed in turn which is basically depend on the integration data mining, knowledge based system and some points of software engineering. The system aims to give direction for the decision makers in improving the current decision support system which is based on report generated monthly. At knowledge acquisition phase the knowledge is extracted from rule based machine learning induction algorithms. The most tedious task is passed at data preprocessing step especially in data Discritization part, detecting outliers and handling missing values. Generally, different challenges are identified in each stage of this thesis.

The proposed rule based system is modeled using PART algorithm rules and the prototype is implemented by JESS programming language. Most of the research time is elapsed during to be familiar with this programming language.

Finally the following conclusions are drawn from the finding with regard to the research questions:

- The proposed RBES solves the current challenges of existing decision support system.
- Knowledge's are extracted from the data mining induction algorithms and represented in the form of if then rules.
- The system provides direction for decision makers by categorizing patients in to different classes.
- To make the system easily maintainable, the MVC design architectural pattern is used to design and implement the proposed RBES.
- System and user acceptance testing are exercised.
- The user interface is flexible.
- The system is updatable when new rules are inserted in to the knowledge base.

7.2. RECOMMENDATION

The recommendation of this thesis is not only for further researchers but also for Adama and Assela hospital ART room officers and MOH. Based on the findings of the study the following recommendations are listed and interested researchers can improve the general ART service i.e. not only counseling service by integrating software engineering approaches with knowledge engineering and knowledge based systems:-

- In recording patient data, so many values are missed and the recorded ones are also misspelled. So, I recommend the concerned bodies to eliminate like such errors.
- Mobile based counseling expert system can be generated for consulting patients.
- Integrating the operations of many hospitals and or healthcare sectors in which the ART treatment is given essential for federal ministry of health for assessing results of ART treatment and make decision based on that.
- In this study an attempt is made to use MVC design architectural pattern for rule based expert system for standalone system. But it can be extended in to web based decision support system that can give consultation for patients by visualizing the clinical and socio-demographic categories.

- In the data preparation step, missing values are filled mostly by their frequent values if the missed records are few. But in some records where the numbers of missed values are higher than the most frequent values, the researcher get confused and put them as it is. So, if further technique of handling this gap is conducted, more accurate rules may be derived from the data.
- The cohesion and coupling of rules are not measured.
- The updating mechanism can be improved from the manual way to automatic way of updating the knowledge based system.

REFERENCES

[1] Joint United Nations Programme on HIV/AIDS (UNAIDS). (2008)

[2] [Online]. 17. https://msbfile03.usc.edu/digitalmeasures/doleary/intellcont/Brown-Oleary-es_tutor-1.htm 11/11/2013

[3] I. O. , 2 Abikoye, O. C., 2 Jimoh, R. G. and 2 Raji, K.S 1 Folorunso, "A Rule-Based Expert System for Mineral Identification," *Journal of Emerging Trends in Computing and Information Sciences*, vol. 3, no. ISSN 2079-8407, p. 205, 2012.

[4] "A RULE BASED EXPERT SYSTEM GENERATION FRAMEWORK USING AN OPEN SOURCE BUSINESS RULE ENGINE," Masters' thesis 2006.

[5] TEKLU URGESSA ABEBE, "APPLICATION OF DATA MINING TECHNIQUES ON ANTIRETROVIRALTHERAPY (ART) DATA: THE CASE OF ADAMA AND ASELLA HOSPITALS," ADDIS ABABA UNIVERSITY, ADDIS ABABA, masters' thesis 2010.

[6] TAGEL ABONEH, "KNOWLEDGE BASED SYSTEM FOR PRE-MEDCAL TRIAGE TREATMENT AT ADAMA UNIVERSITY ASELLA HOSPITAL," ADDIS ABABA UNIVERSITY, Masters thesis 2013.

[7] Wanzahun Godana, Mohammed Taha Amsalu Alagaw, "Factors associated with antiretroviral treatment adherence among adult patients in Wolaita Soddo Hospital," , 2013.

[8] Douglas D. Dankel Avelino J. González, *The Engineering of Knowledge-based Systems: Theory and Practice*.: Prentice Hall, 1993.

[9] Fabio Kurt Schneider Cecilia Vallejos de Schatz, "Intelligent and Expert Systems in Medicine," , 2011.

[10] Jadzia Cendrowska, "PRISM: An algorithm for inducing modular rules," *International Journal of Man-Machine Studies*, vol. 27, pp. 349--370, 1987.

