AN ONLINE BELIEF RULE-BASED GROUP CLINICAL DECISION SUPPORT SYSTEM

A thesis submitted to The University of Manchester for the degree of

Doctor of Philosophy

in the Faculty of Humanities

2011

GUILAN KONG

MANCHESTER BUSINESS SCHOOL

TABLE OF CONTENTS

Word count: 54,079

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS

AHP	analytic hierarchy process
AMI	acute myocardial infarction
ANN	artificial neural network
AUC	area under the ROC curve
BBN	Bayesian belief network
BRB	belief rule base
CBR	case-based reasoning
CCP	cardiac chest pain
CDSS	clinical decision support system
CDSG	clinical decision support guideline
COM	component object model
DBMS	data-base management systems
DLL	dynamic link library
DSS	decision support system
ECG	electrocardiogram
ED	Emergency Department
ER	evidential reasoning
ES	expert system
EVaMACS	Early Vascular Markers of Acute Coronary Syndromes
GI	gastrointestinal
H-FABP	heart type fatty acid binding protein
ICU	intensive care unit
IDE	integrated development environment

IDS	intelligent decision system
IT	information technology
LBBB	left bundle branch block
MADA	multiple attribute decision analysis
MDM	medical decision making
MRI	Manchester Royal Infirmary
NHS	National Health Service
NSAI	non-steroidal anti-inflammatory
QMR	quick medical reference
RIMER	Rule-base Inference Methodology using the Evidential Reasoning approach
ROC	receiver operating characteristics
SBP	systolic blood pressure
SDLC	system development life cycle
SQL	structured query language
STEMI	ST segment elevation myocardial infarction
TnI	Troponin I
TNR	true negative rate
TPR	true positive rate
UMLS	unified medical language system
XML	eXtensible Markup Language

ABSTRACT

The University of Manchester
Guilan Kong
PhD in Business Administration
Thesis title: "An online belief rule-based group clinical decision support system"
April 2011

Around ten percent of patients admitted to National Health Service (NHS) hospitals have experienced a patient safety incident, and an important reason for the high rate of patient safety incidents is medical errors. Research shows that appropriate increase in the use of clinical decision support systems (CDSSs) could help to reduce medical errors and result in substantial improvement in patient safety. However several barriers continue to impede the effective implementation of CDSSs in clinical settings, among which representation of and reasoning about medical knowledge particularly under uncertainty are areas that require refined methodologies and techniques. Particularly, the knowledge base in a CDSS needs to be updated automatically based on accumulated clinical cases to provide evidence-based clinical decision support.

In the research, we employed the recently developed belief Rule-base Inference Methodology using the Evidential Reasoning approach (RIMER) for design and development of an online belief rule-based group CDSS prototype. In the system, belief rule base (BRB) was used to model uncertain clinical domain knowledge, the evidential reasoning (ER) approach was employed to build inference engine, a BRB training module was developed for learning the BRB through accumulated clinical cases, and an online discussion forum together with an ER-based group preferences aggregation tool were developed for providing online clinical group decision support.

We used a set of simulated patients in cardiac chest pain provided by our research collaborators in Manchester Royal Infirmary to validate the developed online belief rule-based CDSS prototype. The results show that the prototype can provide reliable diagnosis recommendations and the diagnostic performance of the system can be improved significantly after training BRB using accumulated clinical cases.

DECLARATION

No portion of the work referred to in the thesis has been submitted in support of an application for another degree or qualification of this or any other university or other institute of learning.

COPYRIGHT STATEMENT

i. The author of this thesis (including any appendices and/or schedules to this thesis) owns certain copyright or related rights in it (the "Copyright") and s/he has given The University of Manchester certain rights to use such Copyright, including for administrative purposes.

ii. Copies of this thesis, either in full or in extracts and whether in hard or electronic copy, may be made **only** in accordance with the Copyright, Designs and Patents Act 1988 (as amended) and regulations issued under it or, where appropriate, in accordance with licensing agreements which the University has from time to time. This page must form part of any such copies made.

iii. The ownership of certain Copyright, patents, designs, trade marks and other intellectual property (the "Intellectual Property") and any reproductions of copyright works in the thesis, for example graphs and tables ("Reproductions"), which may be described in this thesis, may not be owned by the author and may be owned by third parties. Such Intellectual Property and Reproductions cannot and must not be made available for use without the prior written permission of the owner(s) of the relevant Intellectual Property and/or Reproductions.

iv. Further information on the conditions under which disclosure, publication and commercialisation of this thesis, the Copyright and any Intellectual Property and/or Reproductions described in it may take place is available in the University IP Policy (see http://www.campus.manchester.ac.uk/medialibrary/policies/intellectual-property.pdf), in any relevant Thesis restriction declarations deposited in the University Library, The University Library's regulations (see http://www.manchester.ac.uk/library/aboutus/regulations) and in The University's policy on presentation of Theses.

ACKNOWLEDGEMENT

First of all, I would like to express my deep gratitude to my supervisors, Professor Jian-Bo Yang and Dr. Dong-Ling Xu. Without their invaluable support, guidance, and encouragement, I could not have been able to carry on this work which is extremely demanding for me.

I am very grateful to Dr. Richard Body, Professor Kevin Mackway-Jones, and Professor Simon Carley in Manchester Royal Infirmary. Without their priceless support and assistance, the work presented in this thesis has not been possible.

The research presented in the thesis was sponsored by Manchester Business School. I would like to express my sincere thanks to Manchester Business School for the financial support.

I would like to thank my parents for their selfless love, support and helping me to look after my baby daughter, Langqing, from the day I left China for my PhD study till I went back to pick her up nine months later. I would also like to thank my husband, Ziliang, for always being there with support, guidance, encouragement and understanding. The memories of those weekends during my PhD study when Ziliang and Langqing accompanied me in my office or Langqing and I accompanied Ziliang in his office have embedded in my brain for ever.

Finally, I would like to express my gratitude to secretaries in Postgraduate Research Programmes office at Manchester Business School for providing timely help when I need them. Thanks also go to Dr Yu-wang Chen and Mr. Dawei Tang in decision science research group for their friendship and assistance during the period of my PhD study.

PREFACE

GUILAN KONG

Qualifications

BEng	National University of Defense Technology, China	1997
MEng	National University of Defense Technology, China	2000

Grants and Awards

1. Second Class National Scientific Advancement Award, awarded by the Commission for Science, Technology and Industry for National Defence of China, September 2001.
2. Third Class National Scientific Advancement Award, awarded by the Commission for Science, Technology and Industry for National Defence of China, November 2006.
3. Manchester Business School Scholarships and Bursaries (£21,000 per annum, September 2006 – September 2009), awarded by Manchester Business School.

Publications from this research

Journal papers (2):

1. **Guilan Kong**, Dong-Ling Xu, Jian-Bo Yang. "Clinical decision support systems: a review on knowledge representation and inference under uncertainties", *International Journal of Computational Intelligence Systems,* 2008, 1(2): 159 - 167.
2. **Guilan Kong**, Dong-Ling Xu, Xinbao Liu, Jian-Bo Yang. "Applying a new rule-base inference methodology into guideline-based clinical decision support system", *Expert Systems*. 2009, 26(5): 391-408.

Conference presentations (6):

1. **Guilan Kong**, Dong-Ling Xu, Jian-Bo Yang. "A survey of clinical decision support systems", Manchester Business School Annual Conference, Manchester, UK, 3 May, 2007.
2. **Guilan Kong**, Dong-Ling Xu, Jian-Bo Yang. "Clinical decision support systems", the 13th International Conference on Automation and Computing, Staffordshire, UK, 15 September, 2007.
3. **Guilan Kong,** Dong-Ling Xu, Jian-Bo Yang. "A prototype online intelligent group clinical decision support system with belief rule-base inference methodology", the 14th International Conference on Automation and Computing, Brunel, UK, 6 September, 2008.
4. **Guilan Kong**, Dong-Ling Xu, Jian-Bo Yang. "Applying a new belief rule-base inference methodology to development of clinical decision support system", FLINS Conference, Madrid, Spain, 21-24 September, 2008.

5. **Guilan Kong,** Dong-Ling Xu, Jian-Bo Yang. “An online intelligent group clinical decision support system”, 2009 Doctoral School on Decision Theoretic AI, Cork, Ireland, 18-22 April, 2009.
6. **Guilan Kong,** Dong-Ling Xu, Jian-Bo Yang. “An evidence-adaptive belief rule-based decision support system for clinical risk assessment in emergency care”, invited by 23rd European Conference on Operational Research, Bonn, Germany, 5-8 July, 2009.

Book chapters (1):

1. Guilan Kong, Dong-Ling Xu and Jian-Bo Yang. Appling a new belief rule-base inference methodology into clinical decision making. In Da Ruan; Javier Montero; Jie Lu; Luis Martínez; Pierre D'hondt and Etienne E Kerre (eds) *Computational Intelligence in Decision and Control,* pp. 127-148. Singapore: World Scientific Publisher. 2008 (ISBN: 978-981-279-946-3).

Chapter 1

Introduction

1.1. Background

Patient safety incidents or adverse events, which are unintended or unexpected incidents that could have or did lead to harm for one or more patients receiving National Health Service (NHS)-funded health care, represent a serious public health problem and pose a threat to patient safety (Thomas and Brennan, 2001). Research shows that around 10% of patients admitted to NHS hospitals have experienced a patient safety incident, and that up to half of these incidents could have been prevented (Department of Health, 2004). Patient safety incidents cause great harm to not only patients and their families, but also involved clinicians and host hospitals. For example, it is estimated that patient safety incidents cost NHS £2 billion a year in addition to hospital stays, without taking account of human or wider economic costs (Department of Health, 2004).

In clinical governance (Department of Health, 1998), which is a framework through which NHS organizations are accountable for continuously improving the quality of their services and safeguarding high standards of care by creating an environment favourable for the excellence in clinical care to flourish, the reduction of medical errors and the improvement of patient safety have become major priorities since 2000 (Department of Health Expert Group, 2000).

Research shows that an important reason for high rate of patient safety incidents is medical errors that are mostly caused by human factors (Reason, 2001). Take

Emergency Department (ED) for example, causes of medical errors that can happen in ED include clinicians' inexperience or lack of training, interrupted clinical management of patients due to clinicians' shift schedule, clinicians' languished vigilance and alertness of patients' abnormalities, and clinicians' stress which may be caused by patients' various conditions or the working environment in ED (Driscoll et al., 2001). In clinical governance, a very important action is to examine the potential of information technology (IT) to reduce the risks to patients and improve the quality of health care (Department of Health, 2001).

Appropriate increase in the use of IT in health care has proved to help bring substantial improvement in patient safety (Bates et al., 2001). Particularly the introduction of clinical decision support, and appropriate communications between clinical decision support system (CDSS) and the deployed computerised clinical systems have simplified the health care process and substantially facilitated clinical practice and reduced medical errors (Sim et al., 2001, Kawamoto et al., 2005). There are numerous examples of CDSSs in health care which, have successfully improved the quality of health care (de Dombal et al., 1972, Jonsbu et al., 1993, Lin et al., 2006).

Significant research progresses, both theoretical and practical, have been achieved since the idea of computer-based CDSSs emerged. Nonetheless few CDSSs in the literature have been widely applied in practice. The causes for the low popularity of CDSSs include, among others, uncertainties in clinical signs, clinical symptoms and clinical domain knowledge, the complexity of involved inference mechanism, difficulties with domain selection and knowledge base construction and maintenance, and problems with system validation and evaluation (Miller and Geissbuhler, 1999).

A recent survey conducted by Sittig and colleagues (2008) identifies several top challenges in clinical decision support, among which 'prioritised and filtered recommendations to the user' is the one for researchers in decision science area to overcome. This challenge is closely related to inference methodologies used in CDSSs. It requires that the inference mechanism in a CDSS should have the capability of handling different clinical uncertainties and generating possible recommendations with corresponding priorities attached to them and can filter irrational recommendations.

Another challenge in CDSSs identified by researchers is to provide 'evidence-adaptive' CDSSs to better facilitate evidence-based medicine (Sim et al., 2001). A CDSS is evidence-based rather than evidence-adaptive if its clinical knowledge base is derived from scientific evidence, but no mechanisms are in place to incorporate new clinical evidence. On the contrary, a CDSS is evidence-adaptive if its knowledge base is based on current evidence and its recommendations are routinely updated to incorporate new clinical evidence (Sim et al., 2001). As a result, an evidence-adaptive CDSS requires its knowledge base to be adaptive to up-to-date clinical evidence which usually can be obtained from clinical literature and clinical practice. However it is difficult for a CDSS to keep up with clinical literature since the contents of clinical literature are textual and thus not machine-interpretable by present-day CDSSs. Nonetheless, it is not insurmountable for a CDSS to adapt itself to accumulated clinical cases in local clinical practice. The knowledge base, which is usually derived from the best clinical literature or expert domain knowledge, can be updated automatically and routinely based on evidence accumulated in clinical practice in an evidence-adaptive CDSS. This requires that the knowledge representation scheme employed in the evidence-adaptive CDSS should have a corresponding mechanism

which can help the clinical rules embedded in the knowledge base to learn from accumulated clinical cases routinely.

In short, CDSSs is promising in helping facilitate evidence-based medicine and reducing patient safety incidents. But there are some challenges in CDSSs research area that need to be tackled, and these challenges are closely related to knowledge representation scheme and inference methodology. Therefore representation of and reasoning with uncertain medical knowledge are areas that require refined methodologies and techniques (Musen et al., 2006, Lin et al., 2006). Moreover, although some researchers have proposed the idea of developing a CDSS which can provide group decision support (Hatcher, 1990, Rao et al., 1996), few CDSSs in the literature have the capability of providing group clinical decision support.

1.2. Research Questions

To surmount the challenges in CDSSs research as identified from the literature, a recently developed new Rule-base Inference Methodology using the Evidential Reasoning approach (RIMER) (Yang et al., 2006) is employed for developing a CDSS in the research. In the CDSS, belief rule base (BRB) is employed to model specific clinical domain knowledge such as clinical rules for risk assessment of cardiac chest pain (CCP); the evidential reasoning (ER) approach is used as a mechanism to do clinical inference and group clinical decisions aggregation; and an optimization model is used to train or fine-tune BRB through clinical cases accumulated in clinical practice. The research is aimed at answering three main questions:

(1) Is it feasible to employ RIMER for developing a CDSS?

(1-1) What are the system features of the existing CDSSs?

A very important motivation of the research is the lack of a CDSS in the literature which can firstly, handle uncertainties properly in both clinical signs and symptoms and clinical domain knowledge; secondly, provide group or collaborative clinical decision support; and thirdly, have learning capability to automatically update the knowledge base so that the system can be adaptive to clinical practice. To bridge the gap identified in CDSSs literature, 'what are the system features of the existing CDSSs' is the first and most important question we need to answer in the research. System features include domain knowledge representation schemes, clinical inference mechanism, and group clinical decision supporting capability implemented in one CDSS.

(1-2) Is it feasible to employ BRB to model clinical domain knowledge for developing a CDSS?

In the research, we proposed to apply the RIMER methodology for developing a CDSS. It is original to use BRB to model domain knowledge in clinical areas, although BRB has been successfully employed in modelling domain knowledge in areas such as pipeline leak detection (Xu et al., 2007). The feasibility of employing BRB to model clinical domain knowledge should be studied prior to the design and development of a belief rule-based CDSS.

(1-3) Is it reliable to apply the ER approach to build inference engine in the belief rule-based CDSS?

The ER approach has been successfully applied in inference with the BRB model (Yang et al., 2006, Xu et al., 2007). However, it is novel to use the ER approach to do

clinical diagnosis or inference. The reliability of using ER to do clinical inference in a belief rule-based CDSS should be investigated in the research.

(2) How to facilitate online group clinical decision making and arrive at a group combined clinical recommendation in a belief rule-based CDSS?

Groups are often perceived as better equipped than individuals to make difficult decisions (Rangel, 2009). In CDSSs research, the idea of providing group or collaborative decision support for doctors in practice is not new, but few CDSSs in the literature have the capability of supporting group clinical decision making. In the research, we should try to investigate how to facilitate online group clinical decision making and arrive at a group combined clinical recommendation in a belief rule-based CDSS.

(3) How to train belief rule-based CDSS and make its knowledge base to be adaptive to clinical practice?

It is essential for an evidence-adaptive CDSS to have intelligent learning ability so that its knowledge base can be adaptive to clinical practice. In the RIMER methodology, some optimization models can be built to train BRB in belief rule-based systems (Yang et al., 2007). 'How to train belief rule-based CDSS and make its knowledge base to be adaptive to clinical practice' is a necessary research question we need to investigate in the research.

1.3. Research Objectives

Based on the research questions above, the measurable objectives of the research are as follows.

(1) To investigate the existent CDSSs, and to identify system features of the

existing CDSSs.

(2) To acquire target clinical domain knowledge.

Although clinical domain knowledge can be found in medical textbooks, medical journals, clinical practice guidelines, and so on, it is impossible for a non-medicine relevant researcher to fully acquire or understand the knowledge all by himself or herself. Thus it is crucial to acquire and elicit domain knowledge from proper expert clinicians and get in-depth understanding of the domain knowledge through field study.

(3) To investigate the feasibility of employing BRB to model clinical domain knowledge and using the ER approach to do clinical inference in a CDSS.

Concerns would inevitably arise over whether it is feasible to use BRB for modelling clinical domain knowledge and use the ER approach to build inference engine in a CDSS, since employing the RIMER methodology for developing a CDSS is relatively novel. The knowledge modelling methodology most commonly used in the existent CDSSs since the early CDSS MYCIN is traditional 'IF-THEN' rules because of their naturalness and transparency (Spooner, 1999, Spooner, 2007). In the feasibility investigation of this research, inference with BRB using the ER approach is compared to traditional 'IF-THEN' rule-based inference by examining real or simulated cases in selected clinical area.

(4) To design and develop an online belief rule-based group CDSS prototype.

A belief rule-based CDSS should be designed and developed after the preliminary feasibility study for the target clinical areas. In system design, the system architecture can be designed as a web-based three-layer architecture, where system users from the

client layer can access the CDSS through internet, and core system components reside in the server layer, and most of data used in the system can be stored in the back-end data layer. Knowledge base should be constructed using the BRB model, while the kernel algorithm of inference engine and group decision supporting module should be the ER approach, and the knowledge training module can train BRB via learning from accumulated clinical cases.

In system development, Microsoft .NET technology is considered.

(5) To validate the online intelligent CDSS prototype using clinical cases in target clinical areas.

Ideally, after system design and development, the CDSS prototype should be validated by real patients' data collected in clinical practice. If the real data of patients can not be obtained, simulated patients' data can be used for the purpose. A necessary requirement for the simulated data is that it should be close to real patients' data.

1.4. Research Approach

A research methodology consists of the combination of the process, methods, and tools which are used in conducting research in a research domain (Nunamaker and Chen, 1990). A research domain is the subject matter under study in a research project. This research focuses on design, development and validation of an online belief rule-based CDSS prototype. A multi-methodology approach (Nunamaker and Chen, 1990) is employed in the study in that one methodology only is far from sufficient for the current complex research.

Firstly, modelling (Turban and Aronson, 2001) is a key method used in the research. Similar to general DSS, a CDSS should include several models which represent different parts of the clinical decision making problem, and these models are knowledge-based model, inference model, and knowledge base optimization model.

Secondly, field study is employed for study of clinical work flow in NHS hospitals and domain knowledge acquisition in knowledge modelling. By doing field study, manual methods including interviewing and observing (Turban and Aronson, 2001) are conducted in the research for knowledge elicitation.

Thirdly, a system development methodology - prototyping (Turban and Aronson, 2001) is used for CDSS prototype development,

Fourthly, statistical techniques including the receiver operating characteristics (ROC) curve analysis (Metz, 1978) and comparison of the area under the ROC curve (AUC) (Vergara et al., 2008) are used for analyzing the prototype's diagnostic performance in validating the CDSS prototype. Brief introduction to the statistical analysis used in the research will be discussed in Chapter 6.

To sum up, four main research methods including modelling, prototyping, field study, and statistical analysis are used complementarily in the research. Detailed description of the research methods used can be found in Chapter 3.

1.5. Significance

The following research gaps regarding CDSSs have been identified by reviewing the related literature.

(1) **There is a gap in knowledge modelling methodologies used in existent CDSSs** (Lin et al., 2006, Musen et al., 2006).

It is observed that not only clinical signs and clinical symptoms, but also clinical domain knowledge described by individuals is inherent with uncertainties in nature (Szolovits, 1995). To handle clinical uncertainties, various domain knowledge modelling methodologies such as artificial neural networks (ANNs), Bayesian belief networks (BBNs), cases have been applied in existent CDSSs. However, the existent knowledge modelling methodologies have their own inherent drawbacks in representing uncertainties in a transparent and easily understandable way (Lin et al., 2006, Musen et al., 2006)..

(2) **There is insufficient capability of handling uncertainties in inference mechanisms used in existent CDSSs** (Lin et al., 2006, Musen et al., 2006).

Inference mechanisms used in CDSSs are closely related to their corresponding knowledge representation schemes. Through decades of development, researchers have proposed various reasoning methods for handling uncertainties in clinical decision making. Those methods include combining fuzzy logic or certainty factors or Bayesian probabilities with traditional 'IF-THEN' rules, ANNs, etc. However, most inference mechanisms used in the existent CDSSs can not handle clinical uncertainties in a satisfactory way. For example, reasoning in ANN-based CDSS is hard for system users to understand, and the knowledge base is restricted to its learnt zones. Another example is Bayesian rule-based reasoning, which takes advantages of conditional independence, but all conditional probabilities for modelling domain knowledge in a Bayesian rule-based CDSS are hard to acquire or estimate. Therefore an inference mechanism which can well handle clinical uncertainties and process clinical inference

in a transparent way is needed in CDSSs research area (Lin et al., 2006, Musen et al., 2006).

(3) **There is insufficient intellectual capability of updating knowledge base in the existent CDSSs to make them to be adaptive to clinical practice to support evidence-based medicine** (Sim et al., 2001).

In the literature, some non-knowledge-based CDSSs have learning ability to automatically update their knowledge bases. For example, an ANN-based CDSS can train all parameters of its network using a large historical dataset before the system can be put into real use. However, non-knowledge-based CDSSs have drawbacks in clinical reasoning because their knowledge base is restricted to the training data. Knowledge learning functionality is rarely considered in system design and development in knowledge-based CDSSs, which acquire domain knowledge manually to construct knowledge bases. Therefore the intelligence of automatically updating knowledge base according to daily clinical practice in existent CDSSs is not sufficient (Sim et al., 2001).

(4) **There is a lack of a CDSS which can provide group clinical decision support together with individual clinical decision support** (Rao et al., 1994, Rao et al., 1996).

Although the idea of embedding group decision making with individual CDSSs has been proposed by researchers (Hatcher, 1990, Hatcher, 1994, Rao et al., 1994), few CDSSs in the literature have the capability of providing group or collaborative clinical decision support.

The research is, therefore, of significance as it addresses the above gaps effectively.

1.6. Contribution

In trying to fill the research gaps outlined above, the current research contributes to both CDSS research and practical domain application. Major contributions of the research are listed as follows:

(1) **From a CDSS research perspective:**

- The research develops a new CDSS framework which integrates knowledge-based CDSS with automatic knowledge learning functionality and online group decision supporting functionality.

- The research proposes and uses relational database to uniquely store and manage BRB model, and this makes physical knowledge base construction flexible and portable, and it makes it possible to share the knowledge between different clinical systems free of technology barriers thanks to mature relational database technologies.

(2) **From a practical domain application perspective**:

- The research develops a target clinical domain BRB for modelling domain specific knowledge under uncertainty. The BRB can be used not only for generating automatic diagnosis recommendations but also for clinicians' future domain knowledge reference in practice.

- The research develops an ER based inference engine to do inference with input uncertain clinical data and back-end uncertain domain knowledge in the BRB. The inference engine does inference with different clinical uncertainties in a rational way, and can generate prioritised and informative diagnosis

recommendations.

- The research develops an ER based group clinical decision support module. The module provides not only a group diagnosis preferences aggregation mechanism but also a discussion forum for a group of consultants to hold online meetings and discussions or consultations.

- The research develops a BRB training module that can help to update the embedded clinical rules automatically and routinely and help to keep the knowledge base to be adaptive to clinical practice.

- The research implements guideline-based user interfaces which not only facilitates clinicians complying with the practice guidelines, but also makes the integration of CDSS into clinical work flow easily implemented.

1.7. Outline of Contents of the Thesis

The thesis comprises 7 chapters as shown in Figure 1-1.

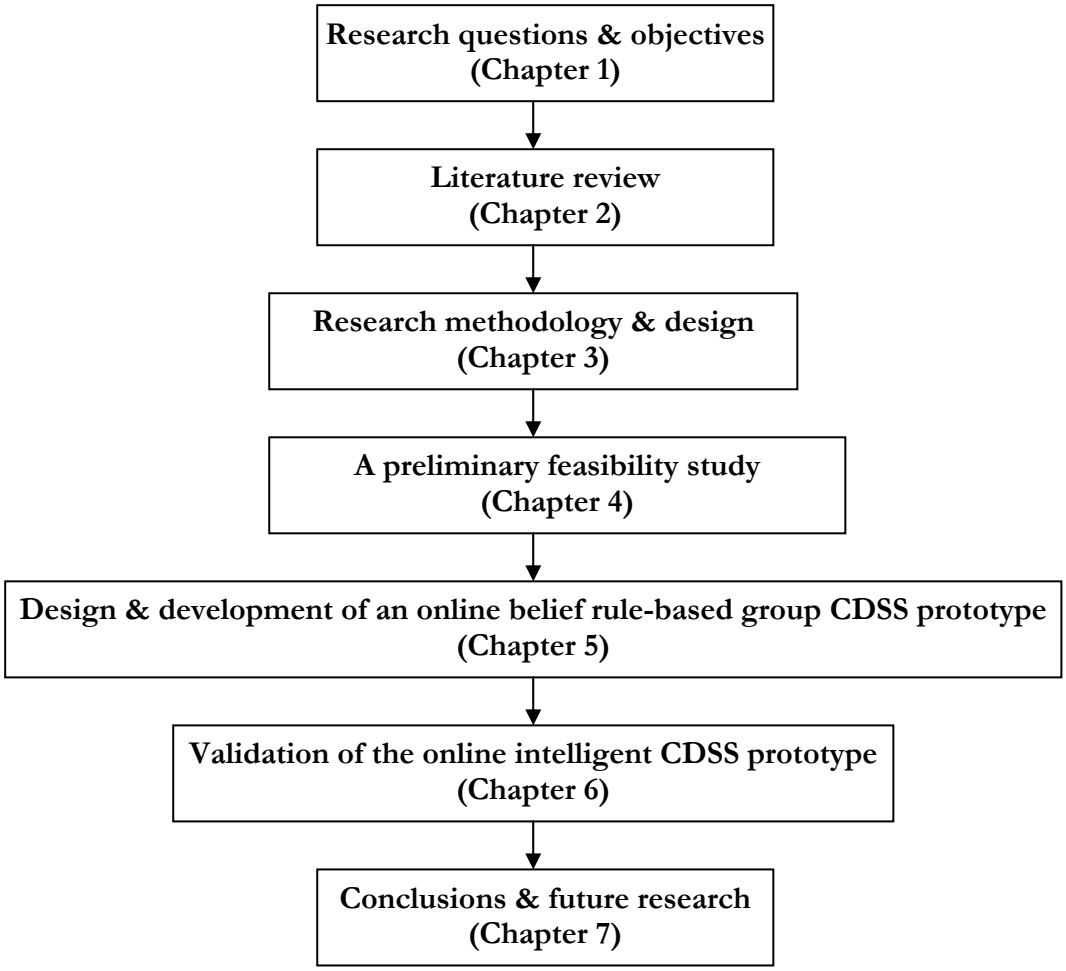


Figure 1-1: Structure of the Thesis

Chapter 1 presents an overview of the research.

Chapter 2 reviews the state of the art of CDSSs.

Chapter 3 discusses methodologies employed in the research and the research design.

Chapter 4 presents a preliminary feasibility study of employing RIMER for development of a CDSS, in which a comparison between a traditional rule-based system and a belief rule-based system in drawing clinical conclusions is conducted.

Chapter 5 describes the design and development of an online intelligent group CDSS prototype which provides individual clinical decision support, group consultation

support, and automatic knowledge updating based on daily clinical cases accumulated in clinical practice.

Chapter 6 discusses the validation of the developed prototype system using simulated clinical cases in target clinical areas.

Chapter 7 summarises the whole thesis and discusses possible future research.

Chapter 2

Literature Review

2.1. Introduction

CDSSs are computer systems designed to impact clinical decision making about individual patients at the point of time that these decisions are made (Berner and La Lande, 2007). With the increased focus on the prevention of medical errors, CDSSs have been proposed as a key element of systems' approaches to improve patient safety (Bates et al., 1998, Kohn et al., 2000). If designed, developed, and implemented properly, CDSSs have potential to improve the quality of health care service and change the way medicine has been practiced (Sim et al., 2001, Kawamoto et al., 2005).

From early generation of CDSSs such as AAPhelp - the Leeds abdominal pain diagnosis system (de Dombal et al., 1972), MYCIN (Shortliffe, 1976), and Quick Medical Reference (QMR) (Miller and Masarie, 1989), to the evolution of modern clinical decision support tools such as EON (Tu and Musen, 1999), PROforma (Fox et al., 1998), and GLIF (Peleg et al., 2000) based on evidence-based clinical guidelines, CDSSs have a history of almost 40 years. Significant research progresses, both theoretical and practical, have been made since the idea of computer-based CDSSs emerged. However, CDSSs are not yet common in patient care settings, and several challenges such as representation of and reasoning about medical knowledge under uncertainty, and integration of CDSSs into clinical workflow continue to impede the effective implementation of CDSSs in clinical settings (Musen et al., 2006).

This chapter provides a review of the literature which is essential in order to ascertain the research work that has been carried out in CDSSs and reveal the topics in which further research can be fruitfully made to advance both the literature and the practice of CDSSs. The chapter is organised as follows. Section 2.2 outlines typical definitions of CDSSs. Section 2.3 presents a review of state-of-the-art of CDSSs, where Section 2.3.1 provides a discussion of different types of CDSSs, Section 2.3.2 discusses sources of different types of clinical uncertainties, Section 2.3.3 and Section 2.3.4 present a review of knowledge-based and non-knowledge-based CDSSs respectively, group CDSSs are discussed in Section 2.3.5, and review of CDSSs validation study is presented in Section 2.3.6. Finally, Section 2.4 concludes the review and identifies the research gaps that this research aims to bridge.

2.2. Definition of CDSSs

There are different types of computerised systems in health care that can provide potential clinical decision support. While traditional CDSSs are defined as systems providing intelligent and automatic diagnostic inference or reasoning to generate patient specific assessment or recommendations to aid clinicians, some medical systems having no reasoning capability can also provide clinical decision support. For example, BestBETs (http://www.bestbets.org/) is a web-based medical system developed and maintained in Manchester Royal Infirmary (MRI) to provide evidence-based clinical decision support, and the web-based system has a large volume of best evidence topics provided by clinicians all over the world, but the system possesses no reasoning capability. This review is based on CDSSs that have intelligent diagnosis or assessment capability. Typical definitions of CDSSs in the literature are given below.

Musen (1997) defines a CDSS as any piece of software that takes information about a clinical situation as inputs and that produces inferences as outputs that can assist practitioners in their decision making and that would be judged as intelligent by the program's users.

Miller and Geissbuhler (1999) defines a CDSS providing diagnostic decision support as a computer-based algorithm that assists a clinician with one or more component steps of the diagnostic process.

Sim et al. (2001) defines CDSSs as 'software that is designed to be a direct aid to clinical decision-making, in which the characteristics of an individual patient are matched to a computerised clinical knowledge base and patient specific assessments or recommendations are then presented to the clinician or the patient for a decision'.

All above cited definitions of CDSSs given by masters in CDSSs research specify three similar key elements of a CDSS, namely (a) information about a clinical situation or an individual patient that acts as the system's inputs, (b) an intelligent diagnosis or assessment mechanism which may contain one or more components, and (c) patient specific assessments or recommendations that are the system's outputs.

2.3. State-of-the-Art of CDSSs

2.3.1. Types of CDSSs

CDSSs can be classified into different types according to different criterion.

Berlin and his colleagues (2006) propose to classify CDSSs according to their technical, workflow, and contextual characteristics.

- Based on internet technology, CDSSs can be classified as stand-alone and web-based systems.

If we classify CDSSs based on their technical characteristics, e.g. internet technology, some early CDSSs such as AAPhelp - Leeds abdominal pain diagnosis system (de Dombal et al., 1972) and MYCIN (Shortliffe, 1976) are stand-alone systems, some recently developed CDSSs such as (Huang and Chen, 2007, Fearn et al., 2007) are web-based or online systems.

- Based on the working environment, CDSSs can be classified as ED CDSS, ICU CDSS, laboratory CDSS, and bed-ward CDSS, etc.

If we classify CDSSs based on the roles that they play in the process of clinical work flow or their clinical working environment, some CDSSs target ED (Roukema et al., 2008, Graham et al., 2008), some CDSSs target intensive care unit (ICU) (Gago et al., 2007, Mack et al., 2009, Kumar et al., 2009), some CDSSs target laboratories (Grams, 1993), some CDSSs target bed-ward (Thilo et al., 2009), and some target medicine prescription (Lin et al., 2009).

- Based on target clinical domain, CDSS can be classified as in different clinical areas.

If we classify CDSSs based on different clinical domains that they have impact on, some CDSSs are for cancer pain management (Thilo et al., 2009), some CDSSs are for acute abdominal pain (de Dombal et al., 1972), some CDSSs are for gynecological diseases (Mangalampalli et al., 2006), and some CDSSs are for heart disease (Yan et al., 2006), and so on.

Metzger and her colleagues (2002) describe CDSS using different dimensions. According to their research, CDSS differ among themselves in the *timing* at which they provide support (before, during, or after the clinical decision is made) and how active or passive the support is, i.e. whether the CDSS actively provides alerts or passively responds to physician input or patient-specific information.

More generally, researchers classify CDSSs based on the way their knowledge bases are constructed. Some are knowledge-based CDSSs in which domain knowledge is acquired from domain experts or medical literature, and the others are non-knowledge-based CDSSs which learn domain knowledge through large historical data (Berner and La Lande, 2007, Spooner, 2007).

- Knowledge-based CDSSs

Before elaborating on the framework of knowledge-based CDSSs, here we provide a brief discussion of general knowledge-based decision support systems (DSSs) first. Klein and Methlie (1995) defined that a knowledge-based system is a computer program that employs knowledge and reasoning to solve problems, and an expert system (ES) is such a knowledge-based system, where knowledge and inference procedures are modelled after human experts. For a traditional DSS, its aim is to provide information in a given application domain by means of analytical decision models in order to support a decision maker in making decisions. The framework of knowledge-based DSSs is resulting from integrating DSSs technologies and ESs technologies. It is based on the paradigm of decision support, but also enables us to incorporate specialized knowledge and expertise into the system, and it can take advantages of numeric computations in traditional DSSs and reasoning functions in

ESs, and the system architecture consists of components from DSSs and ESs (Klein and Methlie, 1995).

In CDSSs research area, many of today's knowledge-based CDSSs arose out of earlier ESs research. What we usually mean by a CDSS is a program that supports a reasoning task carried out behind the scenes and based on clinical data. For example, a program that accepts clinical information about a patient with some clinical signs and symptoms and generates a list of possible diagnoses is what we usually recognize as a diagnostic decision support system which is a particular type of CDSS. The intent of these CDSSs is no longer to simulate an expert's decision making, but to assist a clinician in his or her own decision making. The system was expected to provide information for the user, rather than to come up with "the answer" as was the goal of earlier ESs. The knowledge-based systems cannot simply "learn" how to do the reasoning task from modelling human experts, and the human expert must put the knowledge into the system explicitly and directly (Berner and La Lande, 2007, Spooner, 2007).

Based on the idea of knowledge-based CDSSs as proposed in the literature, a general model of knowledge-based CDSSs in the literature can be depicted as in Figure 2-1. We adopt the general knowledge-based CDSS model in our review.

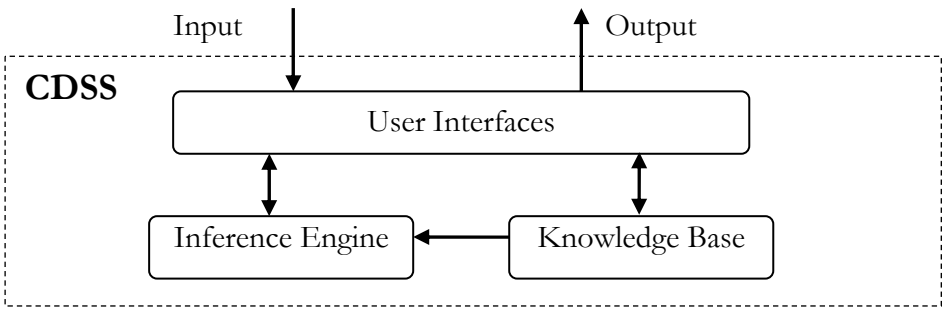


Figure 2-1: A General Model of Knowledge-Based CDSSs (Spooner, 2007)

From the general system structure as depicted in Figure 2-1, there are generally three essential system components in a knowledge-based CDSS. The first component is a knowledge base which includes clinical domain knowledge that is often, but not always, represented in the form of traditional 'IF-THEN' rules. The second one is an inference engine which contains algorithms or formulas for combining or matching clinical rules in the knowledge base to input clinical data. The third part is a user interface which is a communication mechanism between the system users and the system allowing the system users to input the data of patients into the system and get the automatically generated recommendations from the system to help to make final clinical decision.