[11] RC. Chakraborty. (2010, June) Artificial Intelligence : Course Content, Lecture Note, Slides, Text books, References. [Online]. http://www.myreaders.info/03-Knowledge_Representations.pdf

[12] William W., and Yoram Singer Cohen, "A simple, fast, and effective rule learner," in *National Conference on Artificial Intelligence*, 1999, pp. 335-342.

[13] 2012 country progress report on HIV/AIDS response, Ethiopia, country progress report 2012.

[14] Heikki Mannila, Padhraic Smyth D. J. Hand, *Principles of Data Mining*. Cambridge, 2001.

[15] Pedro Domingos, "Toward knowledge-rich data mining," Feb. 2007.

[16] Federal Democratic Republic of Ethiopia, "Country progress report on HIV/AIDS Response," country progress report 2012.

[17] 37. Ernest J. Friedman-Hill, Jess the Rule Engine for the JavaTM Platform, 2007.

[18] Ernest Friedman-Hill, Jess, the rule engine for the java platform, 2008.

[19] SHIRILEY GREGOR,.

[20] SHIRLEY GREGOR,., ch. 13, p. 12.

[21] Anne Håkansson, "A User Interface for the User-Centred Knowledge Model, t-UCK," , Sweden, 2004.

[22] ERNEST FRIEDMA HILL, *JESS in Action*. USA, 2003.

[23] Micheline Kamber, Jian Pei Jiawei Han, *Data Mining Concepts and Techniques*, 2nd ed., 2006.

[24] Dieter Fensel, Rudi Studer Jü rgen Angele, What could the knowledge engineer learn from the software engineer?, 2003.

[25] Aku Kolu, "MVC Frameworks in Web Development," 2012.

[26] C. S., and S. Rajeev Krishnamoorthy, "Artificial intelligence and expert systems for engineers.," vol. 11, 1996.

[27] Daniel T. Larose, *Discovering Knowledge in Data: An Introduction to Data Mining*. Canada, 2005.

[28] Jay Liebowitz, Ed., *The Handbook of Applied Expert Systems*. U.S.A, 1998.

[29] Kurnool, Andhra Pradesh MaddurNagar, "Utilization of Data Mining Techniques for Prediction and Diagnosis of Tuberculosis Disease Survivability," 2013.

[30] Andualem Tesfaye, Segewkal Hawaze Meseret Wube, "Antiretroviral Therapy Regimen Change Among HIV/AIDS Patients in Nekemt Hospital," , Ethiopia, Oromiya, 2013.

[31] Abdulkerim Mohammed, "Towards Integrating Data Mining with Knowledge Based System: The Case of Network Intrusion detection," ADDIS ABABA UNIVERSITY, ADDIS ABABA, Masters' thesis 2013.

[32] "Morbidity and Mortality," U.S department of health and human service, Weekly Report 2014.

[33] Jason Morris, Intro to Jess The Java Expert System Shell, 2003.

[34] Grzegorz J. Nalepa, "Methodologies and Technologies for Rule-Based Systems Design and

Implementation. Towards Hybrid Knowledge Engineering," vol. Volume 102, pp. pp 183-198 , 2008.

[35] MICHAEL NEGNEVITSKY, *Artificial Intelligence: A Guide to Intelligent Systems*, 2nd ed. England, 2005.

[36] Michael Negnevitsky, *Artificial Intelligence: A Guide to Intelligent Systems*, 3rd ed.: Pearson Education Limited, 2011, 2011.

[37] Federal HIV/AIDS Prevention and Control Office, "GUIDELINES FOR IMPLEMENTATION OF THE ANTIRETROVIRAL THERAPY PROGRAMME IN ETHIOPIA," Federal ministry of health, 2007.

[38] P. Hadjinian, R. Stadler, J. Verhees, A. Zanasi, P. Cabena, *Discovering data mining: From concept to implementation*.: Prentice-Hall , 1998.

[39] Prof. Mihaela OPREA PhD, "On the Use of Data-Mining Techniques in Knowledge-Based Systems," vol. 6, p. 21, 2006.

[40] Prof. Mihaela OPREA PhD., On the Use of Data-Mining Techniques in Knowledge-Based Systems, 2006.

[41] D., and D. Simic. Plakalovic, "Applying MVC and PAC patterns in mobile applications.," *arXiv preprint arXiv:1001.3489*, 2010.

[42] C.,Pesado, P. and GarcÃa-MartÃnez, R. Rancan, "TOWARD INTEGRATION OF KNOWLEDGE BASED SYSTEMS AND KNOWLEDGE DISCOVERY SYSTEMS ," vol. 7, p. 93, 2007.

[43] E. and Knight, K., McGraw Hill Rich, "Artificial Intelligence," 1992.

[44] Roger S. Pressman R. S., *Software engineering : a practitioner's approach*, 3rd ed. USA: McGraw-Hill, 1992.

[45] Priti Srinivas, and Rajendra Akerkar Sajja, "Knowledge-based systems for development.," *Advanced knowledge based systems: Model, Application & Research*, pp. 1-11, 2010.

[46] Alex H. B. Duffy , Robert Ian Whitfield, Iain M. Boyle Shaofeng Liu, Integration of decision support systems to improve decision support performance, 2009.

[47] Song, *Data Mining in Biomedical Imaging, Signaling, and System*, Rajendra Acharya U Sumeet Dua, Ed., 2009.

[48] K. P Tripathi, "A Review on Knowledge-based Expert System: Concept and Architecture.," *IJCA Special Issue on Artificial Intelligence Techniques-Novel Approaches & Practical Applications*, pp. 19-23, 2011.

[49] Yusuf, and Gülay Tezel Uzun, "RULE LEARNING WITH MACHINE LEARNING ALGORITHMS AND ARTIFICIAL NEURAL NETWORKS.," *Journal of Selcuk University Natural and Applied Science* , vol. 2, p. 54, 2012.

[50] WHO, "CONSOLIDATED GUIDELINES ON THE USE OF ANTIRETROVIRAL DRUGS FOR TREATING AND PREVENTING HIV INFECTION," 2013.

[51] Chaur G. Wu, "Modeling Rule-Based Systems with EMF," 2004.

[52] Yang, and Yanjing Luo Zhang, "An architecture and implement model for Model-View-Presenter pattern.," *3rd international conference on computer science and information technology*, vol. vol. 8, pp. pp. 532-536, 2010.