- Non-knowledge-based CDSSs

Unlike knowledge-based CDSSs that get clinical domain knowledge from expert clinicians or medical literature, non-knowledge-based CDSSs use a form of artificial intelligence called machine learning that allows the system to learn from past experience and/or to recognise patterns in clinical data (Marakas, 2003).

CDSSs are classified as knowledge-based systems and non-knowledge-based systems in the following review. Sources of different types of uncertainties in medical decision making are reviewed first, since uncertainties in both clinical domain knowledge and clinical situation are inevitable, and CDSSs are entangled with uncertainties since the very early CDSSs such as MYCIN.

2.3.2. Sources of Uncertainties in Medical Decision Making

Uncertainty exists in almost every stage of a clinical decision making process (Szolovits, 1995). Uncertainties may arise from the following circumstances.

- Patients can not describe exactly what has happened to them or how they feel.
- Doctors and nurses can not tell exactly what they observe.
- Laboratories report results may be with some degrees of error.
- Physiologists do not precisely understand how the human body works.
- Medical researchers can not precisely characterise how diseases alter the normal functioning of the body.
- Pharmacologists do not fully understand the mechanisms accounting for the effectiveness of drugs.
- No one can precisely determine one's prognosis.

The above sources of uncertainties in both medical domain knowledge and clinical symptoms during the process of medical decision making can be summarised as in Table 2-1, where all roles involved in medical decision making are listed, and the resultant uncertainties in medical domain knowledge or clinical symptoms related to each role along with their causes are described as well.

Table 2-1: Sources of Uncertainties in Medical Decision Making (Szolovits, 1995)

Roles Involved in Medical Decision Making	Causes of Uncertainties	Resultant Uncertainties
Patients	Can not describe exactly what has happened to them or how they feel.	Uncertainties in clinical symptoms
Doctors	Can not tell exactly what they observe and may produce laboratory results with some degrees of error.	Uncertainties in clinical symptoms
Nurses	Can not tell exactly what they observe.	Uncertainties in clinical symptoms
Physiologists	Do not precisely understand how the human body works.	Uncertainties in medical domain knowledge
Medical Researchers	Can not precisely characterise how diseases alter the normal functioning of the body.	Uncertainties in medical domain knowledge
Pharmacologists	Do not fully understand the mechanisms accounting for the effectiveness of drugs.	Uncertainties in medical domain knowledge

As shown in Table 2-1, all those listed roles including patients, doctors, nurses, physiologists, medical researchers, and pharmacologists can have uncertain judgments or observations. In consequence, uncertainties in medical domain knowledge, clinical symptoms' description, and phased diagnosis judgments provided by clinicians are indeed unavoidable.

2.3.3. Knowledge-Based CDSSs

In the general structure of knowledge-based CDSSs as shown in Figure 2-1, core components of knowledge-based CDSSs include user interfaces, knowledge base, inference engine, decision models, and database. The user interfaces are used for acquiring system's inputs and displaying system's outputs. The knowledge base is a structured collection of expert medical knowledge used by the system. The inference engine is a set of computerised algorithms used to match clinical inputs with knowledge base to generate clinical recommendations. The decision models are to

provide decision support for clinicians. The database is for storing system inputs and outputs.

In a review study of what makes a CDSS successful in improving clinical practice (Kawamoto et al., 2005), researchers found that a knowledge-based CDSS is as effective as its underlying knowledge base only. In fact, the effectiveness of the knowledge base is dependent on its knowledge representation scheme. Moreover, the inference method used in the inference engine is also closely related to the representation schemes used in a CDSS. The following subsections will shift to review of the knowledge representation schemes and inference mechanisms used in existent knowledge-based CDSSs.

2.3.3.1. Knowledge Representation Schemes

The goal of knowledge representation is to provide intelligent systems with information about a specific domain in a form that can be processed efficiently, and basically, knowledge representation schemes can be classified into four categories: logic, procedural, graph/network, and structured (Carter, 1999). This section reviews knowledge representation schemes according to these four categories.

- Logic

Firstly, logic seems to be the most common representation format used by researchers in the field of general artificial intelligence in the literature. In general, medical knowledge can be divided into two types, namely declarative knowledge and procedural knowledge. Declarative knowledge includes propositions and sentences. Propositions are statements about the world that are either ‘true’ or ‘false’. These statements may be connected by Boolean operators such as ‘and’, ‘or’, and ‘and not’

to form sentences. Procedural knowledge provides more explicit information about what action can be taken or what conclusion can be drawn from declarative knowledge. For example, "'ElectroCardioGram (ECG) shows ≥2mm ST elevation in two contiguous chest leads' or 'ECG shows ≥1mm ST elevation in two contiguous limb leads'" is declarative knowledge, and "IF 'ECG shows ≥1mm ST elevation in two contiguous limb leads', THEN 'treat the patient as with ST-Segment Elevation Myocardial Infarction (STEMI)'" is procedural knowledge. The logic-based representations are declarative in nature, in that they consist of 'true' or 'false' statements and all questions are resolved through standard logic inference mechanism which is simply a 'look up' of known facts (Carter, 1999).

- Procedural knowledge representation

Secondly, procedural knowledge representation, on the other hand, is not simply a 'look up' of known facts. It offers a 'process' to aid diagnostic and therapeutic decision-making (Carter, 1999). Procedural knowledge in medicine is usually provided in the form of rules in existent CDSSs. Many implemented CDSSs, from the very early CDSSs such as MYCIN, PUFF (Aikins et al., 1983), and IMM/Serve (Miller et al., 1996), to recently developed CDSSs such as Unified Medical Language System (UMLS)-based CDSS (Achour et al., 2001), and Chinese medical diagnostic system (Huang and Chen, 2007) are all rule-based. Actually, rules have been the dominant knowledge representation scheme for medical expert systems since the days of MYCIN (Carter, 1999).

In practice, because of the existence of uncertainty in clinical domain knowledge, clinical signs and symptoms as discussed in Section 2.3.2, some CDSSs embed fuzzy

logic, certainty factors or probabilities into traditional 'IF-THEN' rules to represent knowledge with uncertainties.

Fuzzy logic has been widely applied in CDSSs (Shiomi et al., 1995, Suryanarayanan et al., 1995, Palma et al., 2006). Certainty factors together with rules are employed in chest pain diagnosis support system (Hudson and Cohen, 1987, Hudson and Cohen, 1988, Hudson and Cohen, 2002), and Bayes' rule are used in Lliad (Warner, 1989). However, it is hard to get Bayes' probabilities (Spooner, 1999, Spooner, 2007).

- Network-based knowledge representation

Thirdly, in terms of the network-based knowledge representation schemes used in knowledge-based CDSSs, Bayesian belief network (BBN) is a commonly used representation scheme, many CDSSs in the literature such as (Burnside et al., 2006, Nicholson et al., 2008) are based on BBN. A Bayesian network is a way to put Bayes' rule to work by laying out graphically which events influence the likelihood of occurrence of other events. In CDSS design, the choice of adopting a Bayesian network as representation scheme allows one to explicitly take advantage of conditional independencies from the modelling viewpoint, and to rely on several powerful algorithms for probabilistic inference. But it is really very difficult for researchers to derive all necessary parameter values or probabilities among the network (Stefania Montani, 2006). Decision trees are another network-based knowledge representation schemes used in knowledge-based CDSSs, and they are frequently used in guideline-based CDSSs such as EsPeR system (Colombet et al., 2005) and breast cancer treatment CDSS (Skevofilakas et al., 2005). The advantage of decision trees is that they are simple to understand and interpret, and it is possible that

uncertainties can be incorporated into the decision trees. But it is always hard for CDSSs developers to extract an exact decision tree from domain experts.

- Structural representation

Fourthly, structural representations emphasise the 'packaging' of knowledge into well defined pieces with higher levels of organizations (Carter, 1999, Carter, 2007). 'Frame' was the first widely accepted structural knowledge representation format created in 1970s (Minsky and Haugeland, 1975), and it is an application of the object-oriented approach to knowledge-based systems, and a frame is a data structure containing typical knowledge about a concept or object. Some CDSSs such as earlier CENTAUR (Aikins, 1980) and Arden Syntax (Clayton et al., 1989, Starren et al., 1994), and GASTON (de Clercq et al., 2001) and GLARE (Terenziani et al., 2003) all use frame as one of their representation formats. Since each frame has its own name and a set of attributes or slots which contain values; for instance, the frame for *patient* might contain an *age* slot, *sex* slot, *smoking status* slot, etc. frames can be used to construct semantic network model. An important part of every frame is the pointer to a more general frame. The slots are filled with fillers which can be either atomic values or names of other frames. The slots of the generic frames can have procedures attached to them. The reasoning in frame-based system starts by identifying of a given object as an instance of a generic frame. After that all slot fillers which have not been set explicitly but can be inherited, are inherited. Where available, the procedures for frame can be invoked. Disadvantages of a frame-based system include: it can not process objects which characteristics are not known in advance; it can not process non typical situations; the procedural knowledge is not represented by a frame but by the procedures attached to frames (Grigorova and Nikolov, 2007).

In recent CDSS studies, Data-Base Management Systems (DBMS) offer another structured format for knowledge representation, and there are two types of databases which are frequently found in clinical settings—relational and object-oriented (Carter, 2007). Relational DBMS uses a relational record to store and manage data, and each record has a number of fields. Records are then collected together into tables, while each row in the table represents a unique rerecord and each column represents a feature of the record. However, a column in the relational DBMS can not hold more complex data structures, for example, another record, or a list of numbers. Differing from relational database object-oriented DBMS incorporates object-oriented technology into DBMS, where the data is seen as an object, and it permits greater expressiveness by permitting the storage of data types that can not be handled by relational DBMS (Pinciroli et al., 1992). Therefore some CDSSs use object-oriented DBMS to store complex medical datasets which are limited by data types in relational databases. A major drawback of DBMS is that although its Structured Query Language (SQL) can manipulate 'query', 'add', 'update' and 'delete' to its stored objects, it lacks a specific knowledge inference mechanism to reason and draw logic conclusions from the data (Carter, 2007).

As discussed above, most knowledge representation schemes have their own advantages and drawbacks. Especially in dealing with uncertainties, knowledge representation schemes in existent CDSSs lack a mechanism that can comprehensively incorporate or represent different clinical uncertainties in a satisfactory way.

The choice of an appropriate knowledge representation scheme in the construction of a knowledge base depends on the domain knowledge it represents and the inference

mechanism it uses. Inference mechanisms used in existent CDSSs are reviewed as follows.

2.3.3.2. Inference Mechanisms

An inference mechanism used in a CDSS is closely related to the corresponding knowledge representation scheme employed in the system. From the literature, inference mechanisms commonly used in knowledge-based CDSSs include rule-based, Bayesian based, and frame-based.

In rule-based CDSSs, a set of 'IF-THEN' rules are processed. The forward and backward chaining can be used to conclude a diagnosis and provide diagnostic explanations for clinical users (Shortliffe and Perreault, 1990). Take forward chaining for example, forward chaining is a top-down method taking facts as they become available and attempts to draw conclusions from satisfied conditions in rules. The process of inference using forward chaining involves assigning values to attributes, evaluating conditions, and checking to see if all of the conditions in a rule are satisfied so as to fire satisfied rules. If fuzzy logic is incorporated in rule-based systems for handling uncertainties, compositional rule of inference (Zadeh, 1973) is commonly used for fuzzy rule-based inference. If certainty factor as used in MYCIN (Shortliffe, 1976) is incorporated in rules for representing uncertainties, a threshold value need to be set for assessing whether a rule in the rule chain is fired or not.

Bayesian systems predict the posterior probability of diagnoses based on the prior disease probabilities, and the sensitivity and specificity of confirmed clinical signs and symptoms (Warner, 1979). BBNs are often created as reformulations of traditional Bayesian representations and can provide many of the same browsing and

explanation capabilities of traditional systems (Li et al., 1994). Although Bayesian rules are preferred by researchers in statistics, it is very hard to derive all necessary probabilities and sensitivity and specificity of clinical signs and symptoms in target clinical area.

As to reasoning in frame-based systems, it can not process objects which characteristics are not known in advance. Since most medical knowledge is ill-structured and involves uncertainties, it is difficult to use frames to make clinical inference under uncertainties in CDSSs (Grigorova and Nikolov, 2007).

It is important to note that medical experts will turn to concrete examples to express their knowledge when medical knowledge is difficult to be modelled in the format of logical representation. In such situation, the case-based reasoning (CBR) approach (Althoff et al., 1998, Kumar et al., 2009) is used in CDSSs. The advantage of CBR is that concrete similar empirical clinical cases are more convincing than some other implicit medical knowledge. The disadvantages of CBR include that it is difficult to measure the similarity between cases, especially under different types and degrees of uncertainties, the retrieval process is hard to be accurate and efficient, and the input scheme required by the CDSS based on CBR may not be easily accepted by clinicians.

An important aspect of inference engines implemented in CDSSs is their independence from their knowledge base. Since CDSSs take a great deal of time to design and develop, reusability has been a focus of research (Tu et al., 1995).

2.3.3.3. Challenges of Knowledge-Based CDSSs

As reviewed from the literature, clinical uncertainties are inevitable not only in the process of shaping domain knowledge in one formal format but also in each clinical

role's clinical judgments or observations. Thus one important aspect of knowledge representation scheme and inference mechanism is their capability of representing and reasoning under clinical uncertainties.

For knowledge-based CDSSs, dominant knowledge representation schemes used in the literature that can somehow represent uncertainties include fuzzy rules, traditional Bayesian rules, and BBN. Traditional 'IF-THEN' rules can be used together with some certainty factors or fuzzy logic to represent uncertain clinical rules in target clinical domain. Both traditional Bayesian rules and BBN use conditional probabilities to represent clinical uncertainties to some degree.

However, there are drawbacks in reasoning with rules together with certainty factors, or fuzzy logic, or Bayesian probabilities. Firstly, in rule-based CDSSs that incorporates certainty factor model, the certainty factor in the conclusion of one rule is based on the assumption that the premise is known with a certainty factor of 1, and uncertainties are propagated through the rules of an inference chain. However, it is unlikely that a premise is always perfectly known, and the premise of one rule can be uncertain due to uncertain facts. Usually, in the system, a threshold value for premise certainty factors is defined to prevent rules with too low premise certainty factors from firing. For example, the premise threshold certainty factor is set to be 0.2 in MYCIN, and the system will stop triggering one rule if calculated certainty factor of its premise is 0.2 or less. This causes more or less information loss in the inference process. Moreover, Clancy and Cooper (1984) observed that perturbations in certainty factors led to an incorrect diagnosis in certain cases, and this observation suggests that the certainty factor model may be inadequate for diagnostic systems (Heckerman and Shortliffe, 1992). Secondly, in a fuzzy rule-based system, essential inference steps

include: fuzzification and fuzzy rule inferencing, where fuzzification is for interpreting a crisp numerical input as a fuzzy set with the membership function and fuzzy rule inferencing is the process of reasoning with fuzzy rules. In cases when linguistic result expressed by inferred fuzzy set contains required information, there is no need for any defuzzification. In other cases, when a crisp value is needed for the output variable, defuzzification is required. However, fuzzy rule inference is controversial partly because of its fuzzification and defuzzification processes. Thirdly, for Bayesian rule-based CDSSs, it seems that uncertainties in both clinical domain knowledge and clinical signs and symptoms can be considered in both knowledge representation and inference processes, but since this kind of CDSSs take advantages of conditional probability and all necessary probabilities in the Bayesian rules or BBN are difficult to derive, it is difficult for CDSSs researchers to elicit domain knowledge from domain experts and to develop such CDSSs.

To conclude, although some existent knowledge-based CDSSs such as early MYCIN have taken clinical uncertainties into consideration in system implementation, methods used in existent CDSSs have their limitations in knowledge representation or knowledge inference under uncertainties. For example, it is hard for experts to estimate all parameters in Bayesian model based systems, and certainty factor model may be inadequate for diagnostic systems in other clinical areas where the system performance is sensitive to perturbations in certainty factors. Therefore, representation of and reasoning about medical knowledge particularly under uncertainties are areas that require refined methodologies and techniques (Lin et al., 2006, Musen et al., 2006).

2.3.4. Non-Knowledge-Based CDSSs

2.3.4.1. Machine Learning Technologies

In non-knowledge-based CDSSs, clinical domain knowledge is not extracted from domain expert clinicians or medical literature, instead, it is automatically learned from past experience or past clinical data by the system through some machine learning technologies. Commonly used machine learning technologies used in non-knowledge-based CDSSs include artificial neural networks (ANNs) used by Mangalampalli et al. (2006), Yan et al. (2006), and Tan et al. (2008); genetic algorithms (GAs) used by Grzymala-Busse and Woolery (1994) and Levin (1995); and decision tree learning used by Gerald et al. (2002).

- ANNs

ANNs are frequently used by researchers as inference mechanism in CDSSs because developers are not required to understand the relationship between input clinical data and output clinical diagnosis recommendations during the development of this kind of systems. ANNs are a black box technique that models relationships by learning from historical data, while developers of CDSSs based on Bayesian networks need to have sufficient domain knowledge including related probabilities. Li et al. (2000) compare ANNs with other mathematical models for building a traumatic brain injury medical decision support system in their study, and the results suggest that ANNs may be a better solution for complex, non-linear CDSS than conventional statistical techniques. The major advantage of ANNs is that they have the ability to learn from the observed data. The disadvantage is that they are unable to provide reliable and logical

representation of knowledge beyond their learnt zones, and the rules that the network uses do not follow a particular logic and are not explicitly understandable.

- Genetic algorithms (GAs)

GAs provide an approach to learning that is based on simulated evolution (Mitchell, 1997). The problem addressed by GAs is to search a space of candidate hypotheses to identify the best hypothesis. Here hypotheses are often described by bit strings whose interpretation depends on the application. Instead of searching from general-to-specific or simple-to-complex hypotheses, GAs generate successor hypotheses by repeatedly mutating and recombining parts of the best currently known hypotheses, and the search for an appropriate hypothesis begins with a collection of initial hypotheses. At each step, a collection of hypotheses is updated by replacing some fraction by offspring of the fittest current hypotheses. In GAs the 'fittest hypothesis' is defined as the one that optimizes a predefined numerical measure for the problem at hand. For example, if the learning task is to approximate an unknown function based on a set of training examples with inputs and outputs, the hypothesis fitness can be defined as the accuracy of the hypothesis over this training data. GAs have an advantage that by iteratively extracting the best solutions, an optimal solution which fits best can be reached, but how to define the fitness is a challenge in GA based CDSSs (Spooner, 1999, Spooner, 2007).

- Decision tree learning

In knowledge-based CDSSs, decision trees are used to represent domain knowledge if they can be explicitly acquired from domain experts. While in non-knowledge-based CDSSs, decision tree learning is used as a method to automatically acquire knowledge from previous concrete cases that were already solved by domain experts (Hardin and

Chhieng, 2007, Sam et al., 2008). In the process of learning a decision tree from sample data, there is no need to make prior assumptions of data, and decision trees are easily understandable. However, decision tree learning algorithms are unstable since they can produce drastically different hypothesis from training examples that differ just slightly, and there are limitations about the number and data type of output variable (Mitchell, 1997).

2.3.4.2. Challenges of Non-Knowledge-Based CDSSs

On the one hand, as domain knowledge is learned from clinical data for non-knowledge-based CDSSs, system users do not really know what happens in the learning process and how the system handle those uncertainties in the learnt clinical data, and it is this black-box learning process that hinders the use of non-knowledge-based CDSSs. Take ANN-based CDSSs as an example, because clinicians can not really understand the knowledge represented in ANNs, most clinicians would refuse this type of CDSSs in clinical practice (Spooner, 2007).

On the other hand, in contrast to knowledge-based CDSSs, non-knowledge-based CDSSs have an advantage of providing knowledge learning capabilities. This advantage helps this type of CDSSs to be adapted to past clinical experience or clinical data, while being adaptive to clinical practice is an important characteristic of CDSSs to support evidence-based medicine (Sim et al., 2001).

To conclude, non-knowledge-based CDSSs have learning capabilities which help this type of systems being adaptive to clinical practice, but their knowledge learnt from past clinical experience or data are not easily understandable. This more or less hinders clinicians using the systems. A potential research direction is to combine an

easily understandable knowledge representation scheme with the learning capabilities as used in non-knowledge-based CDSSs so that a knowledge-based CDSS can also own learning capability and can be adaptive to clinical practice to provide evidence-based clinical decision support.

2.3.5. Group CDSSs

Group or collaborative clinical decision making is another important research area of CDSSs, and in the early 1990s, Hatcher (1990, Hatcher, 1994) did research on the uniqueness of group CDSSs and proposed to use analytic hierarchy process (AHP) to arrive at a clinical decision consensus in group CDSSs. In the mean time, Rao and his colleagues (Rao et al., 1994, Rao and Suresh, 1995) found that although group decision making is wide spread in medicine, limitations in technology and other factors limited the growth of group CDSSs for medical decision making (MDM). Rao et al. (1996) did an analysis on the classification of MDM from a group CDSS perspective, and then Rao and Turoff (2000) proposed and developed a hypermedia-based group CDSS to support collaborative MDM. *MEDICALWARE*™ (Rao and Turoff, 2000), which is integrated with the group CDSS, is designed to provide problem-solving support, access to clinical algorithms and procedures, expert inference support and several MDM support tools with hypermedia functionality. In the integrated group CDSS, Delphi method (Linstone and Turoff, 1975) was used for supporting group decision making and achieving a group consensus.

However in the literature, there are currently not many publications on group CDSSs yet apart from the above mentioned studies. But group or collaborative clinical decision making is becoming popular in today's health care (Rao et al., 1996, Christensen and Larson, 1993, Rangel, 2009).

2.3.6. Validation of CDSSs

Validation is a crucial component in the development of any CDSS (Berner, 1999). In the literature, appropriate validation design is considered as an important perspective in formal validation of CDSSs (Miller, 1996). Keith and Greene (1994) studied validation of CDSSs from the following perspectives: (1) validation of the expert knowledge; (2) validation of the integrated system; (3) external validation of the system; (4) in-house online trial; (5) multicenter randomised trial in validation of their system. Thomas et al. (1999) used case scenarios to validate their guideline-based CDSS. Becker and colleagues (1997) validated their CDSS by validating not only the knowledge base, but also the inference mechanism.

As discussed in the literature, a sound CDSS validation study contains the following fundamental components: enough clinical cases for validation; an appropriate validation design; knowledge base validation; and inference engine validation.

2.4. A New Belief Rule-base Inference Methodology Using the Evidential Reasoning Approach

Above review helps to give audience a rough holistic picture of existent CDSSs, and through the review we know that representation of and reasoning about medical knowledge particularly under uncertainties are areas that require refined methodologies and techniques. Motivated by this, we looked into the possibility of using a recently developed new belief Rule-base Inference Methodology using the Evidential Reasoning approach (RIMER) (Yang et al., 2006) to implement a CDSS that can represent uncertain clinical domain knowledge and provide informative clinical diagnosis recommendations.

Next we provide a brief discussion of RIMER and its advantages compared to other knowledge representation and reasoning methods used in existent CDSSs.

2.4.1. What is RIMER?

RIMER contains three main parts, one part is a model for representing uncertain knowledge, one part is the method to do inference with knowledge and observed facts, and one more part is an optimization model for fine-tuning knowledge model. In RIMER, the belief rule base (BRB) is used for modelling target clinical domain knowledge and the evidential reasoning (ER) approach is used to do clinical inference, and a BRB optimization model is designed and used to train the belief rule-based CDSS. BRB is extended from traditional rule base by adding a belief structure to it, in which knowledge representation parameters including rule weights, antecedent attribute weights and belief degrees in consequents are considered. The ER approach (Yang and Sen, 1994, Yang and Singh, 1994, Yang and Xu, 2002) was originally proposed to deal with multiple attribute decision analysis (MADA) problem having both qualitative and quantitative attributes under uncertainty. In the situation of reasoning with BRB, ER is employed to combine all belief rules triggered by observed facts in the inference process, where the uncertainties in both observed facts and belief rules can be rationally preserved and their effects can be represented in the final reasoning results.

Details of BRB, inference with BRB using the ER approach, and the BRB optimization model can be found in Section 3.2 of Chapter 3.

2.4.2. Advantages of RIMER

Compared with alternative knowledge representation and reasoning methods used in existent CDSSs, advantages of using BRB for uncertain domain knowledge modelling and using the ER approach for reasoning with uncertain knowledge are discussed as follows.

2.4.2.1. Advantages of Modelling Clinical Domain Knowledge with BRB

When choosing a modelling methodology to model domain knowledge, several factors including naturalness, uniformity, and understandability; degree to which knowledge is explicit (declarative); modularity and flexibility of the knowledge base; efficiency of knowledge retrieval; and capability of uncertainty representation should be taken into account (Turban and Aronson, 2001). For a knowledge-based CDSS, transparency and explanation ability of the system affect user acceptance. The more transparent the system is, the easier will it be for users to accept it (Tsymbal et al., 2009).

Compared to alternative methodologies used to model clinical domain knowledge in existent CDSSs, BRB has following advantages:

- **Transparent representation of domain knowledge**

Not like ANNs representing domain knowledge in black boxes, belief rules can represent domain knowledge in a transparent way. Take a belief rule 'IF there is new left bundle branch block (LBBB) with possibility of 80%, AND the history and examination are strongly suggestive of STEMI, THEN the patient is diagnosed as

with STEMI with 90% belief degree' for example, the relationship between the antecedent 'new LBBB with possibility of 80%, AND the history and examination are strongly suggestive of STEMI' and the consequent 'diagnosed as with STEMI' is transparent. While in ANNs, there is no transparent knowledge about the conclusion 'diagnosed as with STEMI'. A non-expert has no idea about what happens between the input and output, and even those parameters used in ANNs need to be pre-trained by historical data.

- **Naturalness of representation**

Same as traditional rules, belief rules is a very natural knowledge representation method with a high level of comprehensibility, and they look like a natural language expression. For example, even a non-expert in clinical area who has no knowledge of 'LBBB' and 'STEMI' can understand the logic behind the example belief rule discussed above.

- **Handling different types of uncertainties in clinical decision making**

BRB provides a flexible framework to capture different types of uncertainties in clinical signs, symptoms and clinical domain knowledge. Take the belief rule discussed above for example, it represents the uncertainties in domain knowledge for diagnosing one patient as with 'STEMI' when a clinician can not be 100% sure of the patient's Electrocardiograph (ECG) signs and the patient's history and examination. At the same time, the rule captures uncertainties in clinical symptoms such as 'new LBBB with possibility of 80%' and 'the history and examination are strongly suggestive of STEMI'.

- **Provision of explanations**

Given that explanations in CDSSs are necessary, BRB has the ability to provide explanations for the derived conclusions in a straightforward manner. If a patient is diagnosed as with 'STEMI' to a belief degree of 0.9 by a belief rule-based CDSS, the system can provide an explanation of the diagnosis recommendation by presenting users the inputs about the patient's clinical signs and symptoms and all activated clinical rules in the inference process, which are very straightforward.

2.4.2.2. Advantages of Using the ER Approach for Clinical Inference

As to clinical inference, the ER approach has many advantages compared to alternative reasoning methods used in existent CDSSs.

- **Preserving uncertainties in the inference process**

The ER approach initially aims to provide assessment to MADA problems which have both quantitative and qualitative attributes with uncertainties. In the application of reasoning with clinical BRB model, it is used to combine all belief rules triggered by input facts with different belief degrees. Uncertainties in the inference process may be caused either by uncertain domain knowledge or uncertain clinical data. Firstly, uncertainties in domain knowledge such as incompleteness, and nonlinear causal relationships can be represented in belief rules by belief degrees. Secondly, an input with uncertainties to an antecedent clinical symptom can be transformed into a belief distribution on all referential values of the antecedent with different matching degrees, and the distribution describes the degree of each antecedent being activated. Subsequently, inference using the ER approach takes into consideration of both the rule activation weight and belief degrees in possible consequents, and thus both

uncertainties in domain knowledge and input data can be rationally preserved and their affects can be represented in the final reasoned results.

- **Providing informative and prioritised clinical recommendations**

The ER approach can generate a distributed consequent associated with belief degrees after aggregating all activated rules in the inference process, and the inferred results with belief degrees attached to possible consequents can provide an informative clinical recommendation compared to those recommendations with single result. For example, if a patient is diagnosed as with {(STEMI, 0.9878), (LBBB without STEMI, 0.0122), (neither STEMI nor LBBB, 0)} after matching the patient's clinical data with clinical rules in the BRB by the ER approach, we can see that all possible consequents including 'STEMI', 'LBBB without STEMI', and 'neither STEMI nor LBBB' have been associated with belief degrees in the inferred result, and the belief degrees demonstrate different confidence in corresponding consequents, and such a type of recommendation is more informative than inferred result with only one consequent without belief degree such as {STEMI} or {LBBB without STEMI} or {neither STEMI nor LBBB}.

- **Ranking the severity of patients' illness**

ER-based inferred results can provide a severity ranking of patients' illness, based on the concept of utility and utility interval as proposed by Yang and Xu (2002) for the combined assessment result generated by the ER approach. Let us examine the same example again which is discussed above. Similar to the concept of expected utility value, in the context of clinical diagnosis, we can use a severity score ranged from 0 to 1 to represent the seriousness of patient illness, where 1 represents that the patient is in a most serious status and 0 represents that the patient has no clinical risk at all.

For example, we can assign a severity score of 1 to patients with 'STEMI', a severity score of 0.5 to patients with 'LBBB without STEMI', and a severity score of 0.25 to patients with 'neither STEMI nor LBBB', where 'STEMI' is severer than 'LBBB without STEMI' and 'LBBB without STEMI 'is severer than' neither STEMI nor LBBB'. Thus we can get an overall severity score for a patient with distributed diagnosis recommendation, and the overall severity score is calculated by the following equation: (overall severity score) = (severity score of 'STEMI') * (belief degree in 'STEMI') + (severity score of 'LBBB without STEMI') * (belief degree in 'LBBB without STEMI') + (severity score of 'neither STEMI nor LBBB') * (belief degree in 'neither STEMI nor LBBB'). As to the patient with the distributed diagnosis result {(STEMI, 0.9878), (LBBB without STEMI, 0.0122), (neither STEMI nor LBBB, 0)} as discussed above, the overall severity score of the patient can be calculated by (1*0.9878+0.5*0.0122+0.25*0) = 0.9939. It is this overall severity score that can be used as a measure to rank the severity of patients' illness.

- **Learning capability**

Most existent knowledge-based CDSSs such as traditional 'IF-THEN' rule-based systems and frame-based systems lack knowledge learning capability. While domain knowledge used in existent non-knowledge-based CDSSs such as ANN-based systems can only be learned from historical clinical data.

In belief rule-based CDSSs, domain knowledge can be explicitly modelled using BRB. However, it is difficult to accurately determine the parameters of a BRB entirely subjectively, and a change in rule weight or attribute weight may lead to changes in the performance of a BRB. As such, the ER algorithm used for inference with BRB model can be used to form optimization models to train BRB using accumulated past

clinical data. Therefore, inference with BRB model using the ER approach can possess system features of both knowledge-based and non-knowledge-based systems.

However, RIMER has its limitations. For example, just like knowledge representation schemes used in existent knowledge-based CDSSs such as traditional rule-based systems, it is hard to extract belief rules from experts. Though knowledge representation parameters of BRB models can be fine-tuned by historical data, the accuracy of the initial BRB in a belief rule-based CDSS is very important to the system performance.

2.5. Conclusions

After a critical review of the literature on CDSSs, a conclusion can be drawn that a number of CDSSs have been developed in the past 40 years, many of which show potential for making significant impacts on patient care. However, after decades of the development of these programs, no CDSS is widely used by clinicians (Carter, 1999, Carter, 2007).

Miller and Geissbuhler (1999) identified that there are a number of problems that have limited the ultimate success of CDSSs, and these include difficulties with domain selection and knowledge base construction and maintenance, problems with the diagnostic algorithms and user interfaces, and problems with system validation or evaluation.

In more recent studies, Kawamoto et al. (2005) identified four features of CDSSs as independent predictors of a good CDSS:

(1) Automatic provision of decision support as part of clinical work flow.

(2) Provision of recommendations rather than just assessments.

(3) Provision of decision support at the time and location of decision making.

(4) Computer based decision support.

To achieve those four features as identified by Kawamoto and his colleagues, a CDSS should have (a) a friendly user interface that help the clinicians easily play their roles in part of the clinical work flow; (b) a knowledge base which contains comprehensive clinical domain knowledge including uncertainties; (c) an intelligent diagnostic inference mechanism that can handle medical uncertainties; (d) linkage to the whole clinical work flow; and (e) reliable, informative and prioritised clinical decision recommendations. These requirements are consistent with Miller and Geissbuhler's findings about what accounts for the lack of CDSS application in clinical practice.

The problem of developing an adequate database which can store both patients' clinical data and declarative and procedural knowledge may not be difficult to overcome with the rapid development of networking and database technologies. However, representation of and reasoning about clinical domain knowledge under uncertainty, and keeping the knowledge base be adaptive to clinical practice are still the main challenges in CDSSs.

Based on the review results, the following four issues are identified as research gaps in CDSSs literature.

(1) Current CDSSs need a more informative knowledge representation scheme which can represent uncertain clinical domain knowledge comprehensively and accurately (Musen et al., 2006, Lin et al., 2006).

(2) Current CDSSs need a refined inference mechanism which can reason with

information which has different uncertainties (Musen et al., 2006, Lin et al., 2006).

(3) Current CDSSs need an intelligent learning capability to automatically update reasoning rules by learning from past experience or clinical data to make the system be adaptive to clinical practice (Sim et al., 2001);

(4) Few CDSSs in the literature support both individual and group clinical decision making although group MDM attracted attention for CDSSs researchers (Hatcher, 1990, Hatcher, 1994, Rao et al., 1994, Rao et al., 1996, Rao and Turoff, 2000). There is a need for a CDSS that can also support group or collaborative clinical decision making.

To address the gaps as described above, the recently developed RIMER was investigated and employed in this research for clinical knowledge representation and inference under uncertainties (Kong et al., 2008a, Kong et al., 2008b, Kong et al., 2009). In RIMER, a rule base is designed with belief degrees embedded in all possible consequents of a rule. Such a rule base is capable of capturing vagueness, incompleteness, and nonlinear causal relationships, while traditional 'IF-THEN' rules can be represented as a special case. Inference in such a rule base is implemented using the ER approach.

Chapter 3

Research Methodology and Design

3.1. Introduction

Research is defined as a systematic investigation to establish facts or principles or collect information on a subject (Collins English Electronic Dictionary, 2008). Usually, a study is conducted in sequential research process stages from deciding research topic, defining research objectives, and choosing appropriate research methodologies to collecting data, analyzing data, developing conclusions, and finalizing findings. Choosing appropriate research methodologies and making a good research design before conducting core research is important for a study to produce fruitful research results. This chapter discusses the research methodologies and research design of the study. A multi-methodology approach (Nunamaker and Chen, 1990) is employed in the study to investigate how to design, develop, and validate a belief rule-based CDSS that can provide online, intelligent, group and informative clinical decision support under uncertainties. In the research, modelling and prototyping are main research methods for design and development of the target CDSS. Field study is used to acquire more specific clinical domain knowledge and to get better understanding of clinical work flow. Statistical techniques are used to analyze the generated results in validating the developed prototype system.

This chapter is organised as follows. The modelling methodology is discussed in Section 3.2, where three models used in the research for design and development of an intelligent evidence-based CDSS are discussed together with their advantages

compared to available alternative models used in existent CDSSs. The system development methodology-prototyping (Turban and Aronson, 2001) and its advantages compared to other system development methodology are presented in Section 3.3. Field study and statistical techniques are discussed in Section 3.4 and Section 3.5 respectively. A research design is also discussed in Section 3.6. Finally, the chapter is summarised in Section 3.7.

3.2. Modelling

CDSSs are large systems consisting of interrelated components working together in a coordinated manner. Generally, a knowledge-based CDSS should consist of five essential components if we use a Data-Base Management System (DBMS) to store both inputs and outputs of the system, as shown by the general system structure in Figure 2-1 of Chapter 2. The first component is user interfaces, which facilitate communication between system users and the systems. The second one is a knowledge base, which contains the clinical rules necessary for the completion of its task. The third one is a database, in which data and conclusions can be stored. The fourth component is an inference engine, which matches clinical rules to input data to derive its conclusions. The fifth one is decision models employed in the system to provide different types of decision support. Considering machine learning through past clinical experience and group clinical decision support, we can integrate a machine learning functionality and an online group decision supporting functionality into the knowledge-based CDSS. The system should then contain two more components - knowledge training module and group decision supporting module.

It is found from the literature, as discussed in Chapter 2, that a good CDSS should follow the three principles. Firstly, building appropriate knowledge base and inference

engine to provide intelligent and accurate clinical decision support under uncertainties is important for a CDSS to be successful in practice. Secondly, providing online group clinical decision support is necessary for a CDSS to meet the needs of inevitable group clinical decision making in clinical practice. Thirdly, building a knowledge training or learning module to automatically update knowledge base according to accumulated clinical practice is necessary for a CDSS to support today's evidence-based medicine. Different reasoning methods one could use in arriving at a diagnosis in the literature could be using rules, statistics, neural networks, comparison with past cases and so on. The knowledge representation scheme and the knowledge training model chosen are closely related to the reasoning method. The group decision achieving methods used in existent group CDSSs include Delphi method (Rao and Turoff, 2000) and AHP (Hatcher, 1994).