APPENDIX A

RULES CATEGORIZED BY AGE

```
(defglobal ?*ICount* = 0)
(defglobal ?*IICount* = 0)
(defglobal ?*IIICount* = 0)
(defglobal ?*IVCount* = 0)
(defglobal ?*VCount* = 0)

(defglobal ?*IInstance* = 0)
(defglobal ?*IIInstance* = 0)
(defglobal ?*IIIInstance* = 0)
(defglobal ?*IVInstance* = 0)
(defglobal ?*VInstance* = 0)

(set-reset-globals FALSE)

;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;
;; age rules

;status = W AND
;Religion_ID = Orthodox AND
;MID = Married: stage-IV (7561.0/41.0)

(defrule agerule1
    (status W)
    (religion orthodox)
    (marital-status married)
   =>
    (assert (age stage-IV))
    (bind ?*IVCount* (+ ?*IVCount* 1))
    (bind ?*IVInstance* (+
?*IVInstance* 7561))
    )
;status = W AND
;Religion_ID = Muslim AND
;HasFamily = TRUE: stage-IV
(1941.0/11.0)

(defrule agerule2
    (status W)
    (religion muslim)
    (has-family true)
   =>
    (assert (age stage-IV))
    (bind ?*IVCount* (+ ?*IVCount* 1))
         (bind ?*IVInstance* (+
?*IVInstance* 1941))

    )
;status = W AND
;Religion_ID = Orthodox AND
;MID = Widower: stage-IV (2466.0/9.0)

(defrule agerule3
    (status W)
    (religion orthodox)
    (marital-status widower)
   =>
    (assert (age stage-IV))
    (bind ?*IVCount* (+ ?*IVCount* 1))
         (bind ?*IVInstance* (+
?*IVInstance* 2466))

    )

;status = A AND
```

```
;Religion_ID = Orthodox: stage-IV
(2404.0/31.0)

(defrule agerule3
    (status A)
    (religion orthodox)
   =>
    (assert (age stage-IV))
    (bind ?*IVCount* (+ ?*IVCount* 1))
         (bind ?*IVInstance* (+
?*IVInstance* 2466))

    )

;status = W AND
;Religion_ID = Orthodox AND
;MID = Divorced: stage-IV (2433.0/11.0)

(defrule agerule4
    (status W)
    (religion orthodox)
    (marital-status divorced)

   =>
    (assert (age stage-IV))
    (bind ?*IVCount* (+ ?*IVCount* 1))
         (bind ?*IVInstance* (+
?*IVInstance* 2433))

    )

;status = W AND
;Religion_ID = Protestant: stage-IV
(1913.0/28.0)

(defrule agerule5
    (status W)
    (religion protestant)
   =>
    (assert (age stage-IV))
    (bind ?*IVCount* (+ ?*IVCount* 1))
         (bind ?*IVInstance* (+
?*IVInstance* 1913))

    )

;status = B: stage-IV (1218.0/26.0)

(defrule agerule6
    (status B)
   =>
    (assert (age stage-IV))
    (bind ?*IVCount* (+ ?*IVCount* 1))
         (bind ?*IVInstance* (+
?*IVInstance* 1218))

    )

;status = A AND
;MID = Others AND
;ARTStatus = IN: stage-III (14.0/5.0)
(defrule agerule7
    (status A)
    (marital-status others)
    (art-status IN)
   =>
    (assert (age stage-III))
    (bind ?*IIICount* (+ ?*IIICount*
1))
         (bind ?*IIIInstance* (+
?*IIIInstance* 14))

    )
```

```
;status = A AND
;MID = Others AND
;ARTStatus = OA AND
;OACD4 = class-II: stage-II (12.0/3.0)

(defrule agerule8
    (status A)
    (marital-status others)

    (art-status OA)
    (oacd4 class-II)
   =>
    (assert (age stage-II))
    (bind ?*IICount* (+ ?*IICount* 1))
         (bind ?*IIInstance* (+
?*IIInstance* 12))

    )

;status = A AND
;MID = Others AND
;ARTStatus = OA      AND
;OACD4 = class-I AND
;WHO = others: stage-III (11.0/3.0)


(defrule agerule9
    (status A)
    (marital-status others)
    (art-status OA)
    (oacd4 class-I)
    (WHO others)
   =>
    (assert (age stage-III))
    (bind ?*IIICount* (+ ?*IIICount*
1))
             (bind ?*IIIInstance* (+
?*IIIInstance* 11))

    )

;status = A AND
;MID = Others AND
;ARTStatus = OA AND
;OACD4 = class-III AND
;TerminatedYN = FALSE AND
;WHO = others AND
;Occupation = Unemployed: stage-II
(12.0/3.21)

(defrule agerule10
    (status A)
    (marital-status others)
    (art-status OA)
    (oacd4 class-III)
    (terminated false)
    (WHO others)
    (occupation unemployed)
   =>
    (assert (age stage-II))
    (bind ?*IICount* (+ ?*IICount* 1))
             (bind ?*IIInstance* (+
?*IIInstance* 12))

    )

;status = A AND
;ARTStatus = OA AND
;EducationID = Others: stage-II
(83.0/47.32)

(defrule agerule11
```

```
    (status A)
    (art-status OA)
    (education others)
   =>
    (assert (age stage-II))
    (bind ?*IICount* (+ ?*IICount* 1))
    (bind ?*IIInstance* (+
?*IIInstance* 83))

    )

;status = P AND
;OACD4 = class-I AND
;HasFamily = TRUE AND
;WHO = others AND
;Occupation = Farmer: stage-I
(7.36/2.91)

(defrule agerule12
    (status P)
    (oacd4 class-I)
    (has-family true)
    (who others)
    (occupation farmer)
   =>
    (assert (age stage-I))
    (bind ?*ICount* (+ ?*ICount* 1))
            (bind ?*IInstance* (+
?*IInstance* 7))

    )

;status = P AND
;OACD4 = class-II AND
;HasFamily = FALSE AND
;Occupation = Unemployed AND
;WHO = stage- 3: stage-I (11.73/4.91)

(defrule agerule13
    (status P)
    (oacd4 class-II)
    (has-family false)
    (occupation unemployed)
    (who stage-3)

   =>
    (assert (age stage-I))
    (bind ?*ICount* (+ ?*ICount* 1))
            (bind ?*IInstance* (+
?*IInstance* 11))

    )
```

APPENDIX B

QUESTIONER TO TEST AND VALIDATE THE PERFORMANCE THE PROPOSED SYSTEM

The following are list of questions prepared as Questionnaire to test the performance of the proposed system by users.

1. Is the user interface attractive?
   1. Poor 2. Fair 3. Good 4. V.good 5. Excellent
2. Is the system efficient in time?
   1. Poor 2. Fair 3. Good 4. V.good 5. Excellent
3. How much accurate the system is accurate in categorizing the patient in to correct category?
   1. Poor 2. Fair 3. Good 4. V.good 5. Excellent
4. Does the system incorporated sufficient knowledge to make in decision in a given category
   1. Poor 2. Fair 3. Good 4. V.good 5. Excellent
5. Does the system have contribution in the domain area?
   1. Poor 2. Fair 3. Good 4. V.good 5. Excellent
6. Simplicity to use and interact with the System
   1. Poor 2. Fair 3. Good 4. V.good 5. Excellent