In the research, a new belief rule-based inference methodology called RIMER (Yang et al., 2006) is investigated and employed for the design and development of an online intelligent group CDSS, and the target CDSS can help bridge the research gaps as identified in Chapter 2.

In the clinical BRB model, domain knowledge is represented by a new knowledge representation scheme, i.e. belief rules. They are different from conventional rules in that they are designed with belief degrees embedded in all possible consequents of a rule, and other knowledge representation parameters such as the weights of rules and antecedent attributes are also considered in this scheme. Such a BRB is capable of capturing the vagueness, incompleteness, and nonlinear causal relationships in knowledge (Yang et al., 2006).

In a belief rule-based system, an input to an antecedent attribute is transformed into a belief distribution on referential values of the attribute, and subsequently inference with the BRB is implemented using the ER approach. As a result of the ER-based aggregation of all activated rules in the BRB, all possible consequents in the inferred result are associated with belief degrees.

The ER approach, which is used for aggregating all activated belief rules in RIMER methodology, is also employed to aggregate all group clinicians' diagnosis preferences in the system thanks to its advantages of combining both quantitative and qualitative judgments under various uncertainties.

As for knowledge training in a belief rule-based system, several online and offline BRB training models have been proposed by researchers (Zhou et al., 2009, Zhou et al., 2010, Yang et al., 2007). Some models target both BRB structure and knowledge representation parameters training (Zhou et al., 2010), while some other models target only knowledge representation parameters training (Yang et al., 2007, Zhou et al., 2009). Based on previous research on BRB training in the literature, an optimization model for training the belief rule-based CDSS was implemented in the research.

The following subsections briefly introduce BRB, ER, and the optimization model used for BRB training..

3.2.1. BRB

BRB is extended from traditional rule base by adding a belief structure, in which knowledge representation parameters including rule weights, antecedent attribute weights and belief degrees in consequents are considered.

Conventionally, in a rule base, the *k*th rule in an 'IF-THEN' format can be described as

$$R_k : \text{If } A_1^k \wedge A_2^k \wedge \cdots \wedge A_{T_k}^k, \text{ then } D_k \tag{3-1}$$

where $A_i^k \left(i = 1, \ldots, T_k\right)$ is a referential value of the *i*th antecedent attribute in the *k*th rule, and T_k is the number of the antecedent attributes used in the *k*th rule. D_k is the consequent of the *k*th rule.

If rule weights, antecedent attribute weights, and belief degrees associated with all possible consequents are taken into account, rule described in (3-1) can be extended to a packet rule using a belief structure, which is referred to as a belief rule and can be described as R_k :

$$\begin{aligned} &\text{If } A_1^k \wedge A_2^k \wedge \cdots \wedge A_{T_k}^k, \text{ Then } \left\{\left(D_1, \overline{\beta}_{1k}\right), \left(D_2, \overline{\beta}_{2k}\right), \ldots, \left(D_N, \overline{\beta}_{Nk}\right)\right\} \\ &\left(\overline{\beta}_{ik} \geq 0, \sum_{i=1}^{N} \overline{\beta}_{ik} \leq 1\right), \text{ with a rule weight } \theta_k \text{ and attribute weights } \\ &\delta_{k1}, \delta_{k2}, \ldots, \delta_{kT_k} \quad k \in \{1, \ldots, L\} \end{aligned} \tag{3-2}$$

where $\overline{\beta}_{ik} \left(i = 1, \cdots, N; k = 1, \cdots, L\right)$ is the belief degree originally given by experts to which D_i is believed to be the consequent if in the *k*th belief rule the input satisfies the packet antecedents $A^k = \left(A_1^k, A_2^k, \ldots, A_{T_k}^k\right)$, the attribute weight $\delta_{ki} \left(i = 1, \ldots, T_k; k = 1, \ldots, L\right)$ represents the relative importance of the *i*th antecedent attribute in the *k*th rule, and the rule weight θ_k represents the relative importance of the *k*th rule in the rule base. L is the number of all belief rules in the rule base. T_k is the number of all antecedent attributes used in the *k*th belief rule. *N* is the number of all possible consequents in the rule base.

BRB is a collection of belief rules as described by (3-2). Inference with BRB is implemented using the ER approach, and knowledge representation parameters including rule weights $\theta_k (k = 1,\ldots,L)$, antecedent attribute weights $\delta_{ki} (i = 1,\ldots,T_k; k = 1,\ldots,L)$ and consequent belief degrees $\beta_{ik} (i \in \{1,\ldots,N\})$ can be learned from past experience or data.

3.2.2. The ER Approach

The ER approach (Yang and Sen, 1994, Yang and Singh, 1994, Yang and Xu, 2002) originally aims to deal with multiple attribute decision analysis (MADA) problem having both qualitative and quantitative attributes under uncertainty. The kernel of the ER approach is an ER algorithm which is developed for aggregating multiple attributes based on a belief decision matrix and the evidence combination rule of the Dempster-Shafer (D-S) theory (Shafer, 1976). Different from traditional MADA approaches that describe a MADA problem using a decision matrix, the ER approach uses the belief decision matrix, in which each attribute of an alternative is described by a distribution assessment using a belief structure. How to use the ER approach to do inference with BRB and how to use the ER approach to aggregate group preferences are briefly discussed as follows.

3.2.2.1. Inference with BRB Using the ER Approach

Assume a BRB has L belief rules $\{R_1, R_2, \cdots, R_L\}$, and the kth rule can be described as R_k : If U is A^k , then D is with belief degree β^k , where U represents the antecedent attribute vector, A^k represents the packet antecedents in kth rule, D represents the consequent vector $(D_1, D_2, \cdots, D_N)$ of the rule base, and β^k represents

the vector of belief degrees $(\beta_{1k}, \beta_{2k}, \cdots, \beta_{Nk})$ in the rule base. N is the number of consequents in the BRB and $k \in \{1, \cdots, L\}$. In inference with the BRB using the ER approach, a belief rule expression matrix can be described as Table 3-1.

Table 3-1: A Belief Rule Expression Matrix for the BRB (Yang et al., 2006)

Output	Input					
	$A^1(\omega_1)$	$A^2(\omega_2)$	…	$A^k(\omega_k)$	…	$A^L(\omega_L)$
D_1	β_{11}	β_{12}	…	β_{1k}	…	β_{1L}
D_2	β_{21}	β_{22}	…	β_{2k}	…	β_{2L}
⋮	⋮	⋮	⋯	⋮	⋯	⋮
D_i	β_{i1}	β_{i2}	…	β_{ik}	…	β_{iL}
⋮	⋮	⋮	⋯	⋮	⋯	⋮
D_N	β_{N1}	β_{N2}	…	β_{Nk}	…	β_{NL}

In the belief rule expression matrix, $D_i (i \in \{1, \ldots, N\})$ is the consequent vector D which represents possible consequent in the rule base, and β_{ik} $(i \in \{1, \ldots, N\}, k \in \{1, \ldots, L\})$ represents belief degree associated to the *i*th consequent in the *k*th belief rule in the BRB. ω_k is the activation weight of the *k*th rule, which measures the degree to which the *k*th rule is weighted and activated in the inference process.

The ER algorithm has recursive and analytical formats (Wang et al., 2006), the following brief discussion of inference with BRB is based on the recursive ER algorithm as introduced in the original paper for proposing RIMER (Yang et al., 2006), and the inference process can be described by the following five steps.

- ***Step 1: Transform input clinical data to a distribution on referential values of relevant antecedent symptoms using belief degrees.***

Given an input $U = (U_i, i=1, \ldots, T)$ together with its corresponding belief degree $\varepsilon = (\varepsilon_i, i = 1, \cdots, T)$, where T is the total number of antecedent attributes in the rule base, U_i (i=1, …, T) is the input value of the ith antecedent attribute, and $\varepsilon_i (i = 1, \cdots, T)$ represents the degree of belief assigned to the input value U_i of the ith antecedent attribute, which reflects the uncertainty of the input data. How should the BRB be used to infer and generate output? Before an inference process starts, all input data need to be transformed to a distribution on referential values of each antecedent attribute using belief degrees and this transformation process can be implemented by the rule or utility-based equivalence transformation techniques (Yang, 2001). For example, the input value U_i for the ith antecedent attribute along with its belief degree ε_i can be transformed as

$$S(U_i, \varepsilon_i) = \{(A_{ij}, \alpha_{ij}), j = 1, \ldots, J_i\}, i = 1, \ldots, T \tag{3-3}$$

where A_{ij} is the jth referential value of the ith antecedent attribute, α_{ij} the degree to which the input U_i with belief degree ε_i belongs to the referential value A_{ij} with $\alpha_{ij} \geq 0$ and $\sum_{j=1}^{J_i} \alpha_{ij} \leq 1 (i = 1, 2, \ldots, T)$, and J_i is the number of all referential values of the ith antecedent attribute.

- ***Step 2: Calculate the activation weight of each rule in the BRB.***

After the input transformation, the activation weight $\omega_k (k = 1, \ldots, L)$ which measures the degree to which the packet antecedent A^k in the kth rule is activated, can be calculated with

$$\omega_k = \frac{\theta_k \alpha_k}{\sum_{j=1}^{L} \theta_j \alpha_j} = \frac{\theta_k \prod_{i=1}^{T_k} \left(\alpha_i^k\right)^{\bar{\delta}_{ki}}}{\sum_{j=1}^{L} \left[\theta_j \prod_{l=1}^{T_k} \left(\alpha_l^j\right)^{\bar{\delta}_{jl}}\right]} \quad (k = 1, \cdots, L) \tag{3-4}$$

$\bar{\delta}_{ki} = \frac{\delta_{ki}}{\max\limits_{i=1,\ldots,T_k} \{\delta_{ki}\}}$ ($0 \le \bar{\delta}_{ki} \le 1$) is transformed from antecedent weight $\delta_{ki}\left(i = 1, \ldots, T_k; k = 1, \ldots, L\right)$ representing the relative importance of the *i*th antecedent attribute in the *k*th rule. $\theta_k (k = 1, \cdots, L)$ is the relative weight of the *k*th rule. $\alpha_i^k \left(i = 1, \ldots, T_k\right)$ is the individual matching degree to which the input U_i (i=1,…,T_k) belongs to $A_i^k (i = 1, \cdots, T_k; k = 1, \cdots, L)$ that is the referential value of the *i*th antecedent attribute used in the *k*th rule, and it is generated from the input transformation as described by equation (3-3), with $\alpha_i^k \ge 0$ and $\sum_{i=1}^{T_k} \alpha_i^k \le 1$.

$\alpha_k = \prod_{i=1}^{T_k} \left(\alpha_i^k\right)^{\bar{\delta}_{ki}} \left(k = 1, \ldots, L\right)$ is called the combined matching degree to which the input vector *U* matches the packet antecedent A^k in the *k*th rule. T_k is the total number of antecedents in the *k*th belief rule. *L* is the total number of all belief rules in the BRB.

It can be easily found from equation (3-4) that in the calculation of the combined matching degree $\alpha_k \left(k = 1, \ldots, L\right)$ of input to packet antecedent, all individual matching degrees $\alpha_i^k \left(i = 1, \ldots, T_k; k = 1, \cdots, L\right)$ and all antecedent weights $\delta_{ki}\left(i = 1, \ldots, T_k; k = 1, \ldots, L\right)$ have been taken into account, and then the calculation of the activation weight $\omega_k (k = 1, \cdots, L)$ takes into consideration both rule weights $\theta_k (k = 1, \cdots, L)$ and the calculated combined matching degrees of input to packet antecedent. This means that all knowledge representation parameters play their roles in calculating a rule's activation weight.

- ***Step 3: Update belief degrees to possible consequents in the BRB based on the input information.***

As for the belief degree $\overline{\beta}_{ik}$ ($0 \le \sum_{i=1}^{N} \overline{\beta}_{ik} \le 1$; i=1,…, N; k=1,…, L) which is originally given by experts when a BRB having a collection of rules as described by equation (3-2) is established, if $\sum_{i=1}^{N} \overline{\beta}_{ik} = 1$, the kth belief rule is said to be complete; otherwise, it is incomplete. If $\sum_{i=1}^{N} \overline{\beta}_{ik} = 0$, it means the output of the kth belief rule is completely unknown. In such situation, the incompleteness of the consequent in a rule is caused by a lack of domain knowledge or expert experience, and the inferred result from this incomplete rule should be incomplete according to the properties of the ER approach (Yang and Xu, 2002). In other situation, when the input data is incomplete, for example, the sum of matching degrees of an input to all referential values of an antecedent attribute is smaller than 1, the inferred result from this incomplete input data should be incomplete as well (Yang et al., 2006). For instance, if the input for the antecedent attributes of a rule is completely unknown, a completely unknown consequent will be generated. If the input of antecedent attributes is partially known, the inferred result will also be partially known or incomplete. In the inference process, the incompleteness in input data should be taken into consideration, because an incomplete input for an antecedent attribute will cause an incomplete output after inference with the rule where the attribute plays its antecedent role. Considering the incompleteness of input data, belief degrees in consequents of a rule need to be updated based on the real input. More specifically, the original belief degree $\overline{\beta}_{ik}\left(0 \le \overline{\beta}_{ik} \le 1; \sum_{i=1}^{N} \overline{\beta}_{ik} \le 1\right)$ given to the ith possible consequent $D_i\left(i = 1,..., N\right)$ in the kth rule which is extracted from experts should be updated on the basis of the actual input information in the inference process by

$$\beta_{ik} = \bar{\beta}_{ik} \frac{\sum_{t=1}^{T_k} \left(\tau(t,k) \sum_{j=1}^{J_t} \alpha_{tj} \right)}{\sum_{t=1}^{T_k} \tau(t,k)} \ (i = 1, \cdots, N; k = 1, \cdots, L; t = 1, \cdots, T_k) \qquad (3\text{-}5)$$

where β_{ik} is the belief degree in consequent D_i when the kth rule is activated by the actual input and it is determined by original belief degree $\bar{\beta}_{ik}$ together with the incompleteness of real input data, in which

$$\tau(t,k) = \begin{cases} 1, \text{if the } t^{\text{th}} \text{ antecedent attribute is used in defining } R_k \ (t = 1, \ldots, T_k). \\ 0, \text{otherwise}. \end{cases},$$

and α_{tj} is the degree to which the input U^k belongs to the referential value $A_{tj} (t = 1, \cdots, T_k; j = 1, \cdots, J_t)$ with $\alpha_{tj} \geq 0$ and $\sum_{j=1}^{J_t} \alpha_{tj} \leq 1 (j = 1, \cdots, J_t)$. The transformation from input U^k to A_{tj} is described as equation (3-3), where A_{tj} is the jth referential value of the tth antecedent attribute, and R_k is the kth rule in the BRB. N is the number of consequents in the kth rule. T_k is the number of all antecedents in the kth rule, and J_t is the number of referential values of the tth antecedent attribute in the kth rule.

- ***Step 4: Aggregate all activated rules using the ER approach to generate a combined belief degrees in possible consequents.***

Once the activation weight of each rule and belief degrees in the possible consequents of each rule in a BRB have been determined by the input clinical data, the ER algorithm (Yang and Xu, 2002) can be applied directly to aggregate all activated rules in a BRB to generate the combined degrees of belief in the consequents of a BRB as follows

$$O(U) = \{(D_j, \beta_j), j = 1, \ldots, N\} \quad (3\text{-}6)$$

This equation reads as that given an input to a belief rule-based system in the vector form of $U = \{U_i, i = 1, \ldots, T\}$, the outcome is consequent D_j with a belief degree of $\beta_j (j = 1, \ldots, N)$.

- ***Step 5 (optional): If necessary, calculate expected severity and severity interval of different diagnostic consequents to rank the severity of patients' illness caused by the same disease.***

For example, if the severity score of 'H' clinical risk is set to be 1, 'M' clinical risk set to be 0.5, and 'L' clinical risk set to be 0, a patient's overall severity score would be 0.8 if he/she is assessed as 'H' clinical risk with 60% probability and 'M' clinical risk with 40% probability. The details of the concept and calculation of the expected utility and utility interval of the ER approach can be found in Yang and Xu (2002).

For better understanding, a flowchart illustrated by Figure 3-1 can be used to describe the whole inference process using the ER approach in a belief rule-based CDSS.

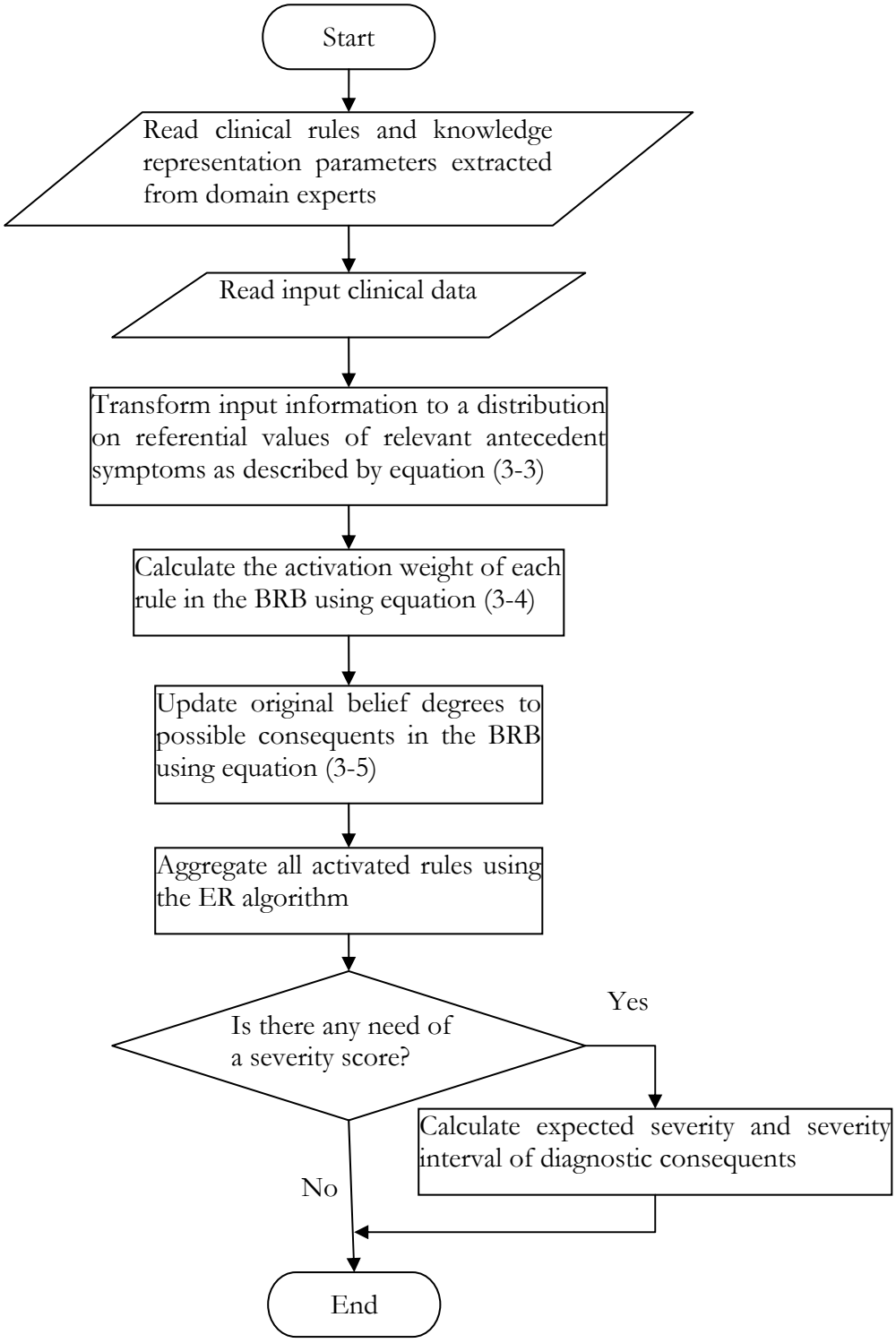


Figure 3-1: Inference with BRB Using the ER Approach in Recursive Format

3.2.2.2. Group Preferences Aggregation Using the ER Approach

Assume there are H consultants ($C_h(h \in \{1,\ldots,H\})$) participating in a group consultation for one patient, and there are R possible diagnosis results ($D_r(r \in \{1,\ldots,R\})$) about the patient. When we use the ER approach for group reference aggregation, the group decision making problem can be expressed in matrix format as follows:

$$D = \begin{array}{c} \\ D_1 \\ D_2 \\ \vdots \\ D_R \end{array} \begin{array}{c} \begin{array}{cccc} C_1 & C_2 & \cdots & C_H \end{array} \\ \begin{bmatrix} \beta_{11} & \beta_{12} & \cdots & \beta_{1H} \\ \beta_{21} & \beta_{22} & \cdots & \beta_{2H} \\ \vdots & \vdots & \vdots & \vdots \\ \beta_{R1} & \beta_{R2} & \cdots & \beta_{RH} \end{bmatrix} \end{array}, W = \begin{bmatrix} \omega_1 & \omega_2 & \cdots & \omega_H \end{bmatrix} \tag{3-7}$$

where C_h denotes the hth consultant, D_r denotes the rth possible diagnosis result about the patient, $\beta_{rh}(r \in \{1,\ldots,R\}, h \in \{1,\ldots,H\})$ represents belief degree provided by the hth group consultant C_h to the rth diagnosis result D_r about the patient, and $\omega_h(h \in \{1,\cdots,H\})$ is the weight assigned to C_h representing the importance of the hth group consultant in the group preference aggregation process. The detailed steps of using the ER approach to aggregate group consultant preferences can be described as follows.

- ***Step 1: Invite a group of clinicians to participate a group consultation for one patient.***

Assume we have a patient, who is with CCP, and the clinician on duty is not sure about what should be taken in the next step based on current status of the patient, then we need to carry out a group consultation for assessing clinical risk of the patient. If there

are *H* experts in CCP field who are available for the group consultation, the clinician on duty can act as a group facilitator role for the group consultation and invite these *H* experts to participate it.

- ***Step 2: Acquire group consultant diagnosis preferences expressed as a distributed assessment of possible diagnoses.***

If possible diagnosis results for one patient is $D_r\left(r \in \{1,\ldots,R\}\right)$ and there are *H* consultants are in the group consultation, the group facilitator can request the *h*th consultant input his/her risk assessment as $\{(D_1,\beta_{1h}),(D_2,\beta_{2h}),\ \ldots,\ (D_R,\beta_{Rh})\}$. Take above mentioned patient with CCP for example, possible risk status of the patient may include: 'very high' (D_1), 'high' (D_2), 'low' (D_3), and 'no' (D_4), and group facilitator can request the *h*th participated clinician express his/her risk assessment as {(D_h, β_{1h}), (D_h, β_{2h}), (D_h, β_{3h}), (D_h, β_{4h})}, for example, {('*very* high', 0.8), ('high', 0.2), ('low', 0), ('no', 0)}.

- ***Step 3: Assign weight to each group consultant based on his/her expertise and reputation.***

The weight $\omega_h\left(h \in \{1,\cdots,H\}\right)$ of each participated clinician represents the importance of each individual clinician based on his/her expertise or reputation in the group consultation. Using the ER approach to aggregate all group preferences can take into consideration of the importance of each group member. For example, there are two experts in CCP who are invited to participate a group consultation, one is an experienced clinician and is in a senior position while the other is less experienced and in a junior position, the group facilitator may assign a weight of 1 to the former and a weight of 0.7 to the latter in the group consultation.

- ***Step 4: Aggregate all group consultant diagnosis preferences using the ER approach to generate a combined belief degrees in possible diagnoses.***

As soon as we get all the parameters such as weight $\omega_h \left(h \in \{1,\cdots,H\}\right)$ of individual consultant and the distributed diagnosis preference $\{(D_1,\beta_{1h}),(D_2,\beta_{2h}),\ \ldots,\ (D_R,\beta_{Rh})\}$ provided by each consultant, we can use the analytic ER algorithm to aggregate the group diagnosis preference and get a combined belief degrees in all possible diagnoses. The analytic ER algorithm is as follows.

$$\beta_r = \frac{\mu \times \left[\prod_{h=1}^{H}\left(\omega_h \beta_{rh} + 1 - \omega_h \sum_{r=1}^{R} \beta_{rh}\right) - \prod_{h=1}^{H}\left(1 - \omega_h \sum_{r=1}^{R} \beta_{rh}\right)\right]}{1 - \mu \times \left[\prod_{h=1}^{H}\left(1-\omega_h\right)\right]}, r = 1,\cdots,R \quad (3\text{-}8)$$

with $\mu = \left[\sum_{r=1}^{R}\prod_{h=1}^{H}\left(\omega_h \beta_{rh} + 1 - \omega_h \sum_{r=1}^{R} \beta_{rh}\right) - (R-1)\times \prod_{h=1}^{H}\left(1 - \omega_h \sum_{r=1}^{R} \beta_{rh}\right)\right]^{-1}$

where $\beta_r \left(r = 1,\cdots,R\right)$ is the final belief degree attached to the *r*th possible diagnosis D_r after combining all group consultant diagnosis preferences in the group consultation, $\beta_{rh}\left(r = 1,\cdots,R; h = 1,\cdots,H\right)$ is the belief degree assigned to the *r*th possible diagnosis D_r by the *h*th consultant, and ω_h is the weight of *h*th group consultant.

In the existent group CDSSs, methods used to achieve a group consensus in the group decision making situation include Delphi method and AHP. The Delphi method seeks to achieve a consensus among group members through a series of questionnaires which requires several rounds for group members to fill in the questionnaires, yet it is actually time consuming and needs carefully designed questionnaires to acquire participants' opinions in each round (Linstone and Turoff, 1975). If we use AHP in the group decision making context, just as using AHP for supporting individual

decision making, we need to acquire comparison information with regard to each pair of objects be determined. Traditionally, there are two approaches that can be used to get the comparison information as to each pair of objects. Firstly, the entire group provides a single numeric value for each pair of objects, for example, object "*i*" compared to object "*j*", and we can take the group provided comparison information as a "consensus" of each pair objects. Secondly, we can request each group member provide a numeric value that reflected her/his view of the relative importance of object "*i*" compared to object "*j*", and then we use geometric mean of all group member input numeric values to get a "consensus" of comparison information of each pair of objects. After we get comparison information for each pair of objects, we can use AHP method to aggregate the comparison information and get a final preference order of the objects. However, using AHP in group decision making context to achieve an aggregated preference order of objects has its limitations. The first option of acquiring a group consensus on comparison information as to each pair of objects suffers from the negative effects of status influence (or power differential problem) which could prevent the realization of real consensus, and a major disadvantage of the second option is that wide disparities in the comparison information could result in the computed 'consensus' matrix being an inaccurate representation of the given situation at human level (Bryson, 1996).

Compared to the above mentioned group decision support methods in the existent group CDSSs, using ER in group clinical decision making context to achieve a group consensus has the following advantages. Firstly, rather than seeking individual group member's diagnosis preference through several rounds of questionnaires as in Delphi method, using the ER approach for group preference aggregation only needs each individual group member's belief degrees on all possible diagnosis results for one

patient. It helps speed up the group decision making process. Secondly, uncertainties in the judgement of all possible diagnosis alternatives can be reflected by the distributed assessment, where the individual group member can provide both complete and incomplete judgements. Note here complete judgement means belief degrees of all possible alternatives add up to 1, while incomplete judgement means the sum of belief degrees of all possible alternatives is smaller than 1. Thirdly, influence of different individual group member on the final group consensus is reflected by weight assigned to individuals in the preferences aggregation process. Fourthly, based on the utility and utility interval of the group aggregated judgement for a patient, the severity of the patient's illness can be calculated.

3.2.3. An Optimization Model for BRB Training

As described in Section 3.2.1, the initial belief rules and knowledge representation parameters including rule weights, attribute weights and consequent belief degrees are originally given by domain experts or randomly generated, and they can not be 100% accurate. To make the BRB to represent clinical domain knowledge more accurately, we need to train or fine-tune the BRB using historical data.

Generally, machine learning comes in two categories: supervised learning and unsupervised learning (Hardin and Chhieng, 2007) as shown in Figure 3-2. In supervised learning the goal of learning is to adjust the knowledge representation model through minimizing the discrepancy between the system results and observed results of the training sample as shown in Figure 3-2 (A). As for unsupervised learning, we know nothing about the knowledge representation model ahead and just let the system learn meaningful structure from a set of historical data as shown in Figure 3-2 (B).

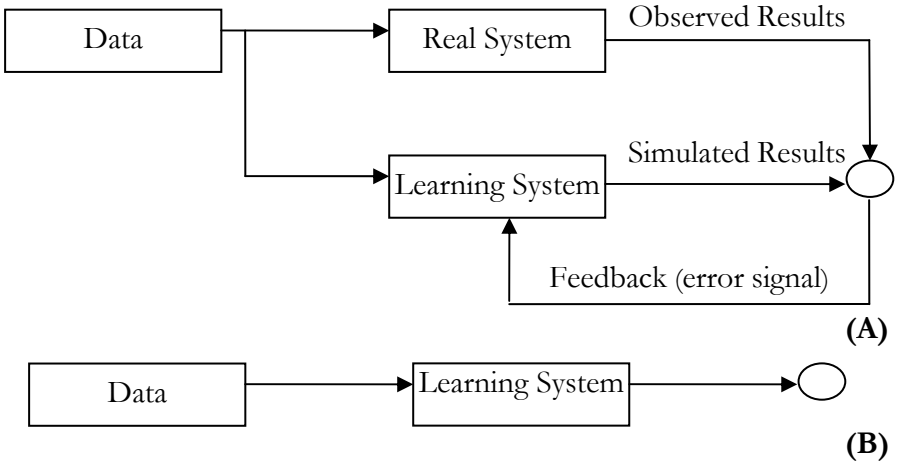


Figure 3-2 (A): Supervised Learning; (B): Unsupervised Learning. (Hardin and Chhieng, 2007)

In terms of BRB training, several online and offline models have been proposed in the literature, where online training models target real-time BRB training by newly input data and offline training models target BRB training by accumulated historical data. All these BRB training methods fall into supervised learning category which requires an initial BRB acquired from domain experts and uses a historical clinical dataset with real clinical outcome to train the BRB.

In the research, target domain BRB is originally constructed based on expert clinicians' experience and knowledge and those clinical rules have been verified by clinical research. Although domain knowledge in medicine keeps changing, it changes at a comparatively low speed, and we do not have to update the domain BRB hourly or daily. Thus, we choose the offline BRB training methods to train the developed belief rule-based CDSS prototype by available accumulated clinical data.

As for the offline BRB training methods, Yang et al. (2007) proposed several optimal learning models for training BRB. Depending on the type of input and output of sample data, the optimal learning model can be constructed in different ways (Yang et

al., 2007). For example, if the output of the training sample is of numerical type, a single objective nonlinear optimization model can be constructed by minimizing the total mean squared error between the simulated output and the observed output of the sample data.

The following is a brief discussion of the BRB training model. The training parameters of the optimization training model may consist of different sets of knowledge representation parameters including rule weights, attribute weights and belief degrees. If necessary, utilities or scores associated with different consequents can be used for training as well. The objective of the optimization model is to minimise the discrepancy between system generated diagnosis results and the observed clinical status of the real or simulated patients, and the discrepancy is calculated by total mean squared error between system and observed results. The constraints of the optimization model are constructed based on the characteristics of those knowledge representation parameters and the utility values. Details of the optimization model are as follows.

The aim of BRB training is to find a set of parameters $(\theta_k, \delta_i, \beta_{jk})(k = 1, \cdots, L; i = 1, \cdots, N; j = 1, \cdots, N)$ of a BRB that can help the BRB to represent domain specific knowledge correctly. The training process is implemented through minimizing the discrepancy between the system generated results and observed results of the training sample. Assume there are *M* cases in the training sample, and the input-output pairs of those *M* cases are $(\hat{x}_m, \hat{y}_m)(m = 1, \cdots, M)$. The process of learning from these *M* datasets can be depicted in Figure 3-3, where the system generated output (y_m) is produced by the system via the inference engine. The real output $(\hat{y}_m)$ is observed by experts or acquired by instruments, and $\xi(P)$

represents the difference between the real output and the system generated output. The objective function of the BRB optimization model is to minimise $\xi(P)$, and the constraint function can be defined based on the conditions that the training parameters must satisfy. As a result of the training process, there will be a new set of $(\theta_k, \delta_i, \beta_{jk})$ for the BRB.

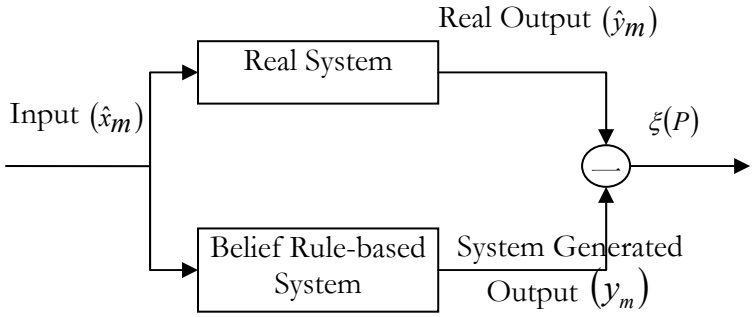


Figure 3-3: Training Process

Regarding the objective function of the training model, we used the total mean squared error $\left(1/M\right)\sum_{m=1}^{M}(y_m - \hat{y}_m)^2$ to represent $\xi(P)$. As an explicit ER aggregation function is required in BRB training, the analytical ER algorithm (Wang et al., 2006) was used to construct the objective function in the training model. When the analytical ER algorithm is applied to inference with BRB, final belief degrees $\beta_j(j=1,\cdots,N)$ attached to all possible consequents $D_j(j=1,\cdots,N)$ are inferred using the following equation:

$$\beta_j = \frac{\mu \times \left[\prod_{k=1}^{L}\left(\omega_k \beta_{jk} + 1 - \omega_k \sum_{j=1}^{N}\beta_{jk}\right) - \prod_{k=1}^{L}\left(1 - \omega_k \sum_{j=1}^{N}\beta_{jk}\right)\right]}{1 - \mu \times \left[\prod_{k=1}^{L}(1-\omega_k)\right]}, j = 1,\cdots,N \quad (3\text{-}9)$$

where $\beta_{jk}(j=1,\cdots,N; k=1,\cdots,L)$ is the original belief degree assigned to the *j*th consequent in the *k*th belief rule,

$\mu = \left[\sum_{j=1}^{N}\prod_{k=1}^{L}\left(\omega_k \beta_{jk} + 1 - \omega_k \sum_{j=1}^{N}\beta_{jk}\right) - (N-1)\times\prod_{k=1}^{L}\left(1-\omega_k\sum_{j=1}^{N}\beta_{jk}\right)\right]^{-1}$, and ω_k is the *k*th rule's activation weight which is calculated using equation (3-4). Here the diagnosis result generated by ER based inference engine for a patient is a set of belief degrees attached to all possible consequents as $(D_j, \beta_j)(j = 1, \cdots, N)$ other than a single numerical output. We need to transform the inferred result of the *m*th case to a single numerical value (y_m) based on the concept of utility and utility interval proposed by Yang and Xu (2002), and the transformation is implemented using $y_m = \sum_{j=1}^{N}\mu(D_j)\beta_j (m = 1, \cdots, M)$, where $\beta_j(m)(j = 1, \cdots, N; m = 1, \cdots, M)$ is generated by the inference engine, and $\mu(D_j)(j = 1, \cdots, N)$ is the utility value or score we set for the *j*th consequent D_j. As for the constraints' setting for the BRB training model, it depends on specific domain knowledge and domain experts' judgements.

3.3. Prototyping

Once the methodology of RIMER has proven feasible to model the clinical domain knowledge and to do clinical inference in target clinical areas, a computerised CDSS prototype can be developed to test whether such a CDSS is really reliable and useful in a real clinical decision making scenario. Prototyping is employed as the system development methodology in the research.

3.3.1. Brief Introduction to Prototyping

Prototyping is an adaptation of the traditional system development life cycle (SDLC). A traditional SDLC starts from some kind of need and results in a completed system, and it consists of four fundamental phases-planning, analysis, design, and implementation which lead to a deployed system (Turban and Aronson, 2001). In a

traditional SDLC, an ideal progression is to follow each phase in order, yet it is possible to return to any phase from any other. While in prototyping methodology, a system is developed sequentially in modules, and it is deployed to users and gains feedback from users for further refinement when each module is completed, so that a prototype system can be quickly developed and demonstrated to users. Figure 3-4 shows a typical prototyping development process.

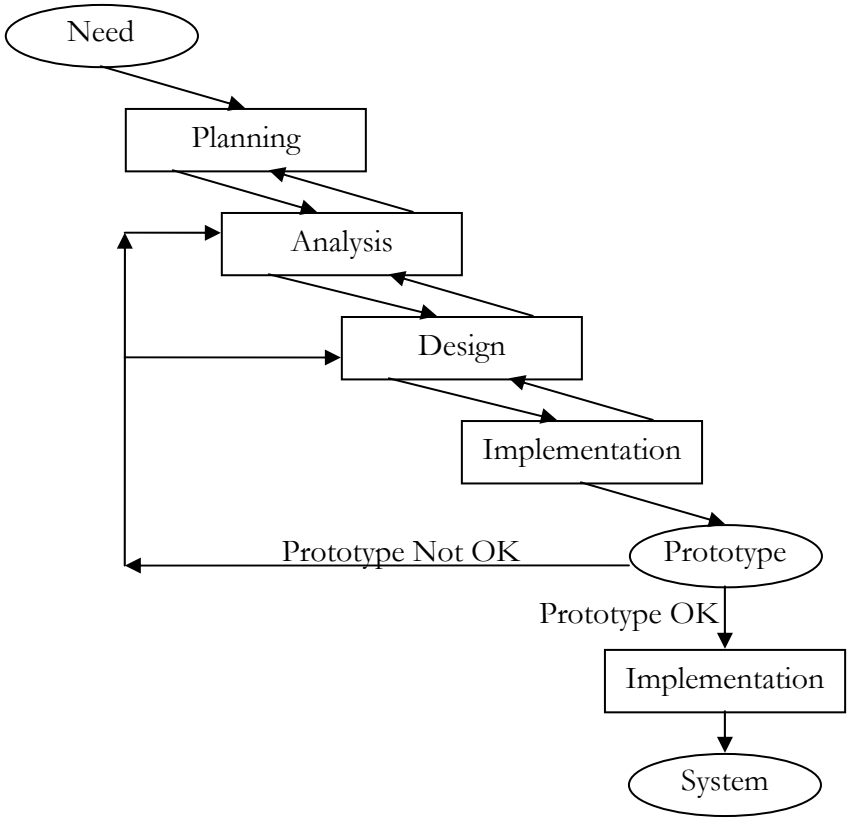


Figure 3-4: Prototyping Development Process (Turban and Aronson, 2001)

Following the development process as shown in Figure 3-4, the analysis, design, and prototype implementation phases are iteratively performed until a small prototype is sufficiently developed.

3.3.2. Alternative System Development Methodologies

There are several alternative system development methodologies which include the traditional SDLC, parallel development which resembles the SDLC, and phased development.

In parallel development, the design and implementation phases are split into multiple sub-ones after the analysis phase, and each of the sub-ones involves development of a separate subsystem. All the sub-ones come together in a single implementation phase, in which a system integrator puts the subsystems together in a cohesive system.

In the phased development methodology, a system is developed sequentially by a series of versions. Each version has more functionality than its previous version, and they evolve into a final system.

3.3.3. Advantages of Prototyping

Compared to alternative development methodologies, prototyping has the following advantages in developing a CDSS:

- **Users are involved in every phase and iteration.**

Unlike in the traditional SDLC, users only play roles in the planning phase when system developers seek information from them. The iterative nature of prototyping allows users to be involved in system design and development, which is important for a CDSS. Getting system users involved in system design and development process enables us to learn from them gradually about the ill-structured clinical domain knowledge and system users' real requirements of a CDSS in clinical practice.

- **Prototype can be developed quickly.**

Unlike a traditional SDLC, prototyping essentially bypasses the formal steps of

requirements acquisition and analysis, since prototyping attempts to clarify users' needs by actively involving them in a fast-feedback development process, and this helps a prototype system to be developed quickly in the prototyping system development methodology.

In the research, we need to design and develop a CDSS first and then use clinical cases to test whether the developed CDSS is really reliable and useful. Due to limited time and human resources, alternative system development methodologies discussed above are not suitable, and prototyping methodology is the appropriate method to develop a CDSS prototype for the study.

3.4. Field Study

In the research, practice guidelines in target areas are important resources for knowledge elicitation in construction of knowledge base. But the domain knowledge in target clinical areas keeps changing, and all the practice guidelines that we obtained at the beginning of the study can be outdated soon. In such situation, we need to hold regular interviews and meetings with expert clinicians for acquiring advanced clinical domain knowledge. Moreover, field study is necessary for us to investigate the daily clinical work flow in NHS hospitals and to acquire real user requirements for a CDSS in clinical practice.

3.5. Statistical Techniques

After system design and development, real or simulated clinical cases are used to validate the developed prototype system. For example, for inference engine validation, real or simulated patients' data can be used as inputs to the prototype system, and then the automatic diagnostic recommendations generated by the prototype system can be

used to compare with real clinicians' conclusions, and finally a conclusion about the prototype system's reliability can be drawn based on the diagnostic performance provided by the system.

Statistical techniques are used to analyze the diagnosis results generated in the prototype validation process. The receiver operating characteristics (ROC) curve (Metz, 1978, Park et al., 2004) is employed to analyze the diagnostic performance of the different diagnosis tests taken in the validation, and the area under the ROC curve (AUC) (DeLong et al., 1988, Mei-Ling Ting and Bernard, 2001) is used to compare diagnostic performance of the different diagnosis tests. Brief introduction of the statistical techniques used in the research can be found in Chapter 6. .

3.6. Research Design

Once research questions and objectives of a study are formulated concretely, a researcher develops a research design as a strategic plan to conduct the study. Research design is a format for detailed steps in a study to tackle previously identified research questions and to achieve already set research objectives.

The research design comprises data collection, prototype CDSS design, development and validation. The research is conducted based on a multiple-methodologies research approach. The research consists of five research stages, the details of which are outlined in Table 3-2.

Table 3-2: Research Design

Stages	Actions	Objectives	Questions to Answer	Research Methodologies
Stage 1	1. Literature review 2. Theoretical investigation	1. To identify research gaps 2. To formulate research questions	Research question (1-1)	Literature review
Stage 2	A feasibility study ⇨ Domain knowledge acquisition ⇨ Compare a traditional rule-based CDSS with a belief rule-based CDSS	1. To acquire target clinical domain knowledge 2. To analyze the feasibility of applying RIMER for development of a CDSS	Research questions (1-2)& (1-3)	1. Literature review 2. Field study 3. Modelling
Stage 3	1. Design of a belief rule-based online intelligent group CDSS 2. Development of the belief rule-based online intelligent group CDSS prototype	To design and develop an online belief rule-based group CDSS with learning capability	Research questions (1-2) & (1-3) & Research questions (2) & (3)	1. Modelling 2. Prototyping
Stage 4	Validation of the developed CDSS prototype ⇨ Domain knowledge acquisition & Data simulation ⇨ Validation of inference engine: compare system's diagnostic performance with one doctor's ⇨ Validation of training module: compare the system's diagnostic performance before and after BRB training	To test the reliability of the prototype	Research question (1-2) & (1-3)& Research questions (3)	1. Field study 2. Statistical analysis
Stage 5	1. Finalizing main results 2. Presenting the prototype system and system manual	1. To draw conclusions about the feasibility of employing RIMER for developing a CDSS 2. To present the final prototype system and the user manual		

(1) Identifying Research Problem

At research stage 1, a research topic on CDSSs is chosen, and information with respect to the topic and its problems is initially extracted from the literature. Existing CDSSs are reviewed for its further requirements and research questions are formulated based on the identified research gaps. The potential of employing advanced methods and technologies in decision making areas such as RIMER to address the research problems is investigated. The research during this stage includes the following activities:

- Review and analyze existent CDSSs;
- Provide a key feature analysis in uncertainty handling for the existent CDSSs;
  i. Identify challenges of knowledge representation schemes and inference mechanisms in existent knowledge-based CDSSs;
  ii. Identify challenges of knowledge learning and representation mechanisms in existent non-knowledge-based CDSSs;
  iii. Identify challenges of group decision support in existent CDSSs;
  iv. Explore the possibility and creativeness of using advanced methods and technologies from decision making area to design and develop an intelligent group CDSS that can handle different uncertainties well;
  v. Formulate research questions;
- Set detailed measurable research objectives for the research.

(2) A Preliminary Feasibility Study

At research stage 2, a clinical domain is chosen as a target clinical area for a preliminary feasibility study, and specific domain knowledge need to be extracted from domain medical literature or/and acquired from expert clinicians through field

study. Subsequently, the feasibility of using RIMER for the design and development of a belief rule-based CDSS is investigated. The research at this stage includes the following activities:

- Acquire target clinical domain knowledge through literature review and field study;
- Construct BRB models based on the domain knowledge;
- Construct an inference model using the ER approach;
- Compare the belief rule-based system with traditional rule-based system in drawing clinical conclusions about simulated cases in target clinical areas.

(3) Design and Development of a Belief Rule-based Online Intelligent Group CDSS Prototype

At research stage 3, we design and develop a belief rule-based online intelligent group CDSS. The system design and development need to consider the details of system architecture, back-end database, user interfaces, knowledge base, inference mechanism, group decision supporting module, and training module. The basic prototyping system development process as shown in Figure 3-4 is applied in the prototype development. The research at this stage includes the following activities:

- Identify the prototype's characteristics in varying aspects: web-based architecture, programming languages, software environment, and key components or functionalities of the system;
- Design and develop the back-end database schema to store clinical data and to physically construct BRB models;
- Design and develop web-based user interfaces based on the clinical work flow depicted in target clinical practice guidelines;

- Design and develop ER-based inference engine;
- Design and develop ER-based group decision supporting module;
- Design and develop a BRB training module;
- Integrate all key components together and present the online belief rule-based group CDSS prototype to target system users;
- Improve the prototype based on users' feedback.

(4) Validation of the Online Intelligent CDSS prototype

At research stage 4, clinical data in target clinical areas is used for validating the developed online intelligent CDSS prototype. In the validation, two core components: inference engine and training module are validated respectively. The activities at this stage include:

- Acquire target clinical domain knowledge through field study;
- Collect second-hand real patients' data or invite expert clinicians to simulate clinical data;
- Test the system's diagnostic performance using the acquired dataset;
- Split the dataset into training and test sets based on some criteria;
- Train BRB model in the system using the training dataset;
- Test the trained system using the test dataset;
- Analyze the system's diagnostic performance before and after BRB training.

(5) Present the Final Prototype CDSS and User Manual

This final research stage summarises the results, draws conclusion about the feasibility and viability of applying RIMER in development of a CDSS and presents the final prototype system and system manual.

3.7. Summary

As identified in Literature Review chapter, representing and reasoning with clinical domain knowledge under uncertainties are areas that require refined technologies, and there are strong needs for a CDSS to provide group clinical decision support and automatic knowledge base updating in addition to individual clinical decision support. Thus we propose to employ RIMER methodology to implement a belief rule-based CDSS, while group clinical decision support and automatic clinical knowledge base updating are also taken into account. In this chapter, the main models we used for domain knowledge representation, clinical inference, and clinical knowledge base training are briefly discussed first, and followed by a brief description of the system development method - prototyping, and then we provide a brief discussion on field study which we used for domain knowledge acquisition and clinical practice observation, and the statistical method – ROC analysis which we employed to analyse the system validation results, and finally, a concrete research design is provided based on proposed methods. The advantages of using BRB for modelling or representing domain knowledge, using ER for individual clinical diagnosis and group clinical preferences aggregation, and using an offline BRB training model for automatic knowledge base updating include: (1) BRB can help transparently represent domain knowledge under uncertainties in an natural 'IF-THEN' rule format with a belief structure, which can be provided as explanation if required; (2) inference with BRB using ER can help preserve uncertainties rationally in the inference process and represent their effects in the finally conclusion, while aggregating group preferences using ER can help speed up the group decision making process compared to Delphi method. In the following chapter, we will provide a preliminary study on feasibility of employing RIMER for implementation of a CDSS.

Chapter 4

A Preliminary Feasibility Study

4.1. Introduction

Prior to the design and development of an online intelligent CDSS based on the newly developed belief rule-base inference methodology-RIMER, a preliminary study on feasibility of employing RIMER for developing a CDSS was conducted.

This chapter discusses a detailed investigation for the feasibility of using RIMER in developing a CDSS. In the feasibility study, clinical risk assessment of acute upper Gastrointestinal (GI) bleed was chosen as target clinical area, and a patient with acute upper GI bleed was simulated by an expert clinician in MRI for the investigation. In the study, knowledge-based systems including traditional rule-based system and belief rule-base system are constructed for investigation.

In a traditional rule-based CDSS, forward chaining or backward chaining is used as the inference method, while in a belief rule-based CDSS, the evidential reasoning (ER) approach is employed as the inference method. Equipped by an intelligent decision system (IDS) (Xu and Yang, 2005), which is a Windows-based multiple criteria assessment system that implements the ER approach, we obtained a diagnostic recommendation through inference with the belief rule base (BRB) using ER in a belief rule-based CDSS. Through comparison of diagnosis recommendation generated by belief rule-based system and traditional rule-based system, we find that the diagnosis recommendation generated by a belief rule-based CDSS is more

informative than the diagnosis conclusion drawn from a traditional rule-based system when there is uncertainty in clinical data.

The chapter is organised as follows. Section 4.2 discusses target clinical domain knowledge acquisition for the feasibility study. Section 4.3 presents the domain knowledge modelling, where the modelling of the domain knowledge using traditional 'IF-THEN' rule base and BRB are discussed in Section 4.3.1 and Section 4.3.2, respectively. A description of a simulated patient in target clinical area will follow in Section 4.4. Inference with the constructed knowledge base and the clinical data is discussed in Section 4.5, where inference with the traditional rule base using forward chaining method is discussed in Section 4.5.1, and inference with the BRB using the ER approach is discussed in Section 4.5.2. Then conclusions about the feasibility of using RIMER for developing a CDSS are provided in Section 4.6, where the advantages of a belief rule-based system compared with a traditional rule-based system is discussed as well. Finally, Section 4.7 summarises the chapter.

4.2. Domain Knowledge Acquisition

Knowledge acquisition is a very important starting procedure for the construction of knowledge bases in CDSSs. The first step of knowledge acquisition is to select the target clinical area and select expert clinicians to gain domain specific knowledge, and then transfer the domain knowledge into computer interpretable knowledge based on the designed knowledge representation scheme.

Target clinical domain selection and domain specific knowledge elicitation for the feasibility study are described as follows in Section 4.2.1 and Section 4.2.2 respectively.

4.2.1. Clinical Domain Selection

Upper GI bleeding is a significant and potentially life-threatening worldwide problem. Despite advances in diagnosis and treatment, mortality and morbidity have remained constant (Marc et al., 2000). In the UK, acute upper GI bleed is a common medical emergency with an incidence of approximately 100 per 100,000 adults per year and a mortality among unselected cases of 14% (Rockall et al., 1995).

Patients with upper GI bleeding vary in severity from those with exsanguinating haemorrhage from oesophageal varices to those with simple streaking due to Mallory-Weiss tear caused by retching after too much alcohol the night before. Thus to provide proper management of patient with acute upper GI bleed, it is important for ED doctors to make evidence-based decisions about the clinical risks involved to ensure that appropriate, timely treatment is provided and that investigation is carried out in an appropriate time-frame. Patients at high risk should be resuscitated and undergo emergency endoscopy immediately. Patients at moderate risk should have Intravenous (IV) access established, have their blood grouped and serum saved and should have endoscopy performed rapidly (Central Manchester and Manchester Children's University Hospitals NHS Trust, 2003b).

Motivated by the above facts, we chose clinical risk assessment of acute upper GI bleed as target clinical domain for investigation in the feasibility study.

4.2.2. Target Domain Knowledge Elicitation

Generally, for acquiring medical domain specific knowledge, many computerised knowledge acquisition tools have been developed by CDSSs researchers. Among them, some tools such as a Unified Medical Language System (UMLS)-based

knowledge acquisition tool developed by Achour et al. (2001) is designed for acquiring domain specific medical rules, and other tools such as a guideline acquisition module in a guideline-based CDSS developed by Terenziani et al. (2001) is designed specially for the acquisition of clinical guidelines which can be used as the best and standardised clinical procedures.

During this study we have expert clinicians in MRI as research collaborators who have already published a clinical practice guideline for diagnosis of acute upper GI bleed in ED. Therefore we do not have to use any computerised knowledge acquisition tool which require our research collaborator to input their domain knowledge as clinical rules or clinical guideline. Instead, we elicited rules for assessing clinical risk of acute upper GI bleed first from the published practice guideline, and then we invited our collaborators in MRI to verify those clinical rules before applying them to construct knowledge base for target CDSS.

The clinical decision support guideline (CDSG) 2003-05 for acute upper GI bleed (Central Manchester and Manchester Children's University Hospitals NHS Trust, 2003b), which we used for eliciting clinical rules for risk assessment of upper GI bleed, was developed by clinicians in MRI, and were originally published on central Manchester and Manchester children's university hospitals NHS trust intranet in 2003. Their content was reviewed by Clinical Effectiveness Committee of the British Association for Emergency Medicine in 2005. As presented in the practice guideline, rules for assessing clinical risk of acute upper GI bleed are as follows in Table 4-1, where SBP stands for ‘systolic blood pressure’, and NSAIs represents ‘non-steroidal anti-inflammatory drugs’.

Table 4-1: Rules for Clinical Risk Assessment of Upper GI Bleed

Clinical signs and symptoms, laboratory tests, and medical history	Clinical Risk		
	High (H)	Moderate (M)	Low (L)
Known or suspected oesophageal varices	√		
Pulse > 130 bpm	√		
SBP < 90 mm Hg	√		
Postural SBP drop > 20 mm Hg	√		
On NSAIs or anticoagulants		√	
Major co-morbidity (eg cardiac or renal		√	
Stigmata of liver disease		√	
Witnessed acute fresh red blood in vomit		√	
Over 75 years old		√	
Urea > 8		√	
None of the above			√

Apart from acquiring domain knowledge from the CDSG as mentioned above, we did field study in MRI to observe one expert clinician's clinical practice and held regular meegtings and discussions with expert clinicians to get correct and deep understanding of these clincial rules.

Tradtitional rule base and BRB are constructed, in the following section for developing a CDSS, based on the clinical rules as described in Table 4-1.

4.3. Domain Knowledge Modelling

4.3.1. Modelling with Traditional 'IF-THEN' Rule Base

Traditional 'IF-THEN' rule base is the dominant knowledge modelling methodology in CDSSs (Carter, 1999). If we use traditional rules to represent the rules for assessing clinical risk of upper GI bleed as described in Table 4-1, the rule base as shown in Table 4-2 can be constructed.

Table 4-2: Traditional Rule Base for Risk Assessment of Acute Upper GI Bleed

Number	Antecedent	Consequent
1	(F^1 is Y)	R is H
2	(F^2 is Y)	R is H
3	(F^3 is Y)	R is H
4	(F^4 is Y)	R is H
5	(F^1 is N ^ F^2 is N ^ F^3 is N ^ $F^4$is N ^ F^5 is Y)	R is M
6	(F^1 is N ^ F^2 is N ^ F^3 is N ^ F^4 is N ^ F^6 is Y)	R is M
7	(F^1 is N ^ $F^2$is N ^ F^3 is N ^ F^4 is N ^ F^7 is Y)	R is M
8	(F^1 is N ^ F^2 is N ^ F^3 is N ^ F^4 is N ^ F^8 is Y)	R is M
9	(F^1 is N ^ F^2 is N ^ F^3 is N ^ F^4 is N ^ F^9 is Y)	R is M
10	(F^1 is N ^ F^2 is N ^ F^3 is N ^ F^4 is N ^ F^{10} is Y)	R is M
11	(F^1 is N ^ F^2 is N ^ F^3 is N ^ F^4 is N ^ F^5 is N ^ F^6 is N ^ F^7 is N ^ F^8 is N ^ F^9 is N ^ F^{10} is N)	R is L

In above table, Y stands for 'yes', N stands for 'no', '^' is a logical connective to represent the 'AND' relationship, and the meaning of other symbols are as follows.

R represents 'the clinical risk of acute upper GI bleed'.

F^1 represents 'known or suspected oesophageal varices'.

F^2 represents 'pulse > 130 bpm'.

F^3 represents 'SBP < 90 mm Hg'.

F^4 represents 'postural SBP drop > 20 mm Hg'.

F^5 represents 'on NSAIs or anticoagulants'.

F^6 represents 'major co-morbidity (eg cardiac or renal impaiment)'.

F^7 represents 'stigmata of liver disease'.

F^8 represents 'witnessed acute fresh red blood in vomit'.

F^9 represents 'over 75 years old'.

F^{10} represents 'urea > 8'.

Note that the same set of symbols will be used in the following sections to describe BRB which is constructed by extending the traditional rule base using the belief structure.

4.3.2. Modelling with BRB

In the traditional rule base as described in Table 4-2, there is no uncertainty in the rules' antecedent or consequent. However, there are indeed at least three circumstances in which uncertainties may arise. Firstly in a real clinical environment, conditions in a rule may not always be met with 100% accuracy by patients' clinical data because there are uncertainties in doctors' subjective judgements about one patient's specific clinical symptoms. Secondly, diagnosis conclusions drawn from different clinicians about one patient may not be the same due to the fact that different clinician in the same clinical area may own different domain knowledge and different practice experiences. Finally even two patients are diagnosed as having the same disease, the severity of the two patients may be different, and accordingly, the two patients may need to be treated in a different time order. To deal with the uncertainties, belief rules may provide an alternative solution to accommodate different types of uncertainties in representing both clinical data and clinical domain knowledge.

If those traditional rules described in Table 4-2 are extended using the belief structure for more precisely imitating human reasoning knowledge in rule-based CDSSs, the corresponding BRB representing clinical rules for risk assessment of upper GI bleed can be described as in Table 4-3. For details of the belief structure in BRB, readers can refer to Section 3.2.1 in Chapter 3.

Table 4-3: BRB for Clinical Risk Assessment of Acute Upper GI Bleed

Number	W	Antecedent	Consequent
1	1	(F^1 is Y)	R is {(H, 1)}
2	1	(F^2 is Y)	R is {(H, 1)}
3	1	(F^3 is Y)	R is {(H, 1)}
4	1	(F^4 is Y)	R is {(H, 1)}
5	1	(F^1 is N ^ F^2 is N ^ F^3 is N ^ F^4 is N ^ F^5 is Y)	R is {(M, 1)}
6	1	(F^1 is N ^ F^2 is N ^ F^3 is N ^ F^4 is N ^ F^6 is Y)	R is {(M, 1)}
7	1	(F^1 is N ^ F^2 is N ^ F^3 is N ^ F^4 is N ^ F^7 is Y)	R is {(M, 1)}
8	1	(F^1 is N ^ F^2 is N ^ F^3 is N ^ F^4 is N ^ F^8 is Y)	R is {(M, 1)}
9	1	(F^1 is N ^ F^2 is N ^ F^3 is N ^ F^4 is N ^ F^9 is Y)	R is {(M, 1)}
10	1	(F^1 is N ^ F^2 is N ^ F^3 is N ^ F^4 is N ^ F^{10} is Y)	R is {(M, 1)}
11	1	(F^1 is N ^ F^2 is N ^ F^3 is N ^ F^4 is N ^ F^5 is N ^ F^6 is N ^ F^7 is N ^ F^8 is N ^ F^9 is N ^ F^{10} is N)	R is {(L, 1)}

In the BRB as shown in Table 4-3, the weight of each rule (represented by **W** in the second column) and the weight of each antecedent attribute are assumed to be 1, which means all rules possess the same importance and all antecedent attribute play similar roles in assessing one patient's clinical risk. Besides, each rule has a consequent only with a belief degree of exactly one, which means if one patient's clinical data meets one rule's antecedent condition, the patient will be 100% with a clinical risk at the level as the rule's consequent describes.

Although based on experts' experience, we have not obtained rules like 'IF (F^1 is 80% Y) THEN R is {(H, 0.8)', which can explicitly represent uncertainties in clinical domain knowledge and clinical data. The rules as described in Table 4-3, which represent certain clinical rules in assessing risk of upper GI bleed, are special cases of belief rules, and they can be used to inference with uncertain clinical data.

4.4. Description of Clinical Data

Due to the strict data protection regulations in the UK, we used simulated patient data to demonstrate the risk assessment process in both the traditional rule-based and belief

rule-based CDSSs. The made-up patient's data have been verified by an expert clinician.

The detailed information about a simulated patient with acute upper GI bleed is given in Table 4-4, where only 'esophageal varices' are judged as 'strongly suspected', and all other clinical signs or symptoms are with certain judgments or exact numerical values.

Table 4-4: One Simulated Patient with Acute Upper GI Bleed

Disease	Clinical signs and symptoms
upper GI bleed	strongly suspected oesophageal varices; pulse is 131 bpm; SBP is 90 mm Hg; postural SBP drop is 20 mm Hg; currently is on anticoagulants; no major co-morbidity; no stigmata of liver disease; no fresh red blood in vomit; 76 years old; urea is 8.

Note that the same patient's data has been used in the published paper of Kong et al. (2009) for illustration of inference with BRB.

Inference with the simulated patient's data and constructed knowledge base is described in the following Section.

4.5. Inference with Knowledge Base

4.5.1. Inference with Traditional Rule Base

There are two methods of inference often used in traditional rule-based CDSSs, namely *forward* and *backward chaining* (Shortliffe and Perreault, 1990). In the study, *forward chaining* is used for reasoning with the traditional rule base as described in Table 4-2. Forward chaining 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.

Inference with the traditional rule-based system using forward chaining involves assigning values to attributes, evaluating conditions, and checking to see if all of the conditions in a rule are satisfied. A general algorithm for forward chaining method can be described as in Figure 4-1.

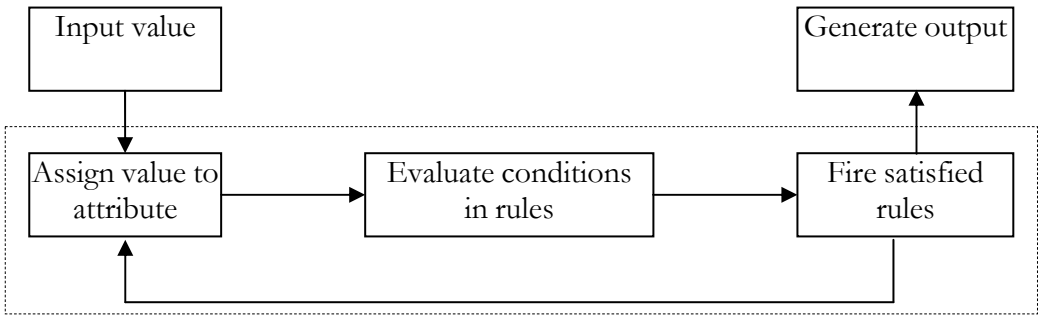


Figure 4-1: A General Algorithm for Forward Chaining in Rule-Based System

In the traditional rule-based system as investigated in the study, all rules are chained according to the real work flow shown in the practice guideline. To acquire inputs, the system would provide a chain of enquiries regarding clinical signs and symptoms which are necessary for specific diagnosis. For example, enquiries for clinical risk assessment in diagnosis of acute upper GI bleed may include:

1) Are there known or suspected oesophageal varices?
2) Is pulse > 130 bpm?
3) Is SBP < 90 mm Hg?
4) Is postural SBP drop > 20 mm Hg?
5) On NSAIs/anticoagulants or not?
6) Is there major co-morbidity (eg cardiac or renal impairment)?
7) Is there a stigmata of liver disease?

8) Is there witnessed acute fresh red blood in vomit?

9) Is the patient over 75 years old?

10) Is the patient's urea > 8?

Only answers Y and N are provided as options for system users to choose for all enquiries like those listed above in traditional rule-based system.

If we need to assess the simulated patient's clinical risk using the traditional rule base as described in Table 4-2, our work will be based on the general algorithm as described in Figure 4-1. Firstly, the matching between the clinical data and antecedent attributes of the traditional rule base can be described as (F^1 is Y)^(F^2 is Y)^(F^3 is N)^(F^4 is N)^(F^5 is Y)^(F^6 is N)^(F^7 is N)^(F^8 is N)^(F^9 is Y)^(F^{10} is N). Secondly only condition of Rule 1 and Rule 2 in Table 4-2 are satisfied by the data. As a result, the inferred clinical risk of the patient generated by the traditional rule-based CDSS is H.

4.5.2. Inference with BRB

Inference with BRB using the ER approach also involves assigning values to attributes, evaluating conditions and checking to see if all of the conditions in a rule are satisfied. However, inference with BRB using the ER approach is different from inference with traditional rule base using forward chaining in many aspects. Firstly, value assignments in the ER approach are different from forward chaining due to an input transformation process. In a rule base, each antecedent attribute has a set of referential values, and individual referential value is used in different rule as an element of antecedent (Yang et al., 2006). Specifically, the *k*th rule in a traditional 'IF-THEN' rule base can be described as R_k : IF $A_1^k \wedge A_2^k \wedge \cdots \wedge A_{T_k}^k$, THEN D_k, where

$A_i^k \left(i = 1, \cdots, T_k\right)$ is a referential value of the *i*th antecedent attribute A_i in the *k*th rule, T_k is the number of the antecedent attributes used in the *k*th rule, and D_k is the consequent in the *k*th rule. In traditional rule-based systems, input data is usually with certainty, and it can be matched directly with antecedents of rules in the system. While in belief rule-based systems, input clinical data can be with uncertainty, and the relationship between an input and each referential value in the antecedents of a rule needs to be determined before an inference process can start (Yang et al., 2006). This process is to transform an input into a distribution on referential values of one antecedent attribute using belief degrees. Secondly, the condition evaluation process is different. In the ER approach, since inputs can be transformed to distributed referential values as described above, conditions of more than one rule may be satisfied by one input patient's data in parallel to different degrees. While in forward chaining, there is only an activated rule or a chain of rules activated in sequence by an input patient's data. Thirdly, conclusions are derived by an aggregation process. Using the ER approach, the conclusions generated by all the activated rules need to be aggregated to generate an overall conclusion. While in forward chaining, there is no rule aggregation or combination process.

Section 4.5.2.1 to 4.5.2.4 will provide a detailed description of the inference process carried out by the belief rule-based CDSS for the simulated patient.

4.5.2.1. Input Transformation

In the BRB outlined in Table 4-3, {Y, N} is used as a set of referential values for all clinical symptoms. The distributed values transformed from original clinical data of the simulated patient described in Table 4-4 in Section 4.4 are as follows: F^l: {(Y,

0.8), (N, 0.2)}, F^2: {(Y, 0.67), (N, 0.33)}, F^3: {(Y, 0.5), (N, 0.5)}, F^4: {(Y, 0.5), (N, 0.5)}, F^5: {(Y, 1), (N, 0)},F^6: {(Y, 0), (N, 1)}, F^7: {(Y, 0), (N, 1)}, F^8: {(Y, 0), (N, 1)}, F^9: {(Y, 0.6), (N, 0.4)}, and F^{10}: {(Y, 0.5), (N, 0.5)}. In the clinical data transformation process, a rule based transformation method (Yang, 2001) is adopted for transforming both qualitative and quantitative input data. Here are the details of the transformation.

- **Qualitative Input Transformation**

As for the qualitative input information regarding F^1, F^5, F^6, F^7 and F^8, 'Y' and 'N' are the set of referential values for all these clinical symptoms. For F^1, {known, strongly suspected, maybe, suspected with a low degree, no} are used as its input options, and transformation rules should be set between the input options and the referential values of {Y, N}. According to an expert clinician's advice, the following transformation rules are used for F^1 related inputs transformation: 'known' means 100% 'Y', 'strongly suspected' means 80% 'Y' and 20% 'N', 'maybe' means 50% 'Y' and 50% 'N', 'suspected with a low degree' means 20% 'Y' and 80% 'N', and 'no' means 100% 'N'. In real life application, the options set for acquiring original information about qualitative clinical symptoms should be set depending on the domain experts' knowledge and experiences. There is no need to establish transformation rules for clinical symptoms of F^5, F^6, F^7 and F^8, because the referential values 'Y' and 'N' are options for acquiring original input. Based on the clinical data as described in Table 4-4, the transformed values for F^1, F^5, F^6, F^7 and F^8 are F^1: {(Y, 0.8), (N, 0.2)}, F^5: {(Y, 1), (N, 0)},F^6: {(Y, 0), (N, 1)}, F^7: {(Y, 0), (N, 1)}, and F^8: {(Y, 0), (N, 1)}.

- **Quantitative Input Transformation**

As for quantitative input data regarding F^2, F^3, F^4, F^9 and F^{10}, their inputs are numerical values, and 'Y' and 'N' are also used as referential values for these symptoms. Moreover, each quantitative clinical symptom is associated with two types of threshold values defined by domain experts including an upper limit value and a lower limit value. Transformation rules for these quantitative clinical symptoms should include that (a) if the input is larger than the upper range value, the input can be transformed to 100% 'Y' or 'N'; (b) if the input is lower than the lower range value, the input can be transformed to 100% 'N' or 'Y'; and (c) if the input falls into the range between the lower and the upper limit values, the input can be transformed to $\{(\mathrm{Y}, \alpha_Y = \frac{input\ value - \text{lower range value}}{\text{upper range value} - \text{lower range value}} * 100\%), (\mathrm{N}, \alpha_N = 1\text{-}\alpha_Y)\}$ or $\{(\mathrm{N}, \alpha_N = \frac{input\ value - \text{lower range value}}{\text{upper range value} - \text{lower range value}} * 100\%), (\mathrm{Y}, \alpha_Y = 1\text{-}\alpha_N)\}$, where α_Y stands for the belief degree to which the input value can be transformed to 'Y' and α_N stands for the belief degree to which the input value can be transformed to 'N'.

The reason for us to adopt a saturated linear utility change process as discussed above rather than a step utility change process in transforming original input numerical value of each quantitative clinical symptom is that the transformed inputs can make the inference in the system better imitate human decision making in a real scenario. Take F^2 for example, if the input value of F^2 is larger than (>) 130 bpm, then the patient will be at high risk according to the original rules as shown in Table 4-1. However, it is unknown what will be the judgment if the input value is exactly 130 bpm. In a real clinical risk assessment of patients with upper GI bleed, a clinician would make his assessment about a patient with pulse of 130 bpm based on his experience and other observations instead of using 130 as the only standard to make a judgment. To solve this problem in the belief rule-based CDSS, a value area of (127, 133) is used as an

interval to define a gradual change area in risk assessment for patient's pulse. Therefore, the input value for enquiry about F^2 equal to or higher than (≥) 133 bpm will be transformed to F^2: {(Y, 1), (N, 0)} and the input value for enquiry about F^2 equal to or lower than (≤) 127 bpm will be transformed to F^2: {(Y, 0), (N, 1)}. If the input value lies in the range of (127, 133), it will be transformed to F^2: {(Y, $\alpha_Y = \frac{input\ value - 127(\text{lower range value})}{133(\text{upper range value}) - 127(\text{lower range value})} * 100\%$), (N, α_N = 1-α_Y)}. Similar transformation will be implemented in input information for enquiries about F^3, F^4, F^9 and F^{10} and the linear change area is (85, 95) for F^3, (15, 25) for F^4, (70, 80) for F^9 and (5, 11) for F^{10}.

Based on the clinical data described in Table 4-4, the transformed values for F^2, F^3, F^4, F^9 and F^{10} are F^2: {(Y, $\alpha_Y = \frac{[131(input\ value) - 127(\text{lower range value})]}{[133(\text{upper range value}) - 127(\text{lower range value})]} * 100\% = 0.67$), (N, α_N=1-α_Y=0.33)}, F^u_3: {(N, α_N =(90-85)/(95-85)*100%=0.5), (Y, α_Y=1-α_Y =0.5)}, F^u_4: {(Y, α_Y=(20-15)/(25-15)*100%=0.5), (N, α_N=1-α_Y=0.5)}, F^u_9: {(Y, α_Y=(76-70)/(80-70)*100%=0.6), (N, α_N=1-α_Y=0.4)}, and F^u_{10}: {(Y, α_Y=(8-5)/(11-5)*100%=0.5), (N, α_N=1-α_Y=0.5)}.

4.5.2.2. Rules' Activation Weights Calculation

After the value assignment for antecedent symptoms, the next step should be to calculate the activation weight for each packet antecedent in the rule base. Using $\alpha_k = \prod_{i=1}^{T_k} \left(\alpha_i^k\right)^{\bar{\delta}_{ki}} \left(k = 1, \ldots, L\right)$ as described in equation (3-4), the combined matching degrees of the input patient's data to each rule's packet antecedent are calculated as follows: α_1=0.8, α_2=0.67, α_3=0.5, α_4=0.5, α_5=0.0165, α_6=0, α_7=0, α_8=0, α_9=0.0099,

α_{10}=0.0083, and α_{11}=0, and the activation weights $\omega_k\left(k=1,\ldots,11\right)$ for all rules are generated using $\omega_k=\frac{\theta_k\alpha_k}{\sum_{j=1}^{L}\theta_j\alpha_j}(k=1,\cdots,L)$ as described in equation (3-4) as follows:

ω_1=0.3194, ω_2=0.2675, ω_3=0.1996, ω_4=0.1996, ω_5=0.0066, ω_6=0, ω_7=0, ω_8=0, ω_9=0.004, ω_{10}=0.0031 and ω_{11}=0.

4.5.2.3. Belief Degrees Update

What follows rules' activation weights calculation is to update belief degrees in the possible consequents of the activated rules in the BRB as shown in Table 4-3. According to the activation weights $\omega_k\left(k=1,\ldots,11\right)$ for each rule in the BRB as calculated in above Section, Rules 1, 2, 3, 4, 5, 9 and 10 are activated to different degrees by the simulated patient's data. After updating of belief degrees in consequents using equation (3-5) based on the transformed input values described in Section 4.5.2.1, it can be found that the updated belief degrees in possible consequents of the rules in the BRB remain original values because all the transformed inputs are complete. Here, a complete input means that if the input U^k is transformed to the original distributed referential values with belief degrees as described in equation (3-3), $\sum_{j=1}^{J_i}\alpha_{ij}\left(i=1,\cdots,T_k;j=1,\cdots,J_i\right)$ should be 1. Take F^l for example, the sum of α_Y (0.8*)* and α_N (0.2) of F^l transformed from the simulated input is 1, which means that the input to the antecedent symptom F^l is complete. If inputs related to all antecedent attributes are completes, the packet input will be a complete one and the belief degrees in the consequents of the BRB will not be affected by the inputs and remain as the original values given by domain experts.

4.5.2.4. Rules Aggregation via ER

Finally, IDS (Xu and Yang, 2005), a Microsoft Windows-based multiple criteria assessment system which implements the ER approach, is used as a tool to aggregate all the activated rules. First, we need to model the belief rule-based clinical risk assessment in the ER framework by taking each patient's illness as an alternative to be assessed, taking clinical risk as the top attribute for the assessment of the patient's illness, and taking each activated rule's packet antecedent as a basic attribute for the assessment of the top attribute. In this model, each rule's activation weight acts like a basic attribute's weight, and each possible consequent of the BRB acts like each individual evaluation grade set for the basic attribute. Accordingly, belief degrees in possible consequents in the activated rules act like belief degrees to possible evaluation grades. The model framed in IDS together with the inputs of the activated rules' activation weights are shown in Figure 4-2, in which the clinical risk assessment model of the simulated patient is shown in the upper left side and the dialog box in IDS for acquiring each activated rule's weight is shown in the lower right side where each activated rule is treated as a basic attribute in the model. A dialog box in IDS for acquiring belief degrees in the possible consequents of each activated rule for the clinical risk assessment model is shown in Figure 4-3.

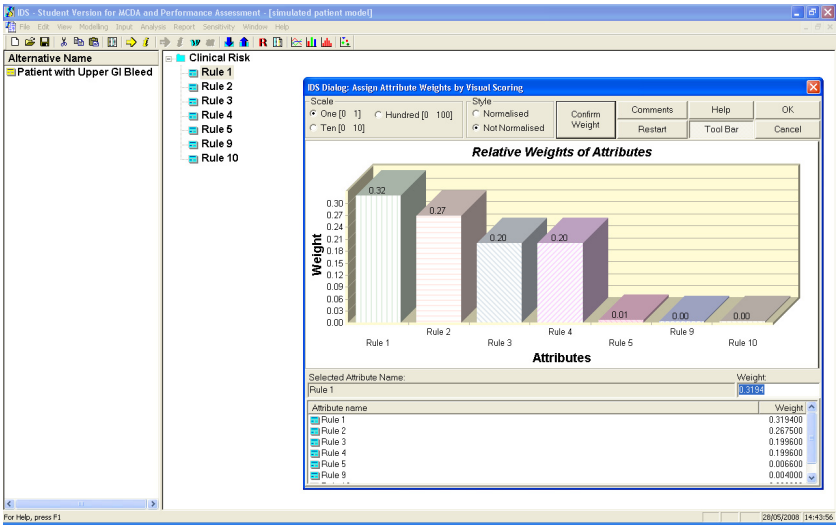


Figure 4-2: Clinical Risk Assessment Model in IDS

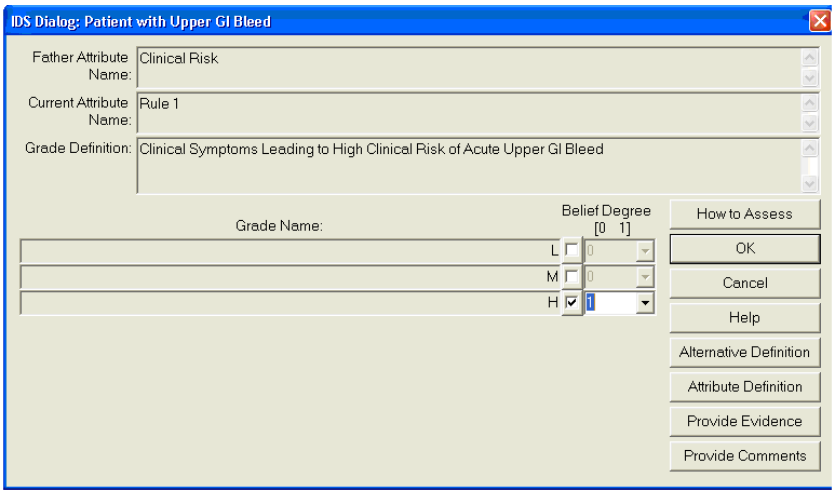


Figure 4-3: IDS Dialog for Acquiring Belief Degrees in Consequents

After the modelling work, we can run IDS to generate assessments, and a final clinical risk assessment for the simulated patient with upper GI bleed can be visually shown in Figure 4-4, which shows that the patient's clinical risk is assessed to be {(H, 0.9935), (M, 0.0065), (L, 0)}. If the severity score of H risk is set to 1, the severity score of M risk to 0.5, and the severity score of L risk to 0, the overall severity score of the simulated patient generated by the ER approach is 0.9968, and this score can be used to tell the severity difference between the patient's illness and other patients' which are also caused by upper GI bleed.

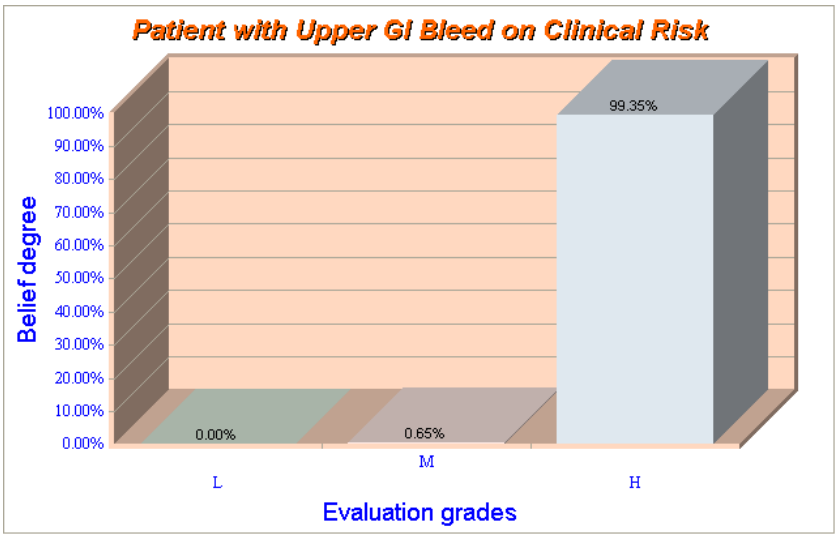


Figure 4-4: Clinical Risk Assessment for Simulated Patient with Upper GI Bleed

4.6. Feasibility Analysis

4.6.1. Advantages of Belief Rule-Based CDSS Compared To Traditional Rule-Based CDSS

After detailed presentation of inference with traditional rule base using forward chaining and inference with BRB using the ER approach for clinical risk assessment of the simulated patient with acute upper GI bleed, a comparison of assessment results for the simulated patient can be made between traditional rule-based and belief rule-based systems, and it is shown in Table 4-5.

Table 4-5: A Comparison between Traditional Rule-based and Belief Rule-based Clinical Risk Assessment

Clinical Risk Assessment Result	
Traditional Rule-based System	**Belief Rule-based System**
H	{(H, 0.9935), (M, 0.0065), (L, 0)}

It can be seen from the comparison that the conclusion reasoned from belief rule-based system are consistent with though not exactly the same as what is inferred from traditional rule-based system. For the simulated patient, his/her clinical risk

assessment result generated by the traditional rule-based system is 'H', while {(H, 0.9935), (M, 0.0065), (L, 0)} is the result reasoned from the belief rule-based system. The reason is that although there is no uncertainty represented in the belief rules as described in Table 4-3 in Section 4.3.2, a belief rule-based system can capture uncertainties in clinical signs and symptoms for the simulated patient, but traditional rule-based systems which have not taken uncertainties into consideration in system design and implementation can only reason with clinical signs and clinical symptoms with 100% certainty.

Based on the above observation, we can draw a conclusion that if there are uncertain or incomplete input data regarding patients' clinical signs and symptoms, the result generated by belief rule-based system with distributed belief degrees attached to different diagnoses is more informative than the one inferred from traditional rule-based system with one certain diagnosis. Meanwhile, if all clinical signs and symptoms can be described with 100% certainty, and there is no uncertainty in clinical rules as well, the diagnosis conclusion derived from these clinical symptoms should also be without any uncertainty. In such situations, belief rules reduce to traditional rules.

Note that clinical domain knowledge for the diagnosis of different diseases may contain different types and degrees of uncertainties, and real life clinical data are actually more complicated than the simulated patient's since information about a patient's clinical signs and symptoms may include ignorance of some symptoms, vagueness or incomplete linguistic description, inexperienced judgements and so on. In such situations, we prefer to using BRB to model uncertain clinical domain

knowledge, and using the ER approach for inference to cope with uncertainties in both clinical rules and clinical data.

Moreover, if two patients are with the same diseases and both are assessed to be at H risk, a very important question in ED would arise as to who should be treated first, which is an important issue in emergency triage system (Mackway-Jones et al., 2005). In the belief rule-based system, a recommendation can be made based on the calculated overall severity scores or severity intervals of the inferred assessments for different patients. For example, if patients P^1 and P^2 are diagnosed by the system simultaneously, and P^1 is assessed to be at {(H, 0.8), (M, 0.2), (L, 0)} risk and P^2 is assessed to be at definitely H risk. If the severity score of H risk is set to 1, the severity score of M risk 0.5, and the severity score of L 0.25, as a result, the overall severity score of patients P^1 is 0.9 and P^2's is 1. If the order of treatment in ED is based on patients' severity ranking, P^2 should be recommended to be treated earlier than P^1 by the belief rule-based system. However, for the traditional rule-based system, it is difficult to give such a recommendation based on the inferred result.

In consequence, compared to traditional rules in developing a CDSS, RIMER has the following advantages for developing a CDSS. Firstly, belief rules can provide a flexible framework to capture uncertainties in both clinical sings and symptoms and clinical domain knowledge. Secondly, inference with belief rules using ER can generate a more informative conclusion which is a combined one. Thirdly, if necessary, the distributed diagnosis recommendations can be used to rank patients' severity. Actually, a traditional rule-based system is a special case of a belief rule based system.

4.6.2. Further Discussions of the Feasibility

From the comparison study conducted above between a belief rule-based CDSS and a traditional rule-based CDSS, we can draw a conclusion that it is logically feasible to model clinical domain knowledge using BRB and to employ the ER approach for clinical inference. Further analysis of the technical feasibility of employing RIMER for developing a computerised intelligent CDSS is discussed as follows.

From a technical perspective, developing a computerised system requires that there should be appropriate computing technologies that can help to implement the system design. As for a computerised belief rule-based CDSS, it needs to consist of at least four fundamental components, namely user interfaces, database, knowledge base, and inference engine. Interfaces are used to acquire inputs, present intermediate or final conclusions and provide necessary explanations. Database is used to store and manage input information, transformed input values and kinds of reasoned results. Knowledge base consists of belief rules extracted from domain knowledge. Inference engine is built with the ER aggregation algorithm.

In the study, the integrated development environment (IDE) available for developing the CDSS is Microsoft Visual Studio .NET 2003 (Beres, 2003). With the aid of IDE, we can not only design web-based system architecture, but also design and develop each system component easily with its visual designers and a range of programming languages. Firstly, friendly web-based interfaces can be easily designed and developed by ASP.NET technology. Secondly, a database built by different types of Data-Base Management Systems (DBMSs) such as Microsoft Access, Microsoft SQL Server, Oracle and so on can be easily connected to the core programs developed in the IDE through ADO.NET technology. Thirdly in terms of the knowledge base, we

can use a unique method to store and manipulate the BRB in a relational database, and it can help reduce the complexity of developing a rule compiler. Fourthly the inference engine can be implemented by programming with languages such as Visual Basic .NET, Visual C++ .NET, Visual C# .NET and so on that are seamlessly integrated in the IDE.

To conclude, it is feasible to employ the RIMER methodology to develop a computerised intelligent CDSS.

4.7. Summary

This chapter describes how to employ the new belief rule inference methodology - RIMER for developing a CDSS, together with a comparison study of belief rules and traditional rules in reasoning out the clinical risk result of a simulated patient with upper GI bleeding. From the comparison study, the following conclusions can be drawn. Firstly, a belief rule-based CDSS can handle different uncertainties in both clinical domain knowledge and clinical data. Secondly, a belief rule-based CDSS can provide a distributed diagnostic recommendation which is more informative than a traditional rule-based CDSS that do not take uncertainties into consideration. Thirdly, if necessary, a severity score or severity interval can be calculated to rank the seriousness of patients' illness in a belief rule-based CDSS. In conclusion, it is feasible to employ RIMER for developing a computerised intelligent CDSS. The design and development of an online intelligent belief rule-based group CDSS is presented in the next chapter.

Chapter 5

Design and Development of

An Online Belief Rule-based Group CDSS Prototype

5.1. Introduction

The preliminary feasibility study presented in Chapter 4 proves that it is feasible to develop a computerised intelligent belief rule-based CDSS. What follows the feasibility study is an implementation of an intelligent belief rule-based CDSS. This chapter describes the design and development of the belief rule-based CDSS prototype.

A new CDSS framework which integrates automatic knowledge learning functionality and online group decision supporting functionality into knowledge-based CDSS has been proposed and employed in the prototype design and development. The developed CDSS prototype helps to bridge the research gaps in the CDSSs literature as described in Chapter 2. Main system features of the prototype CDSS are discussed as follows. Firstly, the prototype has two special functions, namely representation of uncertain clinical domain knowledge using belief rules, and inference with belief rule base (BRB) using the evidential reasoning (ER) approach. The functions enable the prototype to handle uncertainties existing in both clinical signs and symptoms, and clinical domain knowledge. Secondly, apart from providing individual diagnosis support, a group discussion platform and an ER-based group preferences aggregation mechanism are developed for supporting group clinical decision making. Thirdly, a BRB training module is developed and integrated into the prototype, and it enables

the system to automatically update clinical rules in the BRB by learning through clinical cases accumulated in clinical practice. Fourthly, the user interfaces implemented in the prototype are based on clinical guidelines, and the guideline-based information flow can help the system to be integrated in clinical work flow easily, while it can also facilitate system users adhering to clinical guidelines. Fifthly, the BRB is uniquely structured and stored in a relational database in the prototype. Manipulating BRB through a relational database facilitates not only the interaction between knowledge base and other core system components, but also the sharing of clinical domain knowledge between the prototype CDSS and other clinical application systems.

The methodology used for developing the CDSS prototype is prototyping as discussed in Chapter 3. Accordingly, we developed the system in an iterative way. Initially, we developed and presented a preliminary prototype CDSS to experts in MRI based on our system analysis elicited from first several meetings with expert clinicians in MRI. Then we improved the prototype iteratively based on the system users – expert clinicians' feedback of the prototype.

The system development environment is Visual Studio 2003 .NET (Beres, 2003) on platform Windows XP Professional. The programming languages include C#, ASP.NET (Liberty and Hurwitz, 2002), and MATLAB (http://www.mathworks.com/products/matlab/). The Data-Base Management System (DBMS) used for design and development of back-end relational database is SQL Server 2000 (Waymire and Sawtell, 2000).

The chapter is organised as follows. Section 5.2 introduces the system structure, where system architecture design and system component design are discussed

respectively in Section 5.2.1 and Section 5.2.2. Section 5.3 presents detailed design and development of core system components, where inference engine is discussed in Section 5.3.1, group decision supporting module in Section 5.3.2, training module in Section 5.3.3, web-based user interfaces in Section 5.3.4, database in Section 5.3.5, and knowledge base is discussed in Section 5.3.6. Conclusions about the prototype system are provided in Section 5.4. Finally Section 5.5 summarises the chapter.

5.2. System Structure

5.2.1. Architecture Design

World Wide Web (WWW) technologies (Berners-Lee et al., 1994) have transformed the design, development, implementation and deployment of decision support systems (DSSs), and great progress has been made in web-based DSSs in the past decade (Bhargava et al., 2007). Taking advantages of web technologies, a web-based DSS can link multiple decision makers who might be separated in space or time for online group discussion or meeting, and can deliver the suggestions or recommendations generated from the system to a much broader audience of decision makers who is geographically separated (Bhargava et al., 2007). As to web-based DSSs in clinical area, web-based CDSSs have advantages in providing easy accessibility for clinicians in geographically different places and easy dissemination of clinical domain knowledge and patients' clinical data among different clinical application systems. In the research, through regular meetings with expert clinicians in MRI, we know that frontier clinicians have a strong need of an online intelligent CDSS which can help them to act in accordance with practice guidelines in their daily clinical work flow. Motivated by the above facts, we adopt a web-based three-layer client-server architecture (Sommerville, 2007) in the prototype design.

In a client-server architecture, an application is modelled as a set of services that are provided by servers and used by a set of clients (Orfali and Harkey, 1998). In this architecture, clients need to be aware of the servers that are available but usually do not know the existence of other clients, and clients and servers are separate processes. In design of client-server systems, logical structure of the application that is being developed should be reflected in the system architecture (Sommerville, 2007). Usually, an application can be structured into three layers: the presentation layer which is concerned with presenting information to the user and with all user interaction; the application processing layer which is concerned with implementing the logic of the application; and the data management layer which is concerned with all database operations. Figure 5.1 illustrates these three layers.

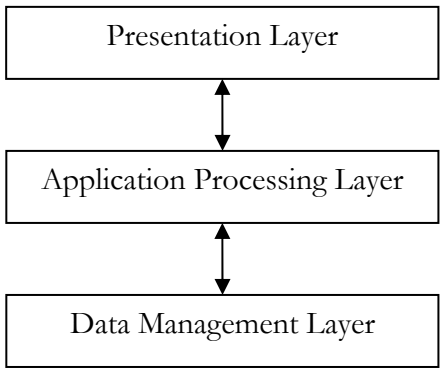


Figure 5.1: Application Layers (Sommerville, 2007)

In the three-layer client-server architecture, the presentation layer, the application processing layer and the data management layer are logically separate processes that execute on different processors (Sommerville, 2007). Generally, the three-layer client-server architecture is composed of three logical parts: system user's own computer with a web browser that can display system's user interfaces is the presentation layer; a web server for providing all services related to the application being developed is the application processing layer; and a back-end database for providing data

management services is the data layer (Sommerville, 2007). The three-layer architecture can be illustrated in Figure 5-2.

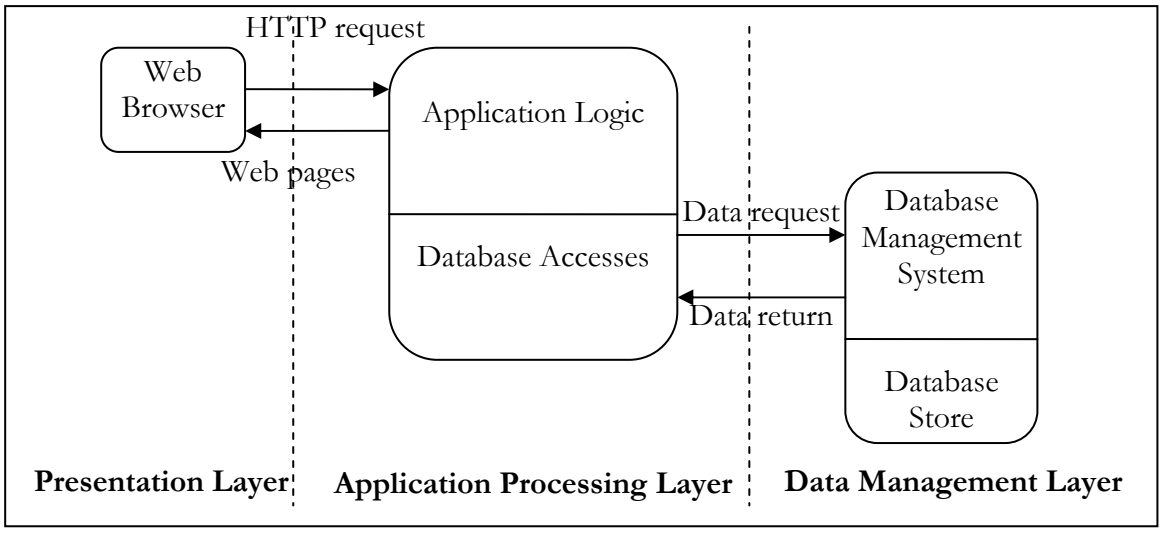


Figure 5-2: Three-Layer Architecture

The prototype CDSS is designed on the basis of the three-layer architecture. In the prototype, system users can access the system through web-based user interfaces, application logic of the system reside in the middle layer which is usually implemented in a web server, and data access technologies such as ADO.NET (Hamilton and MacDonald, 2003) can be used by system components located in the web server layer to communicate with the data management layer, which is usually implemented by a back-end database server directly.

5.2.2. Component Design

As for the system components implemented in the three-layer architecture, there should be generally at least four system components in a web-based CDSS based on the general structure of knowledge-based CDSSs as shown in Figure 2-1 of Chapter 2.

They include friendly web-based user interfaces, inference engine, knowledge base, and database. Both inputs and outputs of the system can be stored in the database.

To address the research gaps identified from the CDSS literature discussed in Chapter 2, the target belief rule-based CDSS should possess at least three capabilities. The first one is the capability of representing and reasoning with uncertain clinical domain knowledge and clinical data. The second one is the capability of providing group decision support. The third one is training or fine-tuning BRB by learning through accumulated clinical cases.

Thus, a new CDSS framework which integrates automatic knowledge learning functionality and online group decision supporting functionality into knowledge-based CDSS is proposed and employed for the prototype design and development. In the new knowledge-based CDSS framework, core components of the prototype system should include inference engine, group decision supporting module, knowledge base training module, database, knowledge base, and web-based user interfaces.

- Inference engine is for matching the system users' clinical inputs with clinical rules in the knowledge base to generate automatic diagnostic recommendations.

- Group decision supporting module provides two important system functionalities. The first one is providing a discussion forum for group clinicians from geographically different places to hold online clinical group discussions and offer individual diagnosis preferences. The second one is providing an ER based diagnosis preferences aggregation mechanism to combine group diagnosis preferences.

- Knowledge base training module is for training BRB by optimizing the knowledge representing parameters of the BRB through accumulated clinical cases.

- Database is used for storage and retrieval of system's input data, some intermediate data, and system output data.

- Knowledge base is for maintaining all clinical rules used in the system, and it is modelled as BRB in the prototype. Specifically, BRB is uniquely designed to be stored and manipulated in the back-end relational database in the prototype.

- Finally, friendly web-based user interfaces are for acquiring system users' inputs and displaying the system's outputs.

The following Figure 5-3 illustrates the actual implementation of the above mentioned core system components in the prototype.

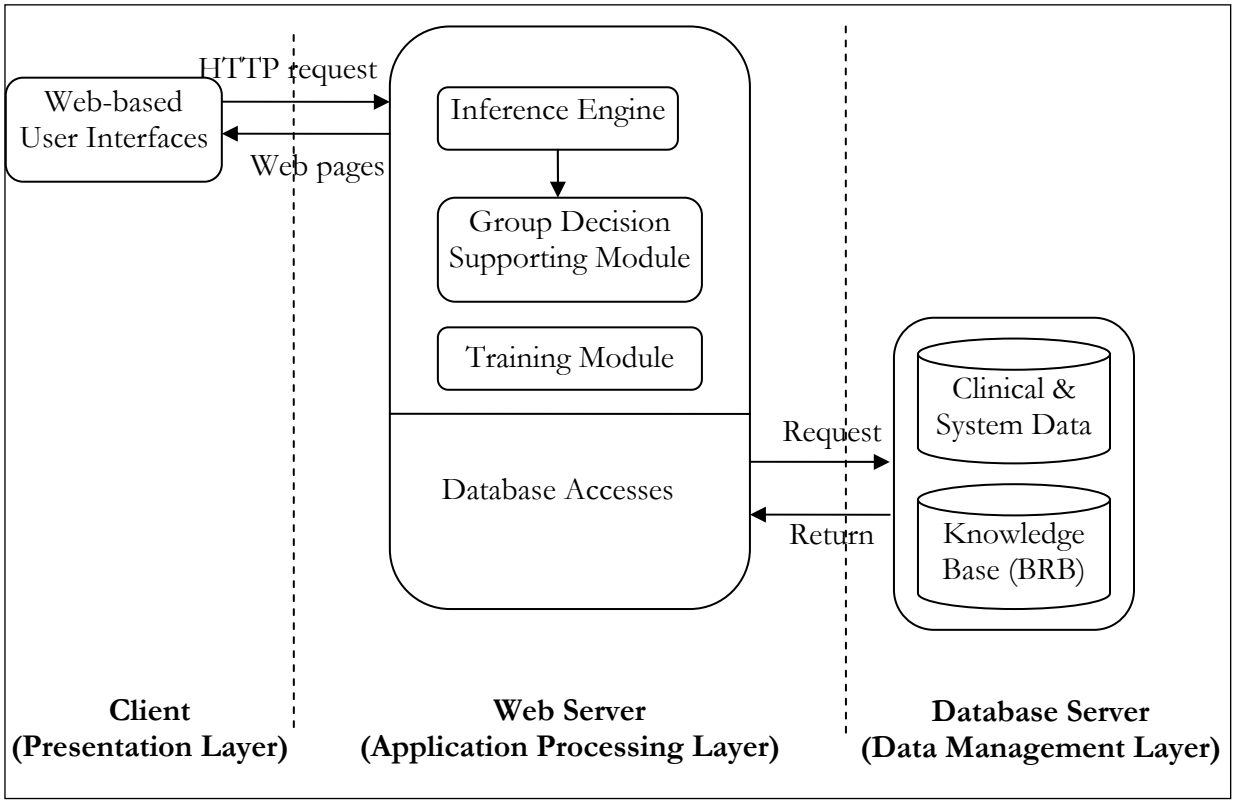


Figure 5-3: Core System Components Implemented in Three-Layer Web Architecture

The system structure as shown in Figure 5-3 has three characteristics. Firstly, thanks to advanced computing technologies, both inference engine and training module can be developed independently of system environment or application domain, so that they can be portable and adaptable to various application areas and different system development or running environments. Secondly, domain specific knowledge modelled as BRB is structured and stored in back-end relational database, which is the same as or separate from the database used for various system data and patients' data storage. Thanks to mature technologies of today's relational DBMS in data analysis and communication with main systems developed with different programming languages, structuring clinical domain BRB in a relational database can facilitate the sharing of domain knowledge, and the interactions between the knowledge base and

other system components including inference engine, training module, and group decision supporting module. Thirdly, the core algorithm of inference engine and group decision supporting module is the evidential reasoning (ER) approach, therefore the computerised ER model which is implemented in inference engine can be reused by the group decision supporting module for group diagnosis preferences aggregation.

The design and development of above discussed core system components in the prototype are presented in the following Section.

5.3. System Components

In this Section, we discuss components implemented in web server layer one by one first, and then we discuss components implemented in the client layer and back-end layer.

5.3.1. Inference Engine

The purpose of inference engine in a knowledge-based CDSS is trying to generate a reasonable clinical decision or recommendation by matching system's input data with domain specific knowledge modelled in the knowledge base. As such, an inference process is an interaction between system's inputs and the knowledge base, and the interaction way is determined by the employed inference algorithm.

The inference engine in the belief rule-based CDSS prototype is implemented using the recursive ER algorithm as described in the RIMER methodology (Yang et al., 2006). The ER based inference engine is actually a non-linear function between a set of parameters and another set of parameters. The former set of parameters include all belief rules' activation weights and consequent belief degrees in one BRB, where rule

activation weights are determined by rule weights, antecedent attribute weights, and matching degrees between system's inputs and rules' antecedents. The latter set of parameters are generated by the non-linear function provided by the ER algorithm, and they represent final belief degrees associated to all possible consequents in the BRB after combining all belief rules activated by the system inputs. The recursive ER algorithm implemented in the inference engine can be described by the flowchart in Figure 5-4.

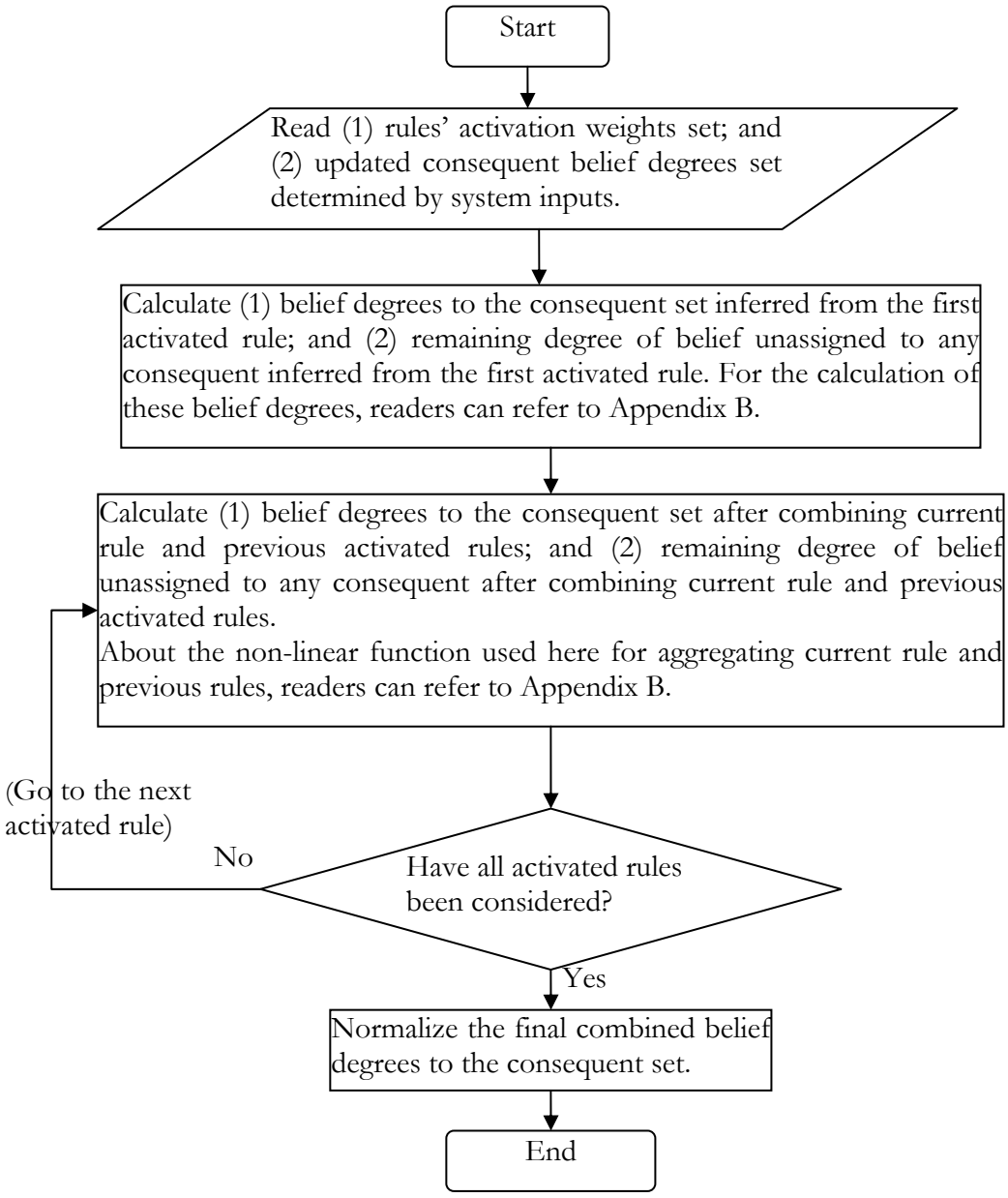


Figure 5-4: The Recursive ER Algorithm (Yang et al., 2006)

There are two other fundamental sub-components that were developed in parallel in the implementation of the inference engine, apart from the core computerised ER algorithm as described in Figure 5-4. The first sub-component of the inference engine is *input transformation sub-component*, and the second one is *rule matching sub-component*.

- Input transformation sub-component

As for the input transformation sub-component, the BRB uses sets of referential values to describe antecedent clinical signs and symptoms in its clinical rules, and the input clinical data about one clinical sign or symptom is a value in qualitative or quantitative or a mixed nature. So there is a demand for transforming the input clinical data to sets of referential values with belief degrees so that the transformed data can be used by the inference engine to do matching with clinical rules in the BRB. Accordingly, the purpose of the input transformation sub-component is to transform qualitative or quantitative or mixed clinical inputs to a set of data that can well represent uncertainties and can be used by the inference engine to do inference with the BRB. The techniques used for the input clinical data transformation are rule-based and can be found in Yang (2001). The details of input transformation have been described in Section 4.5.2.1 of Chapter 4.

- Rule matching sub-component

In terms of the rule matching sub-component, its purpose is to tell the inference engine which rules in the BRB are activated and to what degrees by doing matching between transformed input clinical data and belief rules in the BRB. As can be seen from the ER algorithm outlined in Figure 5-4, if we want to make the core inference algorithm portable and sharable for different kinds of domain application with

different input data and different BRB, we will need to separate it from calculating rules' activation weights and updated consequent belief degrees, so that it can be freely called by the inference engine independent of clinical domain BRB and input clinical data. For this purpose, the rule matching sub-component is designed and developed to complete the following tasks. Firstly, do matching recognition between the transformed system inputs and clinical rules in the BRB. Secondly, calculate activation weights for all rules using equation (3-4) and updated consequent belief degrees using equation (3-5) as described in Chapter 3 based on system inputs. Finally, feed the rule activation weights and updated consequent belief degrees into the core inference algorithm. When the computerised inference algorithm get the data from the rule matching sub-component, all activated rules will be aggregated by the ER algorithm and a combined belief degree set assigned to the consequent set can be automatically generated.

The sub-component structure of the inference engine implemented in the prototype is sketched in Figure 5-5, where the core ER algorithm is implemented as dynamic link library (DLL) by programming with Visual Basic .NET, and the other two sub-components are implemented by programming with C#. The input to the inference engine is clinical data about a patient, and the output of the inference engine is an inferred diagnosis recommendation for the patient based on the input clinical data and the embedded BRB in the system.

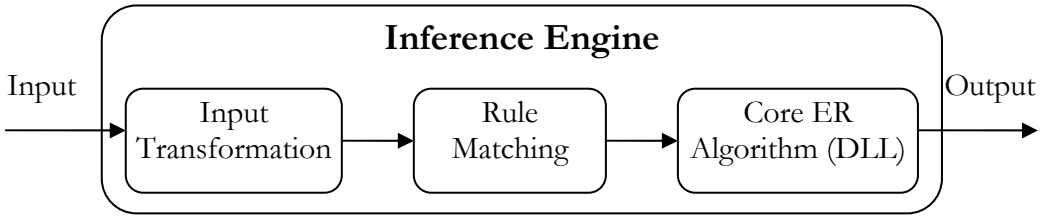


Figure 5-5: Sub-Component Structure of Inference Engine

5.3.2. Group Decision Supporting Module

The group decision supporting module in the prototype is designed to serve two aims.

- One is for providing an online discussion forum for group clinicians or consultants to offer different diagnosis opinions and shape individual diagnosis preferences.

- The other one is for aggregating all group consultants' diagnosis preferences via ER to arrive at a group combined diagnosis recommendation for target patient.

Moreover, to facilitate online group consultants' discussion, both the domain knowledge stored inside the system and those Internet-based domain knowledge resources are designed to be integrated seamlessly with the group module. Thus the group consultation participators can have access to the domain specific clinical rules stored in the system together with the Internet-based resources such as (http://www.bestbets.org/) which can aid the consultation. Details of the implemented group module are presented as follows.

The way that the group decision supporting module works is similar to the real life group consultation in clinical practice. In real life clinical group consultation, there is usually a group facilitator who helps to invite other consultants to participate in the group consultation and facilitates the whole group discussion process, while the group facilitator should have knowledge about all participated group consultants' expertise in advance.

Therefore we designed two types of system user roles in the prototype system, namely group facilitator and consultant, which can login to the group decision supporting

module and use the functionalities that the module provides. What follows are discussion about the above mentioned two system user roles.

- Group facilitator role

The group facilitator role is for initializing and facilitating an online group consultation. The system user acting as a group facilitator should inform participated consultants of key information about target patient before the group consultation, so that a participated consultant can use the key information to identify target consultation group. The main user rights assigned to the group facilitator role include:

(a) inviting group consultants to participate in the group consultation;

(b) having access to target patient's data and various domain knowledge resources;

(c) facilitating the group consultation;

(d) assigning weights to participated consultants based on their expertise;

(e) calling the ER-based aggregation mechanism to combine all consultants' diagnosis preferences.

- Consultant role

The consultant role is for participating in a specific group consultation and providing individual diagnosis preference for target patient. The main user rights assigned to the consultant role include:

(a) joining the online group consultation about one specific patient;

(b) having access to target patient's data and various domain knowledge resources;

(c) providing individual diagnosis preference.

The working flow of each user role acting in an online group consultation supported by the group module is illustrated in Figure 5-6.

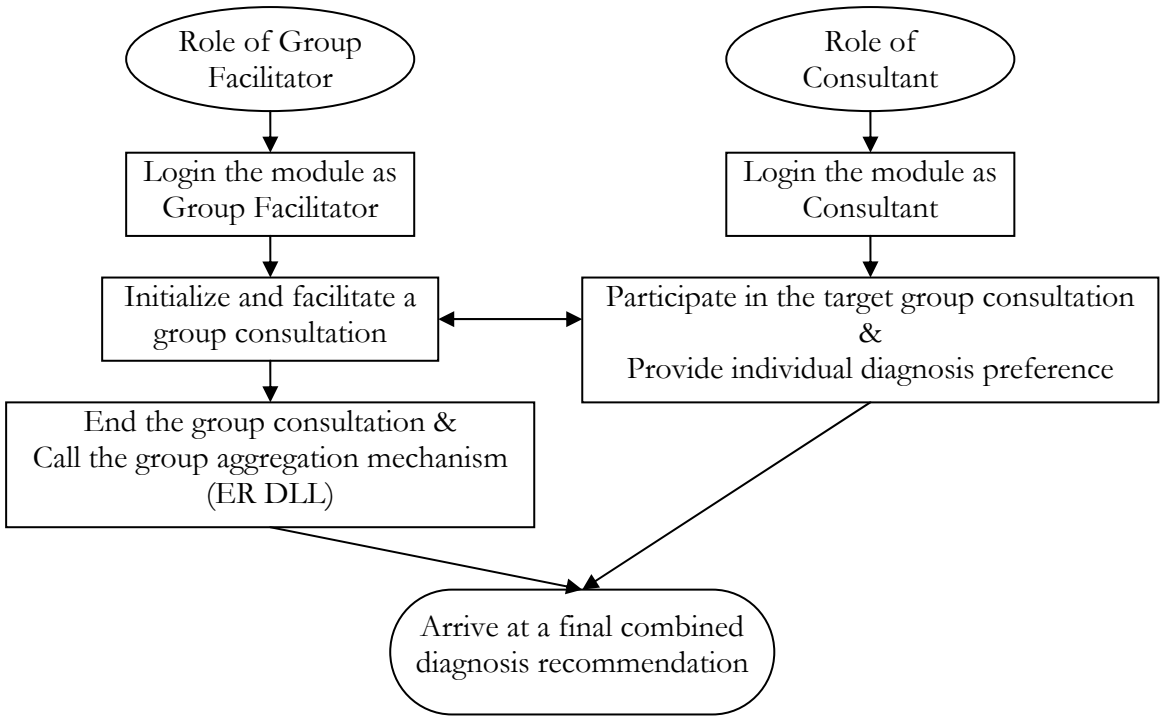


Figure 5-6: Working Flow in Online Group Clinical Consultation

The group decision supporting module based on the ER approach has two main characteristics. First, the ER based group aggregation mechanism can be implemented using the ER DLL that is already developed in the inference engine as discussed in Section 5.3.1. Reusing the ER DLL helps to reduce the complexity in developing the group module, and the ER based group diagnosis preferences aggregation mechanism can be integrated with the discussion forum seamlessly by the DLL technology. Secondly, the capability of the group decision supporting module to be linked directly to target patient's data, the BRB, and Internet resources enables group consultants to easily access target patient's data and various domain knowledge in the consultation process to shape their diagnosis preferences.

The discussion platform implemented in the group module is implemented by programming with ASP.NET and C#. The web-based user interfaces of the group module can guide system users, who act as different user roles, to different work flows in the process of holding an online group consultation and arriving at a final aggregated diagnosis recommendation.

5.3.3. Training Module

Assuming there is a BRB containing L belief rules, T antecedent attributes, and N possible consequents, the parameters of the BRB including rule weights $\theta_k \left(k = 1, \cdots, L\right)$, antecedent attribute weights $\delta_i \left(i = 1, \cdots, T\right)$, and consequent belief degrees $\beta_{jk} \left(j = 1, \cdots, N; k = 1, \cdots, L\right)$ can be originally given by domain experts or generated randomly by systems. However, it is difficult to accurately determine rule weights, antecedent attribute weights, and consequent belief degrees entirely subjectively or randomly (Yang et al., 2007). As such, there is a need to fine-tune or train belief rules originally constructed in a belief rule-based system by accumulated historical data. The training module implemented in the prototype CDSS serves the purpose of training the BRB by learning through accumulated clinical cases.

The core of the training module is a BRB optimization model. We used MATLAB to develop and implement the BRB training model. Assume there are M set of training samples, the mechanism of the training model can be described by Figure 5-7.

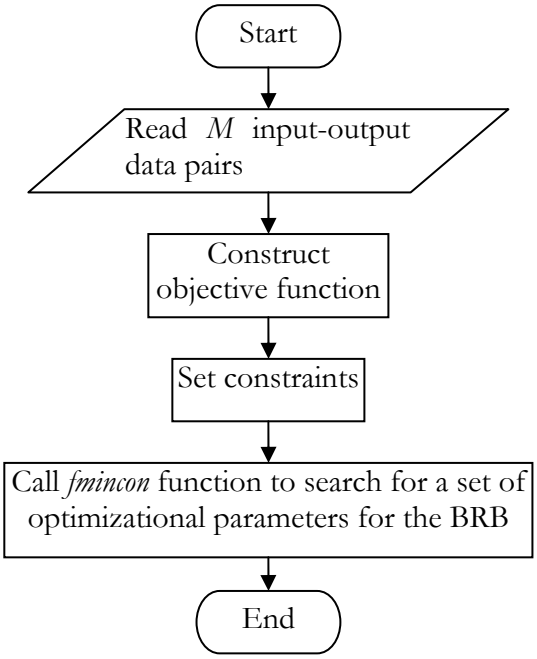


Figure 5-7: Mechanism of BRB Training Model

As described in Figure 5-7, core tasks of the BRB training model include (a) constructing objective function; (b) setting constraints for the training parameters; and (c) calling *fmincon* to search for optimizational parameter set.

In the training module implementation, the training parameter set and constraints for those parameters can be flexibly set by system administrators as required.

Apart from the core training model developed in MATLAB environment, a *parameter transferring sub-component* is developed and implemented in Visual Studio 2003 .NET environment. Its functions include: (1) retrieving training data and initial values of the training parameter set from the back-end database; (2) feeding the initial values of training parameters into the training model; (3) getting the trained parameters from the training model; and (4) putting the trained parameters set into the back-end database.

The implemented sub-component structure of the training module is described in Figure 5-8, where the BRB training model is developed and implemented in MATLAB and packaged as component object model (COM) that can be integrated with the parameter transferring sub-component developed in Visual Studio 2003 .NET seamlessly.

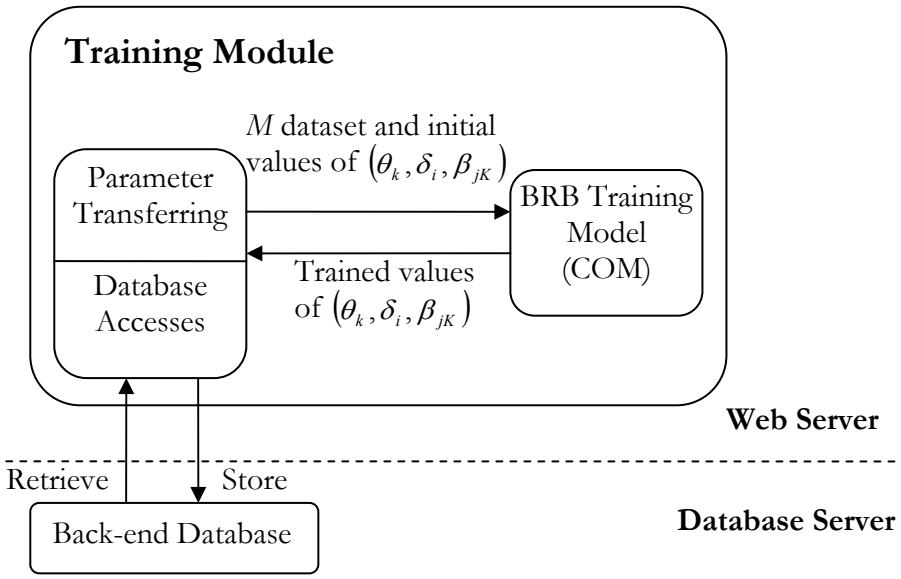


Figure 5-8: Sub-Component Structure of Training Module

The training module is integrated into the prototype CDSS seamlessly and can provide batch-mode automatic BRB training. However, before the prototype CDSS can be put into clinical use, we need to find out the most suitable training parameter set for the training module so that it can bring the most significant performance improvement to the system. Thus, in the later system validation discussed in Chapter 6, we need to compare the performances of the training model with different training parameter sets and identify the most suitable training parameter set for the training module.

5.3.4. User Interfaces

The component implemented in the prototype which concerns the interaction between users and the system is web-based user interfaces.

As identified by CDSSs researchers Sittig et al. (2008), the human-computer interface is one of top challenges in computerised clinical decision support. Human/computer interface is the main point of contact between the user and the computer system, therefore the interface should unobtrusively, but effectively, remind clinicians of things they have truly overlooked and put key pieces of data and knowledge seamlessly into the context of the work flow or clinical decision making process, so that the right clinical decisions can be made in the first place (Berner and Moss, 2005).

Clinical guidelines, as a format of clinical domain knowledge, are increasingly used to improve the quality of care by supporting clinical decision making in recent years. Guideline-based CDSSs have the potential to provide recommendations aimed at each specific patient (Peleg et al., 2003), while conventional text-based guideline can only present population-based recommendations which are aimed at a population with a specific disease. Studies (Grimshaw and Russell, 1993, Johnston et al., 1994, Lobach and Hammond, 1994, Tierney et al., 1995) have shown that computer-based CDSSs can improve clinicians' compliance with clinical guidelines and patient outcomes when developed to provide patient-specific assistances in decision making and integrated with clinical work flow. Development of guideline-based CDSSs has thus been proposed as a strategy to promote the implementation of guidelines (Field and Lohr, 1992, McDonald and Overhage, 1994).

Thus, in the design of user interfaces for the prototype CDSS, we consider two factors. Firstly, in terms of interfaces for individual diagnosis, the information flow embedded in the user interfaces should be the same as the related clinical guideline that clinicians use daily in their practice, so that the CDSS can help clinicians to adhere to the guideline. Secondly, the prototype CDSS should have the capability of being integrated into the clinical work flow seamlessly, so that a right clinical decision can be made in the right place at the right time. By using the prototyping methodology to develop the CDSS prototype, we developed and improved the user interfaces iteratively based on frontier clinicians' feedback about the prototype.

In the following, the information flow of all the user interfaces implemented in the system will be discussed, and for illustration, the user interfaces for individual diagnosis will be described using diagnosis of upper Gastrointestinal (GI) bleed as an example.

- **Information flow of user interfaces**

The information flow of the implemented user interfaces can be sketched in Figure 5-9 based on the system's main functionalities, which include individual diagnosis, group consultation, and automatic knowledge updating.

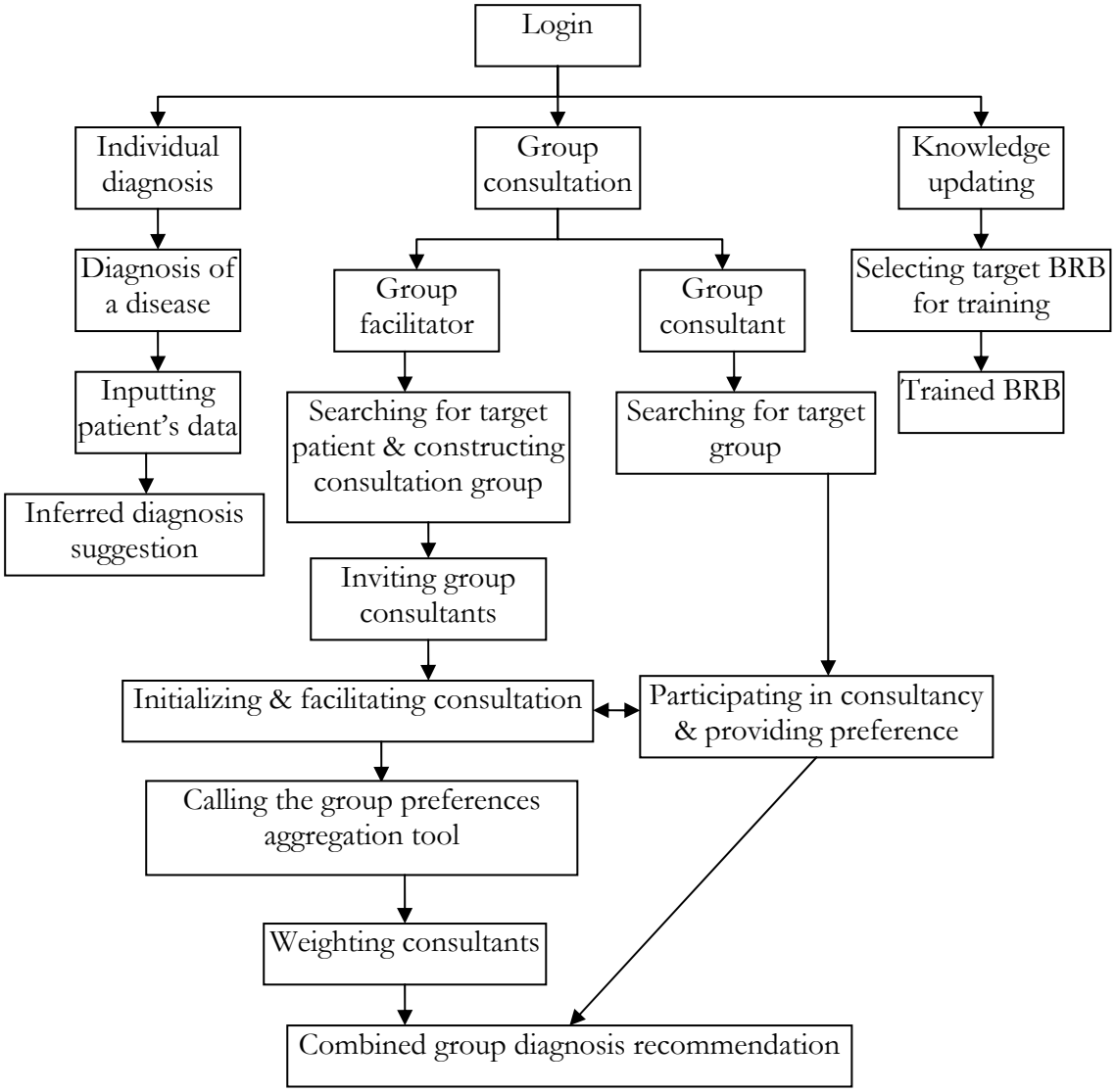


Figure 5-9: User Interface Information Flow Diagram

Note that there are different flows in the user interfaces as shown in Figure 5-9 due to different functionalities that the system can provide. The system users should specify their roles when they login to the system so as to navigate through different information flow. The system user roles include: (1) clinician for individual diagnosis; (2) clinicians for group consultation including group facilitator and consultant; and (3) clinician for updating clinical rules in the BRB. For illustration, user interfaces for individual diagnosis are described as follows.

- **User interfaces for individual diagnosis**

Interfaces for individual diagnosis of one disease were designed based on the guideline to facilitate system users to adhere to clinical guidelines. Take diagnosis of upper GI bleed as an example, main user interfaces for diagnosis and treatment of upper GI bleed can be illustrated as in Figure 5-10.

As can be seen from the interfaces in Figure 5-10, a clinical work flow for diagnosis and management of patients with upper GI bleed in ED is provided in the main diagnosis interface as shown in Figure 5-10(B) after a clinician logs into the system through the login interface as shown in Figure 5-10(A). There are several links in the main diagnosis interface linking to different web forms for acquiring different types of patients' data. For example, '*Please input patient's personal information here*' in the main diagnosis interface is linked to the web form as shown by Figure 5-10(C) for acquiring a patient's personal data. Moreover, the interface can guide the clinician to go back to the main diagnosis interface when necessary data has been input. Then the clinician can proceed to the next diagnosis or treatment step as indicated by the clinical work flow in the main interface. The interface in Figure 5-10(D) is for acquiring a patient's necessary clinical data for risk assessment, and from the interface, the inference engine can be triggered automatically to do inference with the input clinical data, and then another interface as shown by Figure 5-10(E) for showing the inferred result of the patient can be displayed automatically. Note that all these interfaces designed for acquiring patients' data or displaying inferred results have 'Back' links which can guide clinicians back to the main diagnosis interface and to proceed to the next step.

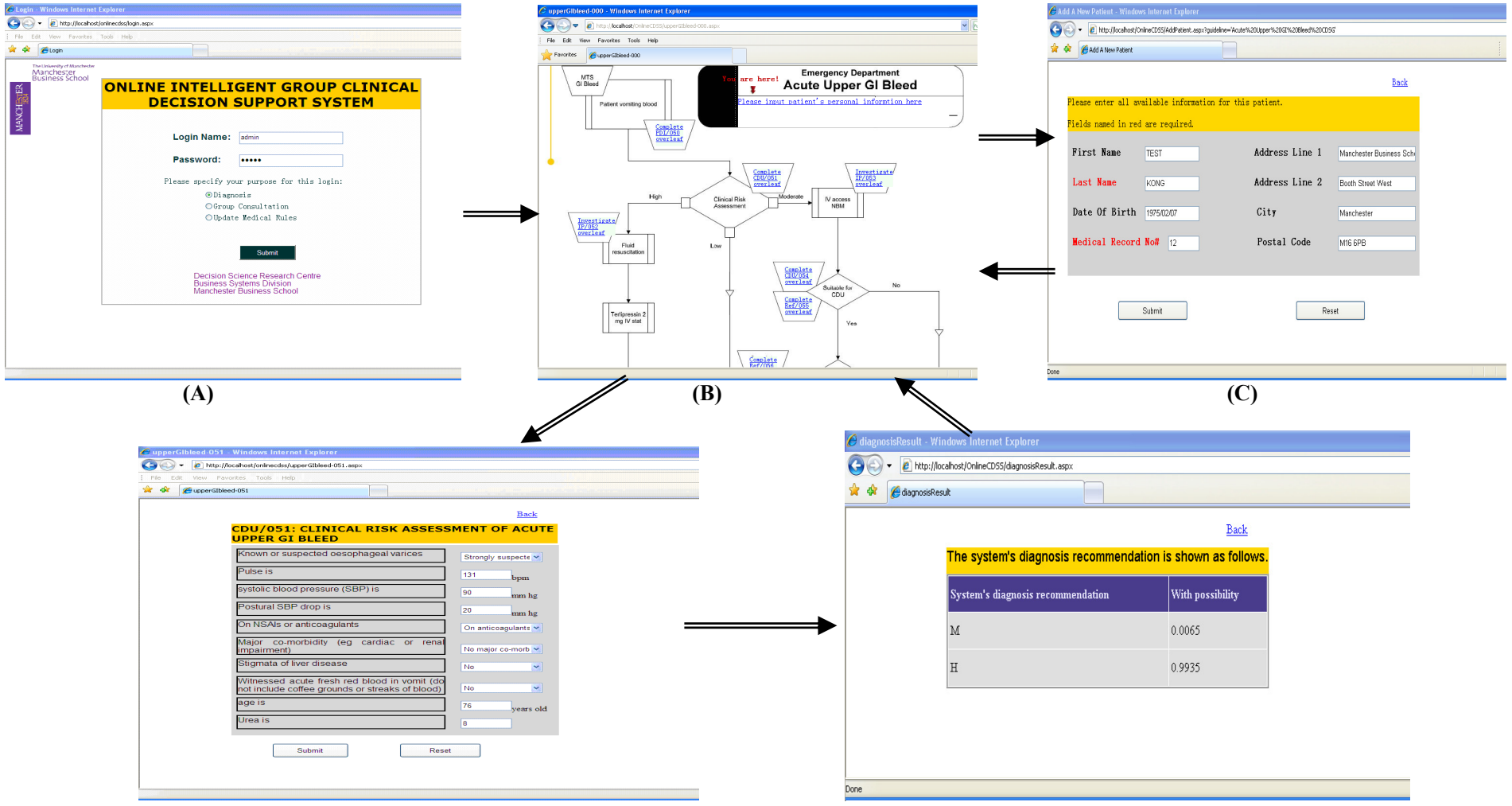


Figure 5-10: Main User Interfaces for Individual Diagnosis of Upper GI Bleed

About the user interfaces designed for group consultation and knowledge updating in the system, readers can refer to a brief user manual in Appendix A.

Thus far we have presented all core system components implemented in the web server layer and client layer. All these components can not work independently from data layer, which is used to store and manage patients' clinical data, system data, and clinical domain knowledge used in the prototype CDSS. What follows is a discussion about the database and the knowledge base implemented in the back-end relational Data-Base Management System (DBMS).

5.3.5. Database

Generally, there are three phases to the database development process, namely logical design, physical implementation, and application development (Hernandez, 2003). Logical design involves determining and defining the structure of the database, which includes tables and their fields, primary and foreign keys of tables, table relationships, and so on. Physical implementation involves using proper DBMS software to implement the database structure we created in logical design phase. Application development involves creating an application that allows system users to interact with the data stored in the database.

As the main application programs developed in the prototype CDSS have been discussed from Section 5.3.1 to Section 5.3.4, we will briefly describe logical design and physical implementation of the database in this Section.

- **Logical design**

In designing the database, we used the database design method introduced by Hernandez (2003). Using the design method, there are seven steps involved in designing the database, namely (1) identify purpose of the database; (2) analyze the data used in clinical environment, and identify *Subjects* that the database needs to keep track of for supporting clinical decision making and *Characteristics* of those subjects; (3) create the data structures, and this step involves establishing *tables* based on identified subjects and associating each table with *fields* that represent characteristics of the table's subject; (4) determine table *relationships*; (5) define rules to set constraints to the data stored in the database; (6) establish *views* to facilitate manipulation of data stored in the database; and (7) review data integrity of the designed database structure.

For clarification, here we give a brief discussion of terms used in this Section, and the term definition is based on Hernandez (2003). *Subjects* represent objects such as persons, places, or events that occur at a given point in time. *Characteristics* represent details of one subject. Take the subject of patient as an example, the patient's first name, last name, age, gender and so on are characteristics of one patient. *Tables* are the chief structures in a relational database, and each table represents a single, specific subject. *Fields* are the structures that actually store data in the database and a field represents a characteristic of the subject represented by the table to which it belongs. A *relationship* exists between two tables if we can in some way associate the *records* of the first table with those of the second. Here a *record* represents a unique instance of the subject of a table. A *view* is a virtual table composed of fields from one or more tables in the database.

In the following discussion, we will not show details of the whole logical design process, we will however briefly discuss first the purpose of the database, and then the subjects together with their characteristics that we identified by analyzing data requirement for the prototype CDSS, and finally the tables we created for the database according to those identified subjects and characteristics.

(1) Purpose of the database

The purpose of back-end database in a CDSS is for the maintenance of various data used by or generated from system components reside in the server layer or client layer as shown in Figure 5-3. For example, inference engine needs patients' clinical data and clinical rules in the BRB for reasoning about a patient's clinical risk level or disease status, while group decision supporting module needs not only patients' data, but also group consultants' data to help a group facilitator organise a group consultation.

(2) Identified subjects and characteristics

To identify subjects and their characteristics that the database need to keep track of, we need to analyze the data requirement for the prototype CDSS first.

As for knowledge modelled as BRB in the system, logical design and physical implementation of BRB in a relational database will be discussed in Section 5.3.6. In this Section, we focus on discussion about other data used by or generated from the system.

Besides domain knowledge modelled as BRB, other necessary data used by or generated from the prototype can be classified into three categories: (a) patients' data, (b) doctors' data, and (c) data generated from group consultation which includes

group discussion content, group consultants' individual diagnosis preferences, and group combined diagnosis suggestion for one patient.

(i) Subjects and characteristics in patients' data

As for a patient's data stored in the prototype, it should contain (a) the patient's key personal data, (b) the patient's clinical data including his/her clinical signs or symptoms, (c) the clinical guideline used by doctors for his/her diagnosis, and (d) system generated diagnosis recommendation about the patient's disease.

We first identified "Patient" as one subject, and its characteristics include (1) key personal information; (2) clinical data; (3) clinical guideline used for his/her diagnosis; and (4) diagnosis recommendation generated by the system or the doctor. We then further identified subjects from these characteristics which have their own characteristics. Thus we treat characteristics of the subject "Patient" including "Clinical sign or symptom", "Clinical guideline", and "Diagnosis recommendation" as subjects. Furthermore, we take "Severity level" which is a characteristic of "Clinical sign or symptom" and "Diagnosis recommendation" as a separated subject.

Identified subjects and characteristics from patients' data can be listed as in Table 5-1.

Table 5-1: Identified Subjects and Characteristics from Patients' Data

Subjects	Characteristics
Patient	(1) Key personal information including Medical Record No. (MRN), last name, first name, gender, age, etc. (2) Clinical data including his/her clinical signs or symptoms (3) Clinical guideline used for his/her diagnosis (4) Diagnosis recommendation generated by the system
Clinical sign or symptom	Name, severity level
Clinical guideline	Name
Diagnosis recommendation	Name, severity level
Severity level	Name, associated clinical sign/symptom or diagnosis recommendation

(ii) Subjects and characteristics in doctors' data

In terms of one doctor's data stored in the system, firstly, the doctor should be a system user of the system with username and password; secondly, the doctor should have his/her user role for each login; and thirdly, the doctor should have his/her expertise domain stored in the system.

We first identified "System user" as one subject, and its characteristics include (1) user name; (2) password; (3) first name; (4) last name; (5) associated user roles; and (6) expertise domains. Since characteristics of "Associated user roles" and "Expertise domains" have their own attributes, we then take these two characteristics as separate subjects.

Identified subjects and characteristics from doctors' data can be listed as in Table 5-2.

Table 5-2: Identified Subjects and Characteristics from Doctors' Data

Subjects	Characteristics
System user	(1) User name (2) Password (3) First name (4) Last name (5) Expertise domains (6) Associated user roles
User role	Name, user right
Doctor expertise	(1) Associated doctor (2) Expertise name (3) Rank order compared to the doctor's other expertises

(iii) Subjects and characteristics in data generated from group consultation

As for data generated from a group consultation, it is mainly include (a) discussion content and individual diagnosis preferences provided by group consultants; and (b) group combined diagnosis suggestion for the target patient.

Identified subjects and characteristics from group consultation data can be listed as in Table 5-3, where the subjects of "Group consultant", "Group discussion content", "Individual diagnosis preference", and "Combined diagnosis preference" are characteristics of subject "Facilitated consultation group". The characteristic of "Belief degree provided by the consultant" associated to the subject "Individual diagnosis preference" represents the belief degree assigned by one consultant to his/her diagnosis preference. For the role of this belief degree played in the group preferences aggregation process, readers can refer to Section 3.2.2 for how to use the evidential reasoning (ER) approach to aggregate group diagnosis preferences with belief degrees. The characteristic of "Aggregated belief degree" associated to the subject "Combined diagnosis preference" represents the belief degree assigned to the final diagnosis preference after aggregating all group consultants' diagnosis preferences.

Table 5-3: Identified Subjects and Characteristics from Group Consultation Data

Subjects	Characteristics
Facilitated consultation group	(1) Group facilitator (2) Group consultants (3) Consulted patient (4) Group discussion content (5) Individual diagnosis preferences provided by group consultants (6) Aggregated diagnosis preference
Group consultant	Name, weight
Group discussion content	(1) Associated consultation group (2) Associated consultant (3) Discussion content
Individual diagnosis preference	(1) Associated consultation group (2) Associated consultant (3) Diagnosis preference (4) Belief degree provided by the associated consultant
Combined diagnosis preference	(1) Associated consultation group (2) Aggregated diagnosis preference (3) Aggregated belief degree

(3) Created tables

After identifying subjects and characteristics from data used by or generated form the system, we designed tables to represent those identified subjects.

(i) Tables for patients' data

In designing tables to represent subjects in patients' data, we used table PATIENTS to represent subject "Patient" with key personal information, table DIAGNOSISITEMS to represent subjects of "Clinical sign or symptom" and "Diagnosis recommendation", table DISEASEDIAGNOSISGUIDELINES to represent subject "Clinical guideline". To make distinction among records of "Clinical sign or symptom" and "Diagnosis recommendation" or other types of clinical data, we used table DIAGNOSISITEMCATEGORIES to represent categories of clinical data.

To represent severity level related to one patient's clinical sign or symptom or diagnosis recommendation, we used table DIAGNOSISITEMEVALUATIONGRADES to represent severity level which is used to describe clinical signs or symptoms, clinical risk or disease. For table linkage, we used table PATIENTDISEASEDIAGNOSISGUIDELINES to link tables of PATIENTS and DISEASEDIAGNOSISGUIDELINES, and we used table PATIENTDIAGNOSISITEMS to link tables of PATIENTS and DIAGNOSISITEMS.

In a BRB, different severity levels associated to one antecedent clinical sign or symptom may be used as a set of referential values for it. Thus to capture matching degree of a doctor's judgement about one patient's clinical sign or symptom to different severity levels associated to the clinical sign or symptom, we used field DIAGNOSISITEMBELIEFDEGREE associated to table PATIENTDIAGNOSISITEMS to capture these matching degrees. If severity levels are not applicable to one clinical sign or

symptom, the field of DIAGNOSISITEMBELIEFDEGREE in table PATIENTDIAGNOSISITEMS can be used to capture one doctor's belief degree in his/her judgement about the clinical sign or symptom.

To represent belief degrees in the system's recommended clinical diagnosis as described in equation (3-6) of Chapter 3, the field of DIAGNOSISITEMBELIEFDEGREE in table PATIENTDIAGNOSISITEMS can also be used to capture these belief degrees.

The created tables together with their fields for representing patients' data can be listed as in Table 5-4.

Note that due to space restrictions, in the following discussion, we can not show all of the fields for a created table, we will however show the fields that are most representative of characteristics belonging to the subject that the table represents.

Table 5-4: Tables for Representing Patients' Data

Tables	Fields
Patients	PatientID, PatientFirstName, PatientMiddleName, PatientLastName, PatientGender, PatientAge
DiagnosisItemCategories	DiagnosisItemCategoryName, DiagnosisItemCategoryID, DiagnosisItemCategoryDescription
DiagnosisItems	DiagnosisItemCategoryID, DiagnosisItemName, DiagnosisItemID, DiagnosisItemDescription
DiagnosisItemEvaluationGrades	DiagnosisItemID, DiagnosisItemEvaluationGradeName, DiagnosisItemEvaluationGradeID, DiagnosisItemEvaluationGradeDescription
DiseaseDiagnosisGuidelines	DiseaseDiagnosisGuidelineName, DiseaseDiagnosisGuidelineID, DiseaseDiagnosisGuidelineDescription
PatientDiagnosisItems	PatientID, DiagnosisItemID, PatientDiagnosisItemID, DiagnosisItemEvaluationGradeID, DiagnosisItemBeliefDegree
PatientDiseaseDiagnosisGuidelines	PatientID, DiseaseDiagnosisGuidelineID, PatientDiseaseDiagnosisGuidelineID

(ii) Tables for doctors' and group consultation data

Based on the identified subjects from doctors' and group consultation data, we used table USERS to represent subject "System users" and "Group consultant", table ROLES to represent subject "User role", table DOCTOREXPERTISES to represent subject "Doctor expertise", table FACILITATEDGROUPS to represent subject "Facilitated consultation group", table GROUPDISCUSSIONS to represent subject "Group discussion content", table GROUPCONSULTANTDIAGNOSES to represent subject "Individual diagnosis preference", and table GROUPFINALDIAGNOSES to represent subject "Combined diagnosis preference".

Since the characteristic "Weight" of subject "Group consultant" is assigned by a group facilitator before he/she calls the group preferences aggregation tool to combine all group consultants' diagnosis preferences, we can not design table USERS with the field WEIGHT. However, we used table GROUPCONSULTANTWEIGHTS to represent the weights of group consultants in one consultation group, and this table has relationships with tables of USERS and FACILITATEDGROUPS.

For table linkage, we used table USERROLES to link table USERS and ROLES.

The created tables together with their fields for representing system users' and group consultation data can be listed as in Table 5-5. The field of BELIEFDEGREE associated to table GROUPCONSULTANTDIAGNOSES is for representing belief degree in one consultant's judgement about one patient's clinical status, and the field DIAGNOSISITEMBELIEFDEGREE associated to table GROUPFINALDIAGNOSES is for representing belief degree in one diagnosis preference after aggregating all group consultants' diagnosis preferences.

Table 5-5: Table for Representing Data about System Users and Group Consultation

Tables	Fields
Users	UserID, FirstName, LastName, UserName, UserPassword
Roles	RoleID, RoleDescription
UserRoles	UserID, RoleID
DoctorExpertises	DoctorID(UserID), DiagnosisItemID, ExpertiseOrder
FacilitatedGroups	GroupID, GroupDesc, FacilitatorID(UserID), PatientID
GroupDiscussions	GroupDiscussionID, GroupID, ConsulatntID(UserID), DiscussionContent
GroupConsultantDiagnoses	GroupID, ConsultantID(UserID), DiagnosisItemEvaluationGradeID, BeliefDegree
GroupFinalDiagnoses	GroupID, DiagnosisItemEvaluationGradeID, DiagnosisItemBeliefDegree
GroupConsultantWeights	GroupID, ConsultantID(UserID), ConsultantWeight

Note that we used different field names for the field of USERID in different tables. For example, in table DOCTOREXPERTISES, we used field DOCTORID to represent the characteristic of "Associated doctor", while DOCTORID is actually the field USERID used in table USERS for identifying different user records stored in the database. We added USERID in parentheses to the field which functions the same but with different name in different table as shown in Table 5-5.

- **Physical implementation**

For physical implementation of the designed tables as described above, we choose Microsoft SQL Server 2000 as the DBMS software. For illustration, the diagrams drawn by SQL Server 2000 for those implemented tables can be shown in Figure 5-11 and Figure 5-12.

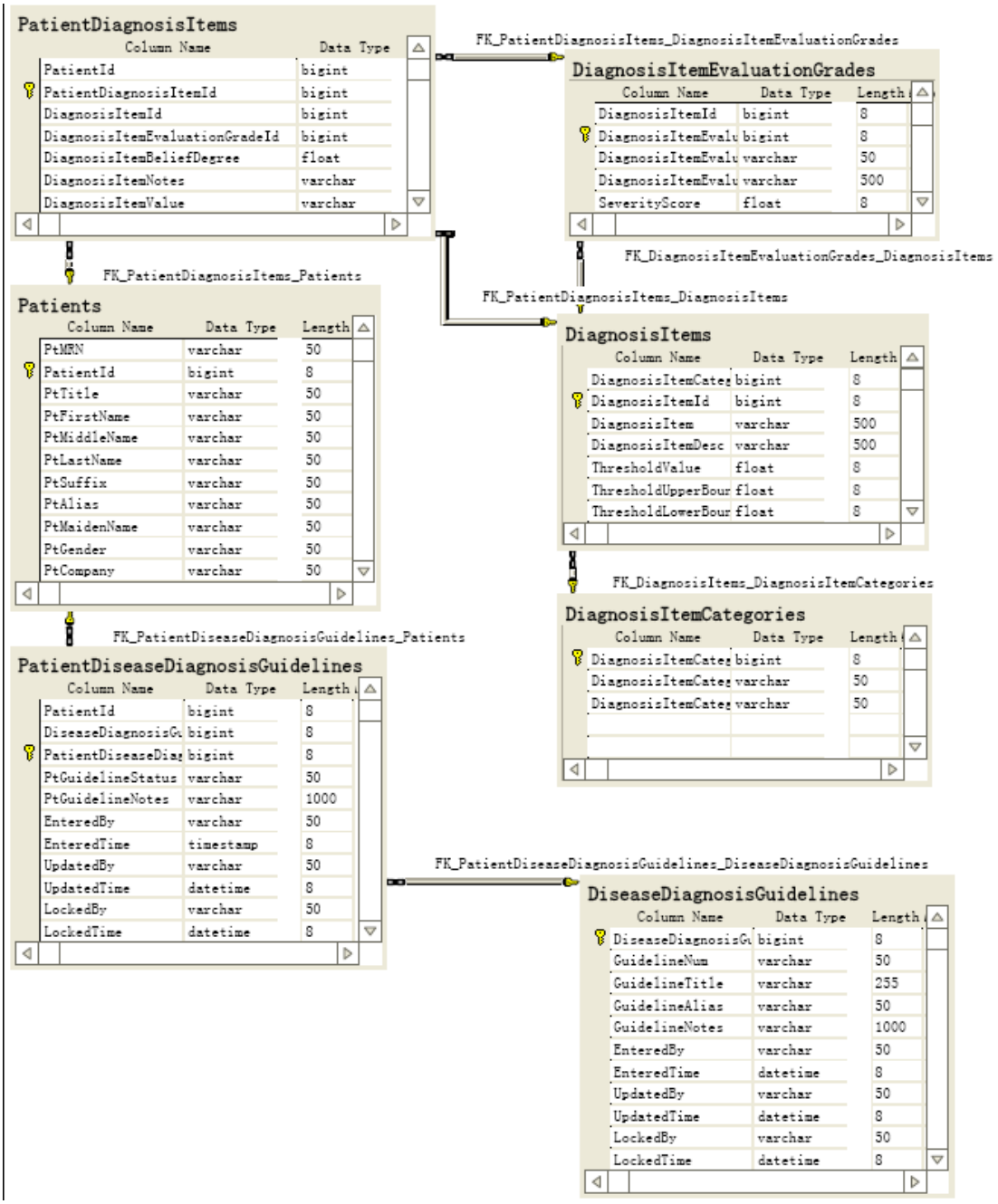


Figure 5-11: Diagram of Tables Representing Patients' Data

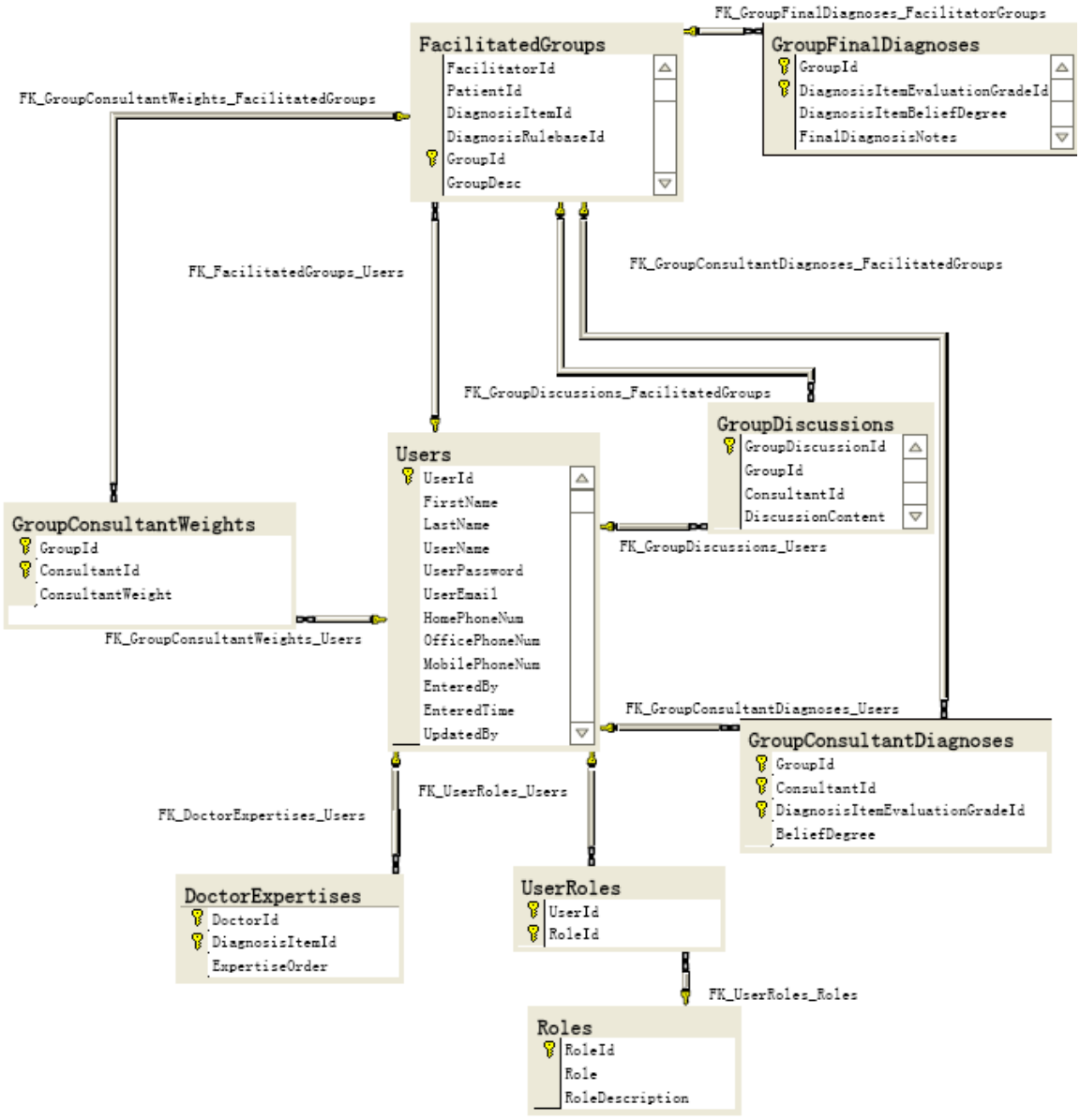


Figure 5-12: Diagram of Tables Representing System Users and Group Consultation Data

Now the logical design and physical implementation of relational models for various data except BRB used in the prototype have been discussed. The following Section 5.3.6 will present brief logical design and physical implementation of BRB in a relational database.

5.3.6. Knowledge Base

The knowledge base constructed in the prototype CDSS is based on belief rules. Traditionally, a knowledge base in a rule-based system is constructed separately from the back-end database which is used for storing system inputs and outputs. For example, some logic programming language such as Prolog (Ivan, 2001) can be applied specifically for rule representation and processing in a rule-based system, and in such a system, knowledge base is implemented as a part of the main program of the system and is separated from the fact data. EXtensible Markup Language (XML) has been recently proposed as a carrier for business rule representation, interchange, and reasoning in Web-based applications because of its easy readability and platform independent attributes (HTTP://RULEML.ORG/).

However, none of the above existent rule base implementation methods is ideal for belief rule-based systems. As to using a specific logic programming language, it will add complexity to system development if we use a specific logic programming language to represent and manipulate belief rules, because the existent logic programming languages can neither well represent belief rules nor provide ER based inference for those rules. As to XML technology, it has advantages in representing business rules because XML can provide a declarative format of rules which can be read by both rule users and computers, but XML is not a widely used technology for rule representation and inference though it has been recommended as a standard for data storage for more than one decade. Compared to relational database, firstly, XML syntax is too verbose for rule owners to design, and different rule owner may have different XML design; secondly, XML-based rules will add complexity to system developers to develop an inference engine to process XML-based rules since there is

no existent standard XML processor to do XML documents processing for applications. Readers can refer to (http://www.w3.org/XML/) for details of XML.

In the research, we propose to store and manage BRB by relational database. We will discuss the knowledge base from perspectives of logical design and physical implementation.

- **Logical design**

To design a relational model for a BRB containing L belief rules as described by equation (3-2) in Chapter 3, as usual, we analyzed BRB first to identify necessary subjects and characteristics that the knowledge base needs to keep track of, and then based on the identified subjects and characteristics, we designed table structures.

(i) Subjects and characteristics identified from BRB

Based on the description of a belief rule R_k as described by equation (3-2) in Chapter 3, the relationship between BRB, belief rule, rule antecedent, and rule consequent can be illustrated with Figure 5-13. For details of the symbols used in the figure, readers can refer to Section 3.2.1 of Chapter 3.

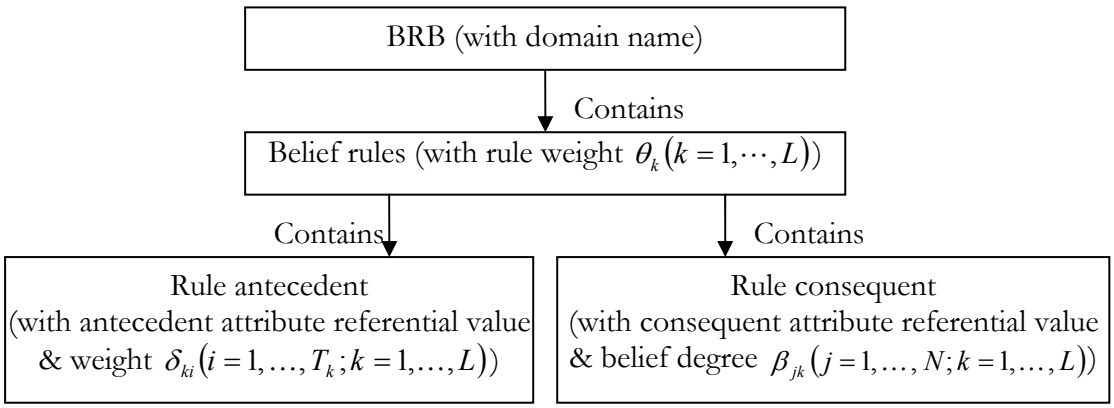


Figure 5-13: Relationship between BRB, Belief Rule, Rule Antecedent, and Rule Consequent

Usually the knowledge representation parameters in a BRB would change after training. The trained set of knowledge representation parameters should be stored in the system for domain experts' check and approval. Furthermore, to keep track of the training history, we need to store the trained knowledge representation parameters after each training round in the database.

The subjects and characteristics that can be identified from a BRB and its training are shown in Table 5-6, where subjects of "Belief rule" and "Training rounds" are characteristics of subject "BRB", subjects of "Rule antecedent" and "Rule consequent" are characteristics of subject "Belief rule", and subjects of "Trained rule weights", "Trained antecedent weights", and "Trained belief degrees" are seen as characteristics of subject "Training round".

Table 5-6: Identified Subjects and Characteristics from a BRB and Training

Subjects	Characteristics
BRB	(1) Domain name (2) Order in the inference process (3) Belief rules (4) Training rounds
Belief rule	(1) Associated BRB (2) Rule number (3) Rule weight (4) Rule antecedents (5) Rule consequents
Rule antecedent	(1) Associated belief rule (2) Antecedent attribute referential value (3) Attribute weight
Rule consequent	(1) Associated belief rule (2) Consequent attribute referential value (3) Belief degree
Training round	(1) Associated BRB (2) Trained rule weights (3) Trained antecedent weights (4) Trained consequent belief degrees
Trained rule weights	(1) Associated training round (2) Associated rule (3) Trained rule weight
Trained antecedent weights	(1) Associated training round (2) Associated BRB (3) Associated antecedent (4) Trained antecedent weight
Trained belief degrees	(1) Associated training round (2) Associated rule (3) Associated consequent (4) Trained belief degree

(ii) Created tables

Based on the identified subjects and characteristics shown in Table 5-6, we designed tables as shown in Table 5-7. As the antecedent attribute in one clinical rule is one clinical sign or symptom and the consequent attribute in the clinical rule is a disease or clinical risk, tables RULEANTECEDENTS and RULECONSEQUENTS are designed to have relationships to tables of RULES and DIAGNOSISITEMEVALUATIONGRADES, where the table DIAGNOSISITEMEVALUATIONGRADES is for representing severity level of one clinical sign or symptom or one disease as discussed in above database

design. For the table of RULEBASEANTECEDENTS, it is designed to have relationship to tables of RULEBASES and DIAGNOSISITEMS, where table DIAGNOSISITEMS is for representing clinical sign or symptom or disease status.

Table 5-7: Tables for Representing BRB

Tables	Fields
Rulebases	DiseaseDiagnosisGuidelineID, RulebaseID, RulebaseDesc, OrderID
Rules	RulebaseID, RuleID, RuleWeight
RuleAntecedents	RuleID, DiagnosisItemEvaluationGradeID
RuleConsequents	RuleID, DiagnosisItemEvaluationGradeID, ConsequentBeliefDegree
RulebaseAntecedents	RulebaseID, DiagnosisItemID, AntecedentAttributeWeight
TrainingRounds	TrainingRoundID, RulebaseID, TrainingRoundDesc
TrainedRuleWeights	TrainingRoundID, RuleID, TrainedRuleWeight
TrainedAntecedentWeights	TrainingRoundID, RulebaseID, DiagnosisItemID; TrainedAntecedentWeight
TrainedBeliefDegrees	TrainingRoundID, RuleID, DiagnosisItemEvaluationGradeID, TrainedBeliefDegree

- **Physical implementation**

We use diagram drawn by Microsoft SQL Server 2000 about designed tables for representing BRB and its training to illustrate physical implementation, and the diagram is shown in Figure 5-14.

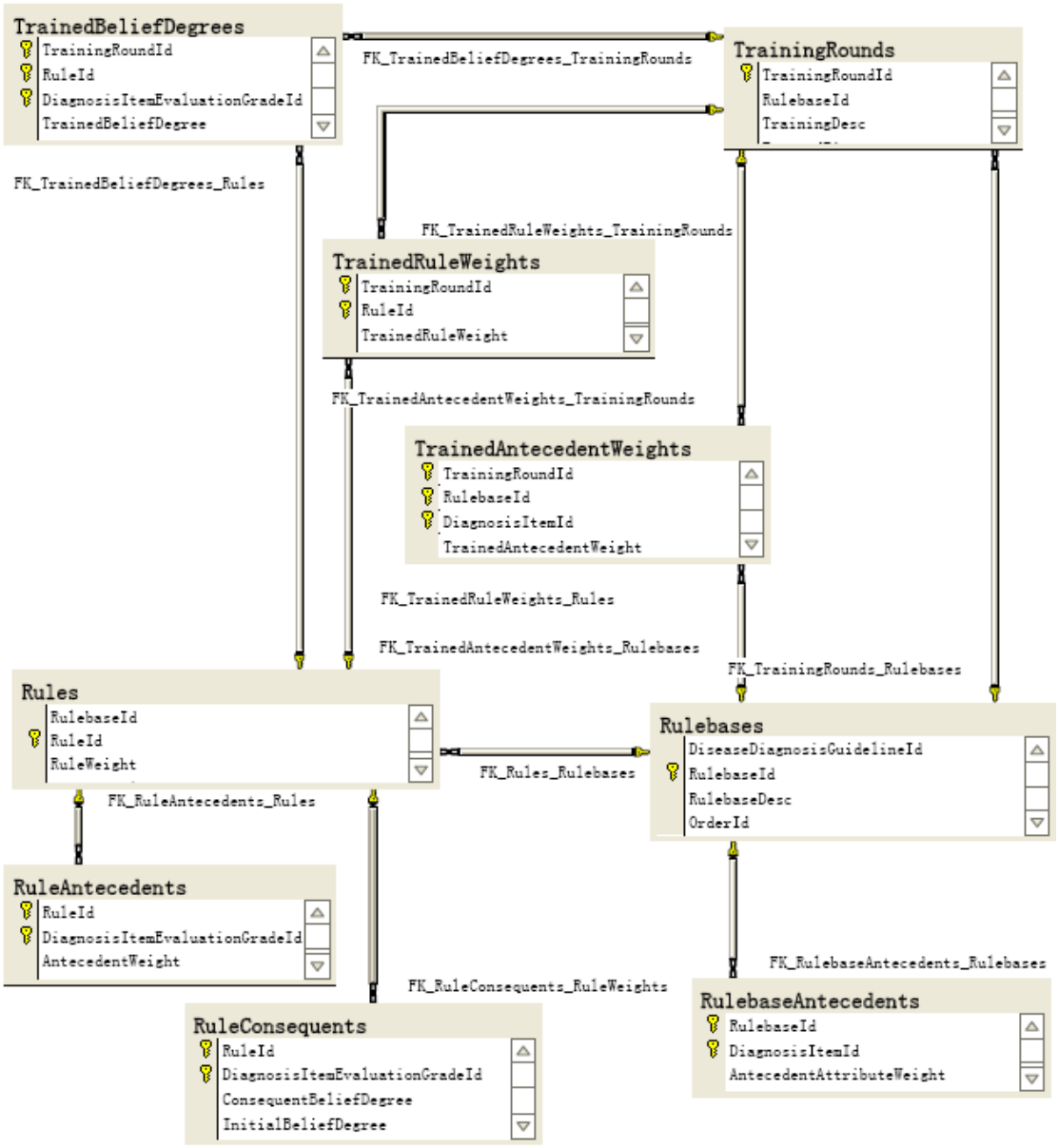


Figure 5-14: Diagram of Tables for Representing BRB

Structuring and storing BRB in relational database can meet the needs of a belief rule-based system. Firstly, it supports dynamics of BRB. The BRB structure is a dynamic one which may change after training with accumulated data. By storing BRB and its training results in relational database as discussed above, changes of BRB can be recorded in the database as well. As a result, the prototype can have a dynamic BRB which can change with different training sample. Secondly, based on the mature relational database technology, the domain specific knowledge, which is modelled by

BRB and stored in relational database, can be freely retrieved, updated, and shared by various computerised clinical systems independent of their platform. Thirdly but not the least importantly, storing BRB in relational database ease the interaction between knowledge base and other core system components thanks to mature database technology.

5.4. Discussions

The system development is really a time-consuming and demanding process. During prototype development, we have demonstrated developed system regularly to expert clinicians in MRI. Based on experts' prompt feedback, we improved the prototype until we obtained the current one which is introduced in this chapter. Experts in MRI gave positive judgements about the system, and they have strong interests in applying a mature system in their clinical practice. The prototype seems to possess six main system features.

(1) Web-based system architecture enables the system to be accessed online from geographically different places, which makes the system have convenient accessibility.

(2) Guideline-based user interface information flow design enables the system to be integrated into clinical work flow relatively easily, and the system can help to improve clinicians' compliance with clinical guidelines.

(3) Modelling domain specific knowledge with BRB and inference with BRB using the ER approach enables the system to represent and reason with clinical domain knowledge under uncertainties in an informative and accurate way.

(4) Group clinical decision supporting module helps the system to provide not only a group discussion platform for experts to hold group meetings, discussions or consultations, but also a ER-based group aggregation mechanism via which a consultant group can arrive at a combined group diagnosis recommendation.

(5) Updating BRB automatically by learning through accumulated clinical cases enables the system to be adaptive to clinical practice, and this functionality helps the system to provide evidence-based clinical decision support.

(6) Structuring and storing BRB in relational database helps to keep the dynamic nature of BRB, facilitates the interactions between knowledge base and other system components, and makes the sharing of domain knowledge between different clinical systems free of technology barriers.

To conclude, the developed prototype CDSS proves that it is feasible and viable to develop a CDSS based on the RIMER methodology, and it helps to bridge the research gaps in the CDSS literature as identified in Chapter 2.

5.5. Summary

Detailed design, development, and implementation of the belief rule-based CDSS is described in this chapter. In the system development, three-layer system architecture is employed in system design, and core system components implemented in the system include: inference engine, group decision supporting module, knowledge base training module, web-based user interfaces, database, and knowledge base, where web-based user interfaces reside in the client layer, inference engine, group decision supporting module, and knowledge base training module reside in web server layer, database and knowledge base are implemented in the back-end layer. We

implemented web-based user interfaces using ASP.NET, implemented inference engine and group decision supporting module using C#, implemented knowledge base training module using MATLAB, and implemented knowledge base and database using Microsoft SQL Server 2000. The developed CDSS has the following three main system functionalities: (1) representing and reasoning with uncertain clinical domain knowledge, (2) offering group clinical decision support, and (3) providing automatic clinical belief rules updating. In addition, it has the following system features: firstly, its three layer system architecture makes it can be accessed easily through Internet or Intranet; secondly, the clinical guideline-based user interfaces can help clinicians comply with clinical guidelines; and thirdly, the knowledge base implemented using relational database can help dissemination and sharing of clinical domain knowledge free of barrier due to mature database technologies. Validation of the developed prototype system is discussed in the following chapter.

Chapter 6

Validation of the Online Intelligent CDSS Prototype

6.1. Introduction

Following the chapter of system design and development, validation of the prototype CDSS is discussed in this chapter. The purpose of system validation is to validate two key features of the system. One is the capability of handling clinical uncertainties and providing reliable diagnosis recommendations. The other one is the system can provide better diagnostic performance via learning from accumulated clinical data. Thus two core components of the system form the focus of the system validation. One is the inference engine, which is responsible for generating diagnosis recommendations by matching input clinical data with clinical rules in the knowledge base. The other is the training module, which is responsible for training or fine-tuning the knowledge base by learning from accumulated clinical data.

For the validation design, we choose CCP as the target clinical area, and the main purpose of the system is set as to aid doctors in ED to assess clinical risks of patients with CCP. Ideally, in the CDSS validation, we should use real patients' clinical data to train BRB and to test the system's diagnostic performance. However, we failed to get the ethical approval of using such data due to the strict data protection regulations in the UK, although our research collaborators in MRI have indeed managed to collect two sets of patients' data in CCP. Instead, we used a simulated dataset of 1000 patients with CCP to validate the developed prototype. The made-up dataset is provided by Dr Richard Body in MRI. All the variables in the dataset including

clinical signs and symptoms and clinical risk status have similar positive response rates to reality.

Initial 'IF-THEN' clinical rules for risk assessment of CCP are provided by our research collaborators in MRI, and the rules are their recent research outcome (Body, 2009). Based on the initial rules, we constructed belief rule base (BRB) for system validation. In inference engine validation, we compared the diagnostic performance of the system with a doctor's in assessing clinical risks of those 1000 patients. In training module validation, we split the simulated data into two sets: one set for training the system, and the other set for testing the trained system's diagnostic performance, and to avoid a trained system to overfit the training data, we tried five rounds of BRB training with different sets of training parameters to seek a set of training parameters that is most suitable for the clinical data.

Three conclusions can be drawn from the system validation study. Firstly, the system built with belief rule-based inference methodology can well handle clinical uncertainties and can provide reliable diagnosis recommendations. Secondly, the system's diagnostic performance can be improved after BRB training, and the most suitable training parameters for the BRB training model contains antecedent attribute weights and belief degrees.

We used receiver operating characteristics (ROC) curve (Metz, 1978, Park et al., 2004) which will be discussed in Section 6.4 to analyze the diagnostic performance of the system before and after each BRB training in the validation, and all ROC curves were plotted by SPSS v 17.0 software (http://www.spss.com/). We used StAR (Vergara et al., 2008, http://protein.bio.puc.cl/cardex/servers/roc/home.php), which is a specific software developed for statistical analysis of ROC curves, to compare the area under

the curve (AUC) of different ROC curves. For details of using AUC to compare ROC curves, readers can refer to (DeLong et al., 1988, Mei-Ling Ting and Bernard, 2001). For application of ROC curve analysis in diagnostic tests, readers can refer to (Body, 2009).

This chapter is structured as follows. Domain BRB employed for the validation is described in Section 6.2. Simulated dataset used in the validation is discussed in Section 6.3. A brief introduction to the ROC curve analysis is provided in Section 6.4. Inference engine validation and training module validation are presented in Section 6.5 and Section 6.6 respectively. Finally, conclusions of the chapter are summarised in Section 6.7.

6.2. Domain Knowledge Base

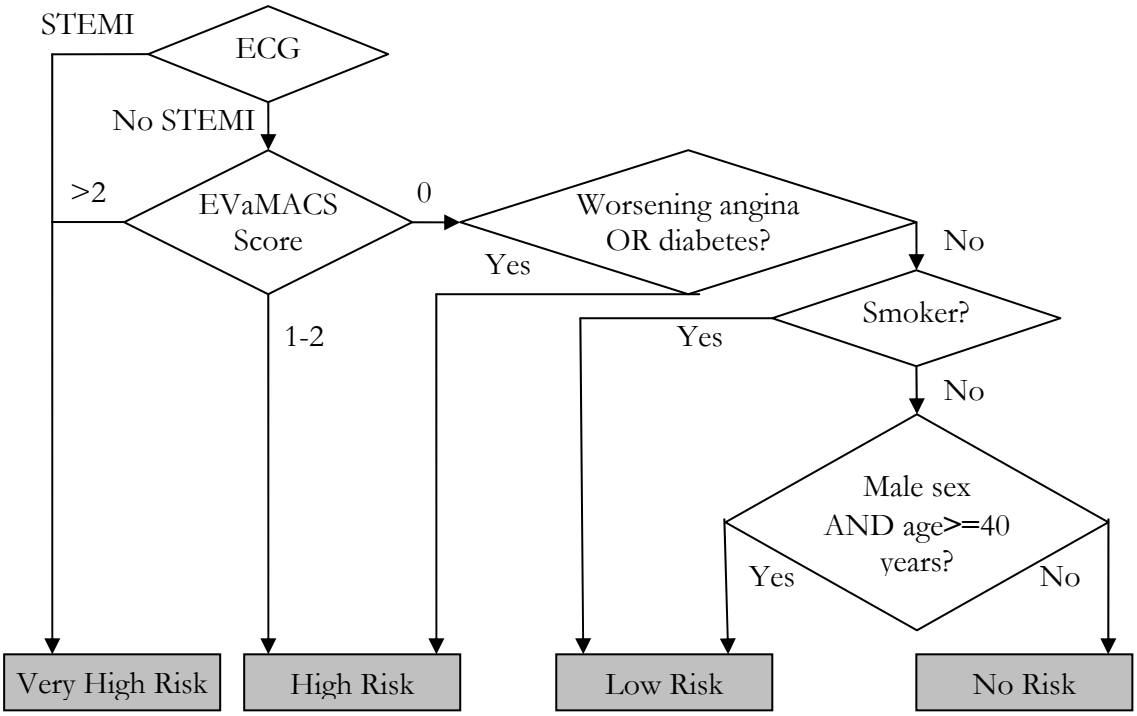
CCP is probably the most frequent serious presentation to ED in the UK. At MRI, some 3% of new attendances are covered by the label of CCP (Central Manchester and Manchester Children's University Hospitals NHS Trust, 2003a). Possible CCP can be viewed as a continuum, ranging from total global acute myocardial infarction (AMI) to simple short lived angina. Within this spectrum lie the acute coronary syndromes with critical cardiac ischaemia and minimal myocardial damage (Mackway-Jones, 2001).

In all these disorders, the risk of death is highest before admission to hospital, with mortality rates of up to 20% (Junghans and Timmis, 2006). Risk remains high after admission to hospital, and although mortality rates have fallen greatly in recent years, up to 7% of patients die before discharge, and risk continues to be high for six months after the ischaemic event (Carruthers et al., 2005).

To minimise the risk of patients with CCP, it is crucial for clinicians in ED to identify patients at high risk early on and treating them with appropriate level of care and medical therapy (Junghans and Timmis, 2006).

Motivated by the above consideration, we chose CCP as target clinical area for system validation, and the purpose of the system was set to provide support for assessing clinical risks of patients with CCP. One of our collaborators, Dr. Richard Body in MRI, has spent years in investigating more accurate and advanced rules for identifying clinical risk of CCP. A set of advanced clinical risk assessment rules as shown in Figure 6-1 are taken from his research (Body, 2009), and the knowledge base constructed for system validation is based on these rules. In Figure 6-1, STEMI stands for 'ST segment Elevation Myocardial Infarction', ECG stands for 'electrocardiography', and EVaMACS represents 'Early Vascular Markers of Acute Coronary Syndromes' and its score can be calculated from results of different clinical tests including Heart type Fatty Acid Binding Protein (H-FABP), Troponin I (TnI), and ECG as displayed at the left bottom of Figure 6-1.

If we use traditional 'IF-THEN' format to represent those rules shown in Figure 6-1, seven 'IF-THEN' rules shown in Table 6-1 can be transformed from Figure 6-1.



The EVaMACS score:
H-FABP>58ng/ml =2
TnI>0.055ng/ml =1
Acute ischaemic ECG features =2

Score>2 – Very high risk
Score 1-2 – High risk
Score 0 – Proceed to next question

Figure 6-1: Rules for Assessing Clinical Risk of CCP (Body, 2009)

Table 6-1: Traditional 'IF-THEN' Rules Transformed from Figure 6-1

No.	Antecedent	Consequent
1	IF ECG shows STEMI	THEN Very High Risk
2	IF ECG shows no STEMI, AND EVaMACS Score is >2	THEN Very High Risk
3	IF ECG shows no STEMI, AND EVaMACS Score is between 1 and 2	THEN High Risk
4	IF ECG shows no STEMI, AND EVaMACS Score equals 0, AND the patient has Worsening angina or diabetes	THEN High Risk
5	IF ECG shows no STEMI, AND EVaMACS Score equals 0, AND the patient has no Worsening angina or diabetes, AND the patient is smoking	THEN Low Risk
6	IF ECG shows no STEMI, AND EVaMACS Score equals 0, AND the patient has no Worsening angina or diabetes, AND the patient is not smoking, AND the patient's sex is Male and the patient ages >=40 years	THEN Low Risk
7	IF ECG shows no STEMI, AND EVaMACS Score equals 0, AND the patient has no Worsening angina or diabetes, AND the patient is not smoking, AND the patient's sex is Female or the patient ages <40 years	THEN No Risk

From rules as shown in Figure 6-1 and Table 6-1, we can find that *ECG status*, *EVaMACS score*, having *worsening angina/diabetes* or not, *smoking status*, *sex* and *age* all are factors that can affect clinical risk level of a patient with CCP. According to these rules, doctors would stop asking further questions or prescribing more tests for a patient if they think they have obtained enough clinical evidence and can make a final decision about the patient's risk status. For example, doctors in ED would judge a CCP patient to be at 'Very High' clinical risk if the patient's ECG shows STEMI, and the doctor would not consider other factors such as *EVaMACS score*, *smoking status*, or *diabetes status*. However, in clinical practice, a careful doctor usually would like to seek all possible clinical data to make conclusion about a patient's risk status due to inescapable uncertainties in clinical decision making.

Specifically, uncertainties occurring in the process of risk assessment of CCP may arise from the following sources. Firstly, doctors may have incomplete or vague knowledge in shaping clinical rules for risk assessment. For example, based on clinical experience, some doctors may provide such a rule as "*IF a patient's ECG is strongly suggestive of STEMI, THEN the patient has a high probability of 'Very High' risk*". Here, 'strongly suggestive' is not a clear cut description of doctors' judgements about one patient's ECG. Secondly, doctors may not be 100% sure of their judgements about patients' clinical symptoms or clinical tests. For example, doctors sometimes can not be 100% sure if a patient's ECG is consistent with STEMI, and they may use "*maybe*" to describe their judgement about the patient's ECG status. Taking these clinical uncertainties into consideration, we choose to use BRB to model clinical domain knowledge.

In our research, doctors did not provide us rules with uncertainties. Traditional rules as described in Table 6-1 for assessing clinical risk of CCP without uncertainty are what we have for system validation. We then extended the initial seven traditional 'IF-THEN' rules using a belief structure, and accordingly, a set of 48 belief rules can be created as in Table 6-2, where A^1, A^2, A^3, A^4 and A^5 represent '*ECG status'*, '*EvaMACS Score*', '*worsening angina or diabetes*', '*smoking*', and '*male sex and age larger than 40*' respectively. Validation of inference engine and training module of the system is based on this BRB.

Table 6-2: BRB for Assessing Clinical Risk of CCP

No.	Antecedent					Consequent
	A^1	A^2	A^3	A^4	A^5	Clinical Risk
1	STEMI	>2	Yes	Yes	Yes	{(Very High, 1),(High 0),(Low, 0),(No, 0)}
2	STEMI	>2	Yes	Yes	No	{(Very High, 1),(High 0),(Low, 0),(No, 0)}
3	STEMI	>2	Yes	No	Yes	{(Very High, 1),(High 0),(Low, 0),(No, 0)}
4	STEMI	>2	Yes	No	No	{(Very High, 1),(High 0),(Low, 0),(No, 0)}
5	STEMI	>2	No	Yes	Yes	{(Very High, 1),(High 0),(Low, 0),(No, 0)}
6	STEMI	>2	No	Yes	No	{(Very High, 1),(High 0),(Low, 0),(No, 0)}
7	STEMI	>2	No	No	Yes	{(Very High, 1),(High 0),(Low, 0),(No, 0)}
8	STEMI	>2	No	No	No	{(Very High, 1),(High 0),(Low, 0),(No, 0)}
9	STEMI	[1 2]	Yes	Yes	Yes	{(Very High, 1),(High 0),(Low, 0),(No, 0)}
10	STEMI	[1 2]	Yes	Yes	No	{(Very High, 1),(High 0),(Low, 0),(No, 0)}
11	STEMI	[1 2]	Yes	No	Yes	{(Very High, 1),(High 0),(Low, 0),(No, 0)}
12	STEMI	[1 2]	Yes	No	No	{(Very High, 1),(High 0),(Low, 0),(No, 0)}
13	STEMI	[1 2]	No	Yes	Yes	{(Very High, 1),(High 0),(Low, 0),(No, 0)}
14	STEMI	[1 2]	No	Yes	No	{(Very High, 1),(High 0),(Low, 0),(No, 0)}
15	STEMI	[1 2]	No	No	Yes	{(Very High, 1),(High 0),(Low, 0),(No, 0)}
16	STEMI	[1 2]	No	No	No	{(Very High, 1),(High 0),(Low, 0),(No, 0)}
17	STEMI	0	Yes	Yes	Yes	{(Very High, 1),(High 0),(Low, 0),(No, 0)}
18	STEMI	0	Yes	Yes	No	{(Very High, 1),(High 0),(Low, 0),(No, 0)}
19	STEMI	0	Yes	No	Yes	{(Very High, 1),(High 0),(Low, 0),(No, 0)}
20	STEMI	0	Yes	No	No	{(Very High, 1),(High 0),(Low, 0),(No, 0)}
21	STEMI	0	No	Yes	Yes	{(Very High, 1),(High 0),(Low, 0),(No, 0)}
22	STEMI	0	No	Yes	No	{(Very High, 1),(High 0),(Low, 0),(No, 0)}
23	STEMI	0	No	No	Yes	{(Very High, 1),(High 0),(Low, 0),(No, 0)}
24	STEMI	0	No	No	No	{(Very High, 1),(High 0),(Low, 0),(No, 0)}
25	No	>2	Yes	Yes	Yes	{(Very High, 1),(High 0),(Low, 0),(No, 0)}
26	No	>2	Yes	Yes	No	{(Very High, 1),(High 0),(Low, 0),(No, 0)}
27	No	>2	Yes	No	Yes	{(Very High, 1),(High 0),(Low, 0),(No, 0)}
28	No	>2	Yes	No	No	{(Very High, 1),(High 0),(Low, 0),(No, 0)}
29	No	>2	No	Yes	Yes	{(Very High, 1),(High 0),(Low, 0),(No, 0)}
30	No	>2	No	Yes	No	{(Very High, 1),(High 0),(Low, 0),(No, 0)}
31	No	>2	No	No	Yes	{(Very High, 1),(High 0),(Low, 0),(No, 0)}
32	No	>2	No	No	No	{(Very High, 1),(High 0),(Low, 0),(No, 0)}
33	No	[1 2]	Yes	Yes	Yes	{(Very High, 0),(High 1),(Low, 0),(No, 0)}
34	No	[1 2]	Yes	Yes	No	{(Very High, 0),(High 1),(Low, 0),(No, 0)}
35	No	[1 2]	Yes	No	Yes	{(Very High, 0),(High 1),(Low, 0),(No, 0)}
36	No	[1 2]	Yes	No	No	{(Very High, 0),(High 1),(Low, 0),(No, 0)}
37	No	[1 2]	No	Yes	Yes	{(Very High, 0),(High 1),(Low, 0),(No, 0)}
38	No	[1 2]	No	Yes	No	{(Very High, 0),(High 1),(Low, 0),(No, 0)}
39	No	[1 2]	No	No	Yes	{(Very High, 0),(High 1),(Low, 0),(No, 0)}
40	No	[1 2]	No	No	No	{(Very High, 0),(High 1),(Low, 0),(No, 0)}
41	No	0	Yes	Yes	Yes	{(Very High, 0),(High 1),(Low, 0),(No, 0)}
42	No	0	Yes	Yes	No	{(Very High, 0),(High 1),(Low, 0),(No, 0)}
43	No	0	Yes	No	Yes	{(Very High, 0),(High 1),(Low, 0),(No, 0)}
44	No	0	Yes	No	No	{(Very High, 0),(High 1),(Low, 0),(No, 0)}
45	No	0	No	Yes	Yes	{(Very High, 0),(High 0),(Low, 1),(No, 0)}
46	No	0	No	Yes	No	{(Very High, 0),(High 0),(Low, 1),(No, 0)}
47	No	0	No	No	Yes	{(Very High, 0),(High 0),(Low, 1),(No, 0)}
48	No	0	No	No	No	{(Very High, 0),(High 0),(Low, 0),(No, 1)}

6.3. Simulated Dataset

The simulated dataset used for system validation is provided by our research collaborator, Dr Richard Body, working in MRI. In the dataset, independent variables, which contribute to clinical risk status of patients with CCP, include clinical signs or symptoms, demographics, and clinical test results of simulated patients, namely '*ECG*', '*Worsening_Angina*', '*Diabetes*', '*Smoking*', '*Sex*', '*Age*', and '*EvaMACS_Score*'. Dependent variable in the dataset is '*Outcome*' which is used for recording the outcome of the simulated patients. Here, the outcome was the composite of AMI or the occurrence of adverse events within six months, where adverse events were defined as death (all-cause), AMI or the need for urgent coronary revascularisation (Body, 2009).

In the dataset, two numerical values including 1 and 0 are used to record outcome of simulated patients, where 1 represents that the patient had AMI or he/she died, had AMI or needed urgent coronary revascularisation within six months, and 0 represents that the patient had no real clinical risk. As for the values used to record '*ECG'*, subjective judgements including '*definitely yes*', '*strongly suggestive*', '*maybe*', '*a little like*', and '*absolutely no*' are used to simulate patients' ECG status which is diagnosed as with STEMI under uncertainties, and in the dataset, Dr Richard Body has transformed these subjective judgements into degrees of belief in STEMI. Specifically, for variable '*ECG'*, 1 represents '*definitely yes with STEMI*', 0.8 represents '*strongly suggestive of STEMI*', 0.5 represents '*maybe STEMI*', 0.2 represents '*a little like STEMI*', and 0 represents '*absolutely no STEMI*'. For other variables including '*Worsening_Angina*', '*Diabetes*', and '*Smoking*', value of 1 represents 'yes' while value of 0 represents 'no'. For variable '*Sex*', 1 represents male

and 0 represents female. Table 6-3 displays five example patients' data extracted from the simulated dataset.

Table 6-3: Example Patients' Data in the Simulated Dataset

No.	Outcome	ECG	Worsening Angina	Diabetes	Smoking	Sex	Age	EVaMACS_Score
1	0	0.2	0	0	0	0	87	0
2	1	1.0	0	0	0	0	81	1
3	1	0.8	0	0	1	0	80	1
4	0	0.0	0	0	0	0	39	0
5	1	0.5	1	0	0	1	61	1

The dataset has two features that are important for the research. Firstly, the dataset is close to reality. All of the variables including clinical signs or symptoms, demographics, clinical test results and outcome in the dataset have similar positive response rates to reality. For example, in clinical practice, around 20% of patients with CCP attended in ED are with STEMI, and among them, some are definitely with STEMI, some are strongly suggestive of STEMI, while some others show a little sign of STEMI. The probabilities of the various STEMI situations are reflected in the simulated data. Secondly, the dataset reflects uncertainties in clinical decision making. In the simulated dataset, doctors' judgement of a patient's ECG can be 'definitely yes with STEMI', 'strongly suggestive of STEMI', 'maybe STEMI', 'a little like STEMI', and 'absolutely no STEMI'.

Note that there are some conflicting cases in the dataset. For example, for the Rule 13 in the BRB as described in Table 6-2, if one patient's clinical data match this rule's conditions, the patient should be at 'Very High' clinical risk. While in the simulated dataset, there are 35 cases that are consistent with the rule, however there are two other cases having no real clinical risk in spite that their ECGs show 'strongly

suggestive' of STEMI and their other clinical data exactly match other conditions of the rule. The latter 2 cases conflict with the former 35 cases.

To enforce our confidence of using the simulated dataset for the system validation, we presented all conflicting data to two more sophisticated expert clinicians in CCP in MRI to seek their advice for handling these conflicting cases in the system validation. Based on their knowledge and clinical experience, the two experts believed that this type of medical errors could happen everyday due to uncertainties, and they suggested us including all conflicting data in the validation. Thus we tried to include all conflicted cases in the simulated dataset for both the inference engine and the training module validation.

6.4. Brief Introduction to Receiver Operating Characteristic (ROC) Curve Analysis

In the prototype validation, a necessary procedure is to compare the diagnostic performance of the system with a doctor's. Moreover, to validate the implemented training module as described in Section 5.3.3 of Chapter 5, we need to compare the diagnostic performance of the system before and after BRB training. As such, a question would arise as how to evaluate the diagnostic performance of different tests.

Usually, in diagnostic research, there are several ways for evaluating performance or accuracy of a diagnostic test such as overall diagnostic accuracy, diagnostic odds ratios, and ROC curve (Body, 2009), where the overall diagnostic accuracy and diagnostic odds have their drawbacks and will be briefly discussed together with ROC curve in Section 6.4.1. As identified by Body (2009), the ROC curve has its advantage in demonstrating diagnostic performances as it can be used to summarise the accuracy

of an investigation with a single number by calculating the area under the curve (AUC). In the literature, the ROC curve has been widely used in evaluating the performance of diagnostic tests or some other classifiers, and the AUC has also been widely used for comparing performance of different diagnostic tests and machine learning algorithms (Metz, 1978) (Body, 2009) (Bradley, 1997, Jin and Ling, 2005). Taking the above into consideration, we opted to use the ROC curve to measure diagnostic performance of all tests in the validation.

In this Section, a brief introduction to the ROC curve and the AUC is presented in Section 6.4.1 and Section 6.4.2 respectively, followed by a brief discussion about comparison of the AUC for different ROC curves in Section 6.4.3.

6.4.1. ROC Curve

Before introducing the ROC curve, we will briefly discuss some other measures for evaluating diagnostic performance of a test first. In measuring a diagnostic test, a decision matrix as described in Table 6-4 can be created to evaluate the diagnostic performance of the test, where *Positive* means having a specific disease while *Negative* means having no the disease.

Table 6-4: Decision Matrix for a Diagnostic Test (Body, 2009)

	True Condition Status				
Test Result	**Positive (+)**	***number***	**Negative (-)**	***number***	**Total**
Positive (+)	True Positive (TP)	a	False Positive (FP)	b	$a + b$
Negative (-)	False Negative (FN)	c	True Negative (TN)	d	$c + d$
Total		$a + c$		$b + d$	

The overall diagnostic accuracy of the diagnostic test is determined as the proportion of cases in which the results of the diagnostic test and the reality are the same:

$$Overall\ diagnostic\ accuracy\ = \frac{a+d}{a+b+c+d}$$

The diagnostic odds ratio is defined as the odds of a positive test result in patients with disease, relative to the odds of a positive test result in patients without disease:

$$Diagnostic\ odds\ ratio = \frac{a/c}{b/d} = \frac{ad}{bc}$$

Both of above described measures have their disadvantages in evaluating diagnostic performance of a test. For example, if the prevalence of disease in the test population is very low, any investigation that returns predominantly negative results will tend to have a high overall diagnostic accuracy. For details, readers can refer to (Body, 2009).

A better way to assess the diagnostic test performance is to use *sensitivity* or *True Positive Rate (TPR),* and *specificity* or *True Negative Rate (TNR)* (Body, 2009). *Sensitivity* is determined by the proportions of patients with the disease who were correctly identified by the diagnostic test. *Specificity* is determined by the proportion of negatives which are correctly identified. These statistics can be calculated as follows:

$$Sensitivity = \frac{a}{a+c}$$
$$Specificity = \frac{d}{b+d}$$

The sensitivity and specificity of a diagnostic test can tell us about the ability of the test to discriminate between healthy and diseased patients. Moreover, these values are independent of the disease prevalence.

When the results of a diagnostic test fall into two obviously defined categories, such as either the presence or absence of a disease, then the test has only a pair of

sensitivity and specificity values using the *sensitivity* and *specificity* calculations as discussed above. However, in many diagnostic situations, making a decision in a binary mode is both difficult and impractical. For example, Some diagnostic test results may be ordinal (for example a risk score with possible values being whole numbers ranging from 0 to 5) or continuous (for example a blood test, where possible values can be anything within the detectable range of the instruments, including decimals). In order to calculate sensitivity and specificity for ordinal and continuous data, the values must first be dichotomised. To do this we must select an appropriate threshold value as a diagnostic cut-off. Values above this cut-off would be considered 'positive' and values below it considered 'negative'. As a result, a single pair of sensitivity and specificity values is insufficient to describe the full range of diagnostic performance of a test (Metz, 1978). In such situations, the ROC curve can be used for evaluation of the diagnostic performance of an investigation (Body, 2009). Details of the ROC curve are briefly discussed as follows.

The ROC curve, which is defined as a plot of test sensitivity or TPR as the y coordinate versus its 1-specificity or false positive rate (FPR), which is determined by the proportion of negatives which are wrongly identified, as the x coordinate, is an effective method of evaluating the performance of diagnostic tests which have ordinal or continuous results (Body, 2009). Each point on the graph represents a pair of sensitivity and 1-specificity based on a different diagnostic cut-off value.

Take a fictional diagnostic test for example, there are 20 patients involved in the test. These patients have different clinical risk status in reality and they have been judged by doctors with different risk scores ranging from 1 to 5. The detailed data are presented in Table 6-5, where value 1 in *Real Status* column means that the patient is

at high clinical risk, while 0 means no clinical risk. To measure the performance of the fictional diagnostic test, a ROC curve as represented by blue line in Figure 6-2 can be generated by SPSS, and all cut-off values used to shape the curve by SPSS is shown in Table 6-6 which is cut from SPSS, where pairs of specificity and 1-specificity values which correspond to each cut-off value are displayed as well. Take the cut-off value of 2.5 for example, based on this cut-off, the calculated sensitivity of the fictional diagnostic test is 0.9167, and 1-specificity of the test is 0.1250.

Table 6-5: A Fictional Diagnostic Test Data

No.	Risk Score	Real Status	No.	Risk Score	Real Status
1	1	0	11	1	0
2	2	0	12	2	0
3	3	1	13	3	0
4	4	1	14	4	1
5	5	1	15	5	1
6	1	0	16	1	0
7	2	1	17	2	0
8	3	1	18	3	1
9	4	1	19	4	1
10	5	1	20	5	1

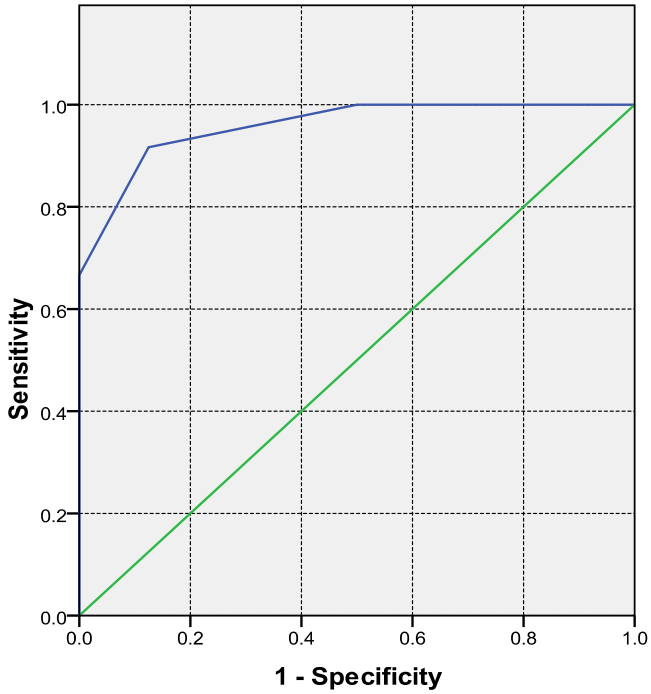


Figure 6-2: ROC Curve Demonstrating the Performance of a Fictional Diagnostic Test

Table 6-6: Coordinates of the Curve (Cut from SPSS)

Diagnostic Cut-off	Sensitivity	1-Specificity
0.0000	1.0000	1.0000
1.5000	1.0000	0.5000
2.5000	0.9167	0.1250
3.5000	0.6667	0.0000
4.5000	0.3333	0.0000
6.0000	0.0000	0.0000

A perfect diagnostic test would have 100% sensitivity and specificity and the ROC curve would therefore intersect the top left hand corner of the graph. If a diagnostic test has no ability to differentiate between healthy and diseased patients, the ROC curve will take the form of a straight line intersecting the bottom left and top right diagonals which is called 'chance diagonal', and it is represented by the green line in Figure 6-2.

6.4.2. Area under the Curve (AUC): a Measure of Overall Diagnostic Performance

An advantage of the ROC curve is that it can be used to summarise the accuracy of a diagnostic test with a single number by calculating the size of the area under the curve (AUC) (Body, 2009). The AUC can take any value between 0 and 1, since both the x and y axes have values ranging from 0 to 1 and size of the square between (0, 0) and (1, 1) is 1. The closer AUC is to 1, the better the overall diagnostic performance of the test. A test with an AUC value of 1 is one that is perfectly accurate, while a test with an AUC value of 0 is one that is perfectly inaccurate. The practical lower limit for the AUC of a diagnostic test is 0.5. Because if we were to rely on pure chance to distinguish those subjects with a particular disease against those without a particular disease, the resulting ROC curve would fall along this diagonal line, which is referred to as the chance diagonal as shown in Figure 6-2, and the line segment from (0, 0) to (1, 1) has an area with size of 0.5.

In the fictional example of clinical risk assessment as discussed in Section 6.4.1, the AUC is estimated to be 0.9583 by SPSS, which suggests that the diagnostic performance of the fictional test is very good as the AUC is very close to 1.

Note that a ROC curve and its AUC can be generated by different methods, namely parametric and non-parametric approaches. If we use parametric to estimate a ROC curve or the AUC, we need to make assumption about the distribution of the diagnostic test's results, and very often bi-normal distribution is assumed (Skalska and Freylich, 2006). While we do not have to do any assumption about the test results if we use non-parametric approach to plot ROC curve or estimate the AUC. In the research, we chose non-parametric approach to do ROC analysis, and the ROC

analysis conducted in this thesis is also based on non-parametric approach. For details of parametric and non-parametric approaches for conducting ROC analysis, readers can refer to (Metz, 1978, Hanley, 1988, Zou et al., 1997, DeLong et al., 1988, Mei-Ling Ting and Bernard, 2001).

6.4.3. Comparing the AUC: Comparing Overall Diagnostic Performance

The overall diagnostic performance of different tests can be compared by comparing AUC of different ROC curves, as AUC is a measure of the overall performance of a diagnostic test. The bigger its AUC is, the better the overall performance of the diagnostic test will have. For example, we can easily find that diagnostic test B0 has better performance than A0 from Figure 6-3, as the AUC in test A0 is 0.9583 and the AUC in test B0 is 0.9844.

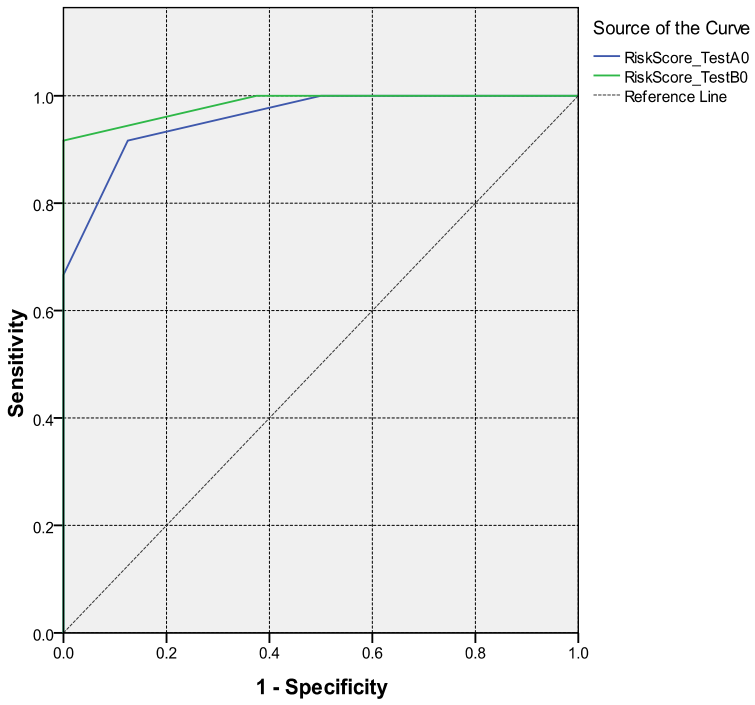


Figure 6-3: Two ROC Curves with Different Values of the AUC (A0-0.9583; B0-0.9844)

Furthermore, to test the statistical significance of the difference between the areas under different ROC curves, some specific software have been developed by researchers for AUC comparison. Frequently mentioned software for AUC comparison in the literature include MedCalc (http://www.medcalc.be/), ROCKIT (http://www-radiology.uchicago.edu/krl/KRL_ROC/software_index6.htm), and StAR (http://protein.bio.puc.cl/cardex/servers/roc/home.php). In the research, we chose StAR to do AUC comparison as StAR is online software which can be accessed freely and can meet our requirements for comparison of paired data. StAR is designed for the ROC analysis of paired data and the core of the software is a non-parametric test for the difference of the AUC that accounts for the correlation of the ROC curves. Here, paired data are data generated from those diagnostic tests in which each case in the studied sample has been tested (Metz et al., 1998).

Generally, we call the difference between the diagnostic performances of two tests that are summarized by AUC statistically significant if there is enough evidence showing that the difference does not occur by chance. When we use statistical software to compare two groups of paired data, we can get a *p*-value from the comparison, where *p*-value is a measure of probability that a difference between two groups of data happened by chance. In statistics, the highest acceptable *p*-value, at which we can still say that a difference between two groups does not happen by chance, is called *significance level* or *α level* (Aczel and Sounderpandian, 2005, Wright, 1997). The difference between two groups of data can be described as statistically significant when a *p*-value is less than the set significance level, and as non-significant when a *p*-value is above the significance level. Conventionally, significance level is set to be 0.05 (Wright, 1997). The lower the *p*-value, the more likely it is that the difference between groups of data does not occur by chance.

In the following discussion, we will use *p*-values generated by StAR in comparing AUC of ROC curves to measure the statistical significance of differences between different diagnostic performances.

6.5. Inference Engine Validation

6.5.1. Method

Validation of the inference engine was basically composed of three main steps. The first one was to produce a doctor's assessment for risks of the simulated 1000 patients and to calculate the doctor's overall diagnostic performance. To facilitate the process of acquiring the doctor's assessment for those 1000 patients' clinical risk status, we produced risk assessment results for the 1000 patients first based on the initial rules as shown in Figure 6-1, and then we invited one of our collaborators in MRI to verify the judgements. The second step included using the simulated patients' clinical data as inputs to the system and triggering the system to assess clinical risk of the patients, and then calculating the system's overall diagnostic performance. The third step was to compare the system's diagnostic performance with the doctor's and draw conclusions about the reliability of the prototype system. The ROC analysis as discussed in Section 6.4 was used to analyze the diagnostic performances of the system and the doctor.

As noted in Section 6.4.1, the ROC curve analysis can effectively measure performances of diagnostic tests having ordinal or continuous results. While in our research, according to the evidential reasoning (ER) approach and the BRB described in Table 6-2, the system's inferred result for each patient should be a belief distribution among four risk levels, namely 'Very High', 'High', 'Low', and 'No'.

Thus we need to transform the inferred diagnosis result, which is distributed in different risk levels, into a value that suits the ROC analysis.

As proposed by Yang and Xu (2002), if necessary, an overall utility value can be estimated from an assessment of both qualitative and quantitative characteristics. Similar to transforming assessments of alternatives under decision into overall utility values, the diagnosis recommendations about patients' risk status provided by the system or the doctor can be transformed into overall severity scores, as mentioned in Section 4.6.1 of Chapter 4. As the overall severity score is numerical and continuous in the range from 0 to 1, ROC curves can be constructed from overall severity scores to demonstrate diagnostic performances of different tests. Therefore in our research, we used the overall severity scores estimated from risk assessment results that are either generated from the system or provided by the doctor to compare diagnostic performances.

To estimate the overall severity score from risk assessment result of each patient, we need to estimate severity scores of those four different risk levels as described in Figure 6-1 first.

With advices from an expert clinician, we assigned a severity score of 1 to 'Very High', 0.67 to 'High', 0.33 to 'Low', and 0 to 'No'. For example, we can estimate a patient's overall severity score as 0.9668 if the risk result generated by the system for the patient is {(Very High, 0.94), (High, 0.04), (Low, 0), (No, 0)}. Therefore, every patient in the simulated dataset can be given an overall severity score automatically by the system or manually based on the risk assessment result produced by one doctor, and we can then use the overall severity scores of those 1000 simulated patients generated in different situations to do the ROC analysis.

6.5.2. Results

In inference engine validation, we obtained two sets of overall severity scores of the simulated 1000 patients. One set was automatically generated by the system and the other set was manually produced based on a doctor's judgements.

In using the ROC curve to analyze diagnostic performances of the system and the doctor, we used the recorded outcome of those 1000 patients as benchmark, and we obtained the following two ROC curves as shown in Figure 6-4, which represent the diagnostic performances of the system and the doctor's. The ROC curve as represented by the blue line in Figure 6-4 is plotted from the severity score set generated by the system, and the AUC is 0.7921 (95% confidence intervals 0.7586 – 0.8257). The ROC curve as represented by the green line in Figure 6-4 is plotted from the severity score set manually produced based on the doctor's judgements, and its AUC is 0.7525 (95% confidence intervals 0.7177 – 0.7873). Here we required SPSS to give a 95% confidence interval estimate together with a single AUC value estimate, where the 95% confidence interval can tell us that the value of the parameter in estimation can lie within the estimated interval with 95% certainty (Aczel and Sounderpandian, 2005).

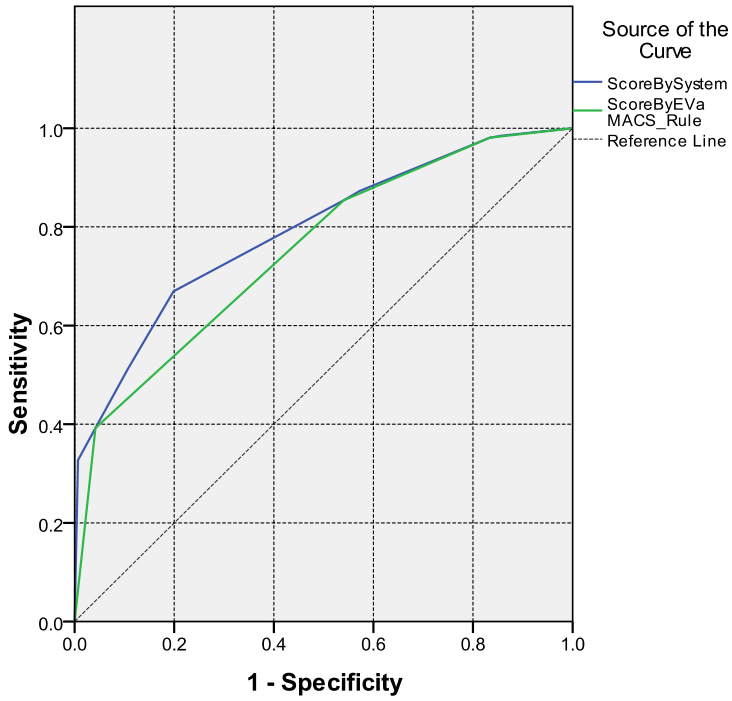


Figure 6-4: ROC Curves Demonstrating the Diagnostic Performances of the System and One Doctor

From the two ROC curves as displayed in Figure 6-4, we can see that the AUC representing diagnostic performance of the system is larger than the AUC representing performance of the doctor. To test whether the difference of performances between the system's and the doctor's is caused by chance or not, we then used StAR to compare the AUC of these two ROC curves and got a *p*-value less than 0.0001.

The results show that under clinical uncertainties, the diagnostic performance of the CDSS prototype implemented using the RIMER methodology is better than manual judgement produced by a doctor, and the performance difference is statistically significant with a perfect *p*-value less than 0.0001.

6.6. Training Module Validation

6.6.1. Method

The kernel model of the training module integrated in the system is the BRB optimization model as described in Section 3.2.3 of Chapter 3 and Section 5.3.3 of Chapter 5. In the validation, we need to validate that the diagnostic performance of the system can be improved after BRB training with accumulated clinical cases. Thus we should test performance of the system after BRB training using a dataset which is independent and not included in the training dataset.

There is no a general rule on how to choose the training examples and the test examples size (Hastie et al., 2001). Conventionally, in machine learning applications, to measure performance of the training system, we can split all available data in half, while one half goes to the training set for system training and the other half goes to the test set for performance measurement after training (Seufert and O'Brien, 2007, Liu et al., 2005, Agarwal et al., 2010, Yang et al., 2007). About the method we used to draw training data and test data in the training module validation, we will discuss it in next Section 6.6.1.1.

As described in Section 5.3.3 of Chapter 5, knowledge representation parameters including rule weights $\theta_k \left(k = 1, \cdots, L\right)$, antecedent attribute weights $\delta_i \left(i = 1, \cdots, T\right)$, and consequent belief degrees $\beta_{kj} \left(k = 1, \cdots, L; j = 1, \cdots, N\right)$ are main training parameters for the training module. In our training module design and development as described in Section 5.3.3 of Chapter 5, we employed the training model proposed by Yang et al. (2007) which is based on numerical system outputs, .thus we need to transform the distributed inferred results $\left(D_j, \beta_j\right)\left(j = 1, \cdots, N\right)$ generated from the

prototype for one patient into a numerical value that can denote the patient's severity status. Here, based on the utility concept $\mu(D_j)(j=1,\cdots,N)$ proposed by Yang and Xu (2002), as described in Section 6.5.1, different severity scores can be assigned to the four consequent risk levels of the BRB in Table 6-2, and thus an overall severity score can be generated by the system based on the inferred distributed result for one patient. In such situations, the severity scores of the four consequent risk levels in the BRB can also be trained by the training data. Here, $D_j(j=1,\cdots,N)$ is the consequents of the BRB, and $\beta_j(j=1,\cdots,N)$ is the inferred belief degree from the prototype associated to the *j*th consequent. *L* is the number of rules in the BRB, and it equals 48 in our system; *T* is the number of antecedent attributes used in the BRB, and it equals 5; *N* is the number of consequents in the BRB, and it equals 4 in the prototype. For details of the BRB, readers can refer to Table 6-2 in Section 6.2.

However, in machine learning, a common issue is overfitting. A trained model is thought to overfit training data if there is some alternative trained model, such that the former model fits the training data better than the alternative one, but the alternative one performs better than the former over a test dataset which is independent of the training data (Mitchell, 1997). Overfitting is especially likely to happen when we give the training algorithm a very rich searching space for the parameters of the model being trained and thus enable the model being trained to overfit the training data (Mitchell, 1997).

In the context of BRB training, we may get a trained BRB which can overfit the training data if we give the training model too many parameters that can vary during the training process.

To avoid overfitting in BRB training, we tried five different sets of training parameters for the training module. In the following discussion, we will use R^1, R^2, R^3, R^4, and R^5 to represent the BRB training with different parameters, where

R^1: training with rule weights $\theta_k (k = 1,\cdots,48)$, antecedent attribute weights $\delta_i (i = 1,\cdots,5)$, consequent belief degrees $\beta_{kj} (k = 1,\cdots,48; j = 1,\cdots,4)$, and severity scores of the four different risk levels $\mu(D_j)(j = 1,\cdots,4)$;

R^2: training with rule weights $\theta_k (k = 1,\cdots,48)$, antecedent attribute weights $\delta_i (i = 1,\cdots,5)$, and consequent belief degrees $\beta_{kj} (k = 1,\cdots,48; j = 1,\cdots,4)$;

R^3: training with antecedent attribute weights $\delta_i (i = 1,\cdots,5)$ and consequent belief degrees $\beta_{kj} (k = 1,\cdots,48; j = 1,\cdots,4)$;

R^4: training with rule weights $\theta_k (k = 1,\cdots,48)$ and consequent belief degrees $\beta_{kj} (k = 1,\cdots,48; j = 1,\cdots,4)$;

R^5: training with belief degrees $\beta_{kj} (k = 1,\cdots,48; j = 1,\cdots,4)$.

There are actually some other combinations of the parameters for BRB training, the reasons for us to try above five combinations are as follows. The main purpose for the training is to find a BRB model that can better represent domain knowledge in assessing clinical risk of CCP, though we set all knowledge representation parameters of the BRB and severity scores of consequent risk levels as training parameters in the first training round R^1, we took severity scores of the four risk levels out of training in round R^2 as too many training parameters may cause the trained system overfit the training data. As it is really hard for domain clinicians to give exact belief degrees to

different consequent severity levels in those 48 rules as shown in Table 6-2, we took consequent belief degrees as core parameters that need to be fine-tuned by training data. Thus after training round R^2 where all knowledge representation parameters were taken as training parameters, we tried to combine consequent belief degrees with antecedent attribute weights and rule weights respectively as training parameters in training rounds R^3 and R^4, and finally in training round R^5, we tried to put only consequent belied degrees as training parameters to avoid overfitting.

Taking above mentioned into consideration, we designed the training module validation as follows.

Firstly, we drew representative training examples and test examples from the simulated 1000 cases.

Secondly, we trained the system by simulated cases in the training set with above mentioned five different set of training parameters, and we set the same initial values for the training parameters of each training round.

Thirdly, we test the performance of the system before and after different BRB training over simulated cases in the test set.

Fourthly, we analyzed the system's diagnostic performance using ROC curves as described in Section 6.4.

Finally, based on the system performance analysis results, we drew conclusions about the training module.

Further details regarding how to split the simulated cases into training set and test set and how to set initial values and constraints for training parameters will be discussed

in the following Sections 6.6.1.1 and 6.6.1.2.

6.6.1.1. Training Set and Test Set

First and foremost, we need to draw training examples and test examples from those simulated 1000 clinical cases. Usually, in supervised machine learning as described in Section 3.2.3 of Chapter 3, training set can be drawn from available data randomly (Dale et al., 2010, Tong and Koller, 2001). In choose training examples, an important attribute of training examples is how well it represents the distribution of test examples over which performance of the trained system be measured, and generally, learning is most reliable when the training examples follow a similar distribution to that of the future test examples (Mitchell, 1997, Freund et al., 1997).

In the context of BRB model training and validation, to ensure the reliability of BRB training, we need to draw similar data into both training set and test set. As in BRB training, it is not uncommon that the following two situations about the training would happen. Firstly, in some cases, parameters related to some clinical rules can be trained from the training data, but there are no cases in the test set that can activate the trained rules in testing performance of the trained system, and then it would affect the performance evaluation of the training module since not all trained rules have made contributions in the test of system performance after training. Secondly, in some other cases, parameters related to some clinical rules can not be trained by the training data due to lack of training examples in these regions where BRB was designed to operate, but there are cases in the test set that can activate the untrained rules in system performance testing after training, and then it would also affect performance evaluation of the training module as untrained rules would lead to irrational conclusions if they were initially assigned randomly or without care (Yang et al.,

2007). Take Rule 1 in Table 6-2 for example, if knowledge representation parameters related to Rule 1 can be trained from the training data, then in the test set, there should be some patients' data fall in the region that Rule 1 was designed to operate so that the trained Rule 1 can play its role in system performance test after BRB training, and thus the performance of the training module can be rationally evaluated.

To draw representative data into training set and test set for training and validating all rules in the BRB model, we analyzed the matching status between all simulated cases and clinical rules in the BRB first, and then we randomly split matched cases of each rule probably into half (if there are enough cases for both training and test), while probably one half goes to training set and the other half for test set. The details are as follows.

- **Analyzing matching status between simulated cases and clinical rules in the BRB**

In analyzing the matching status between simulated patients and clinical rules, we checked the matching status between each simulated patient's clinical data except recorded outcome and one specific rule's antecedents. Here, the degree of matching between one simulated patient's data and one clinical rule need not be with 100% certainty, because inference with BRB as described in Section 4.5.2 of Chapter 4 and Section 5.3.1 of Chapter 5 can consider different matching degrees between input data and one rule's packet antecedent.

In the simulated dataset as described in Section 6.3, we have uncertain judgements about patients' ECG status: 'definitely yes with STEMI', 'strongly suggestive of STEMI', 'maybe STEMI', 'a little like STEMI', and 'absolutely no STEMI'. Thus for some clinical cases, they may be consistent with two different clinical rules'

antecedents to different degrees because of the uncertain ECG judgements. In calculating the number of matched cases to one clinical rule in the BRB, we considered all cases whose matching degrees to the rule are larger than 0. The number of all matched clinical cases to each clinical rule in the BRB as described in Table 6-2 is reported in Table 6-7.

Table 6-7: Number of Matched Clinical Cases to Each Clinical Rule in the BRB

Rule No.	Number of Matched Cases	Rule No.	Number of Matched
1	2	25	0
2	1	26	0
3	2	27	1
4	2	28	2
5	0	29	2
6	1	30	0
7	1	31	0
8	1	32	1
9	42	33	34
10	16	34	18
11	58	35	64
12	37	36	45
13	35	37	49
14	13	38	39
15	36	39	87
16	40	40	97
17	4	41	16
18	3	42	14
19	15	43	38
20	10	44	31
21	8	45	59
22	8	46	59
23	13	47	132
24	18	48	128

- **Splitting simulated cases into training set and test set**

From Table 6-7, we can find that the number of matched cases to rules numbering from 1 to 8 and from 25 to 32 is 0 or 1 or 2, which means there are no enough cases falling in the regions where these rules were designed to operate for both training and testing these rules. Specifically, for Rule 5, Rule 25, Rule 26, Rule 30, and Rule 31, there is no matched case in the dataset, and thus it is impossible to use the dataset to

train these rules. For Rule 2, Rule 6, Rule 7, Rule 8, Rule 27 and Rule 32, there is only one clinical case matched to each of these rule's antecedents, and this means if we put the matched case in the training set, there would be no case in the test set to activate the rule after training. For Rule 1, Rule 3, Rule 4, Rule 28, and Rule 29, there are two cases in the dataset that match each rule's antecedents, and the only option for us to do both training and testing for these rules is to put one case in the training set and put the other case in the test set. However, if two clinical cases matched to one clinical rule have extremely different recorded outcomes, it will negatively affect the reliability of the training module if we put one case in the training set and put the other case in the test set.

Taking above into consideration, we put aside the 12 cases in the simulated dataset that are matched to rules numbering from 1 to 4, from 6 to 8, from 27 to 29, and Rule 32 as shown in Table 6-7, and randomly split the remaining 988 cases, with one half for training and the other half for test. During the data splitting, we tried to make that probably half of matched cases to each rule goes for training and the other half goes for test. We did so to ensure that each rule that has been trained in the training process can make contribution to system performance test after training.

As a result, rules numbering from 1 to 8 and rules numbering from 25 to 32 have not be trained in the training process or tested in the test process, and the final diagnostic performance analyses were based on clinical cases that match rules numbering from 9 to 24 and rules numbering from 33 to 48 in the BRB as illustrated in Table 6-7.

6.6.1.2. Initialization of the Training Model

- **Initial values of the training parameters**

For BRB training, a necessary task is to set initial values and constraints for the training parameters. We set the initial values of the training parameters for each BRB training round of R^1, R^2, R^3, R^4, and R^5 as follows.

(1) Severity scores of four risk levels $\mu(D_j)(j=1,\cdots,4)$: $\mu(D_1(\text{Very High}))=1$, $\mu(D_2(\text{High}))=0.67$, $\mu(D_3(\text{Low}))=0.33$, and $\mu(D_4(\text{No}))=0$;

(2) Rule weights $\theta_k(k=1,\cdots,48)$: $\theta_k(k=1,\cdots,48)=1$;

(3) Antecedent attribute weights $\delta_i(i=1,\cdots,5)$: $\delta_i(i=1,\cdots,5)=1$;

(4) Consequent belief degrees $\beta_{kj}(k=1,\cdots,48; j=1,\cdots,4)$: we set the initial values of the consequent belief degrees in the BRB based on the statistical calculation of the training dataset. For example, for Rule 11 in the BRB in Table 6-2, there are 9 cases in the training dataset that match the rule's antecedents with 100% certainty, and the recorded outcome of those 9 cases show that 8 of them were at 'Very High' clinical risk while 1 case had 'No' clinical risk. Thus we set the initial values of consequent belief degrees in Rule 11 as {(Very High, 8/9=0.8889), (High 0), (Low, 0), (No, 0.1111)}. Here, with advice from experts, we assigned 'Very High' clinical risk to patients with outcome of 1 and 'No' clinical risk to patients with outcome of 0 in the simulated dataset.

The initial values of severity scores of four risk levels, antecedent attribute weights, rule weights, and consequent belief degrees will be displayed together with corresponding trained values after different BRB training rounds in Table 6-8, Table 6-9, Table 6-10, and Table 6-11 respectively in Section 6.6.2.1. Readers can refer to

the beginning part of Section 6.6.1 for details about the training parameters used in training rounds R^1, R^2, R^3, R^4, and R^5.

- **Constraints of the training parameters**

In terms of the constraints for training parameters, we set constraints for training parameters as follows.

(1) Severity scores of four risk levels $\mu(D_j)(j=1,\cdots,4)$: $1 \geq \mu(D_j)(j=1,\cdots,4) \geq 0$, and $\mu(D_1(\text{Very High})) \geq \mu(D_2(\text{High})) \geq \mu(D_3(\text{Low})) \geq \mu(D_4(\text{No}))$;

(2) Rule weights $\theta_k(k=1,\cdots,48)$: $1 \geq \theta_k(k=1,\cdots,48) \geq 0.01$, here we set the lower bound of rule weight to be 0.01, because we want to keep each rule's weight to be larger than 0 after training to ensure that each simulated patient in the test set could be diagnosed with trained rules in the BRB. For example, if there is a patient in the test set whose data is 100% matched to one clinical rule, the patient would not be diagnosed by the system in the performance test process if the clinical rule's weight is trained to be 0, which means the rule is of no importance in diagnosis. To avoid such situations, we set 0.01 as a lower bound to rule weight, and 0.01 can represent low importance of the rule while the rule with trained weight 0.01 can still be activated in the performance test process. We have tried other similar values as low bounds of rule weight, and it actually makes no difference to the system's performance after training if it is set to be other similar small values.

(3) Antecedent attribute weights $\delta_i(i=1,\cdots,5)$: $1 \geq \delta_1$ (ECG status) ≥ 0.5, and $1 \geq \delta_i(i=2,3,4,5) \geq 0$, here we set the weight of the ECG status between 0.5 and 1, because we know the ECG status is a very important risk factor in daily clinical risk

assessment of CCP, and this constraint can help to make the importance of ECG status not be less than half of other antecedent attributes' importance in diagnosis process.

(4) Consequent belief degrees $\beta_{kj}\left(k=1,\cdots,48; j=1,\cdots,4\right)$: $1 \geq \beta_{kj}\left(k=1,\cdots,48; j=1,\cdots,4\right) \geq 0$, and $1 \geq \sum_{j=1}^{4} \beta_{kj}\left(k=1,\cdots,48\right) \geq 0$.

What follows is discussion about the results generated from the training module validation.

6.6.2. Results

In this Section, the comparison of BRB model in the system before and after training is discussed in Section 6.6.2.1. Comparison of the system's performance on test set before and after BRB training with different training parameters is discussed in Section 6.6.2.2.

6.6.2.1. Comparison of the BRB before and after Training

In the BRB training process, with the same training data set, different training parameters brought different changes to the BRB model in the system. As described in Section 6.6.1, R^1, R^2, R^3, R^4, and R^5 are used to represent BRB training with different parameters, and we will use these five symbols in the following discussion.

The severity scores of the four consequent risk levels were trained once in training R^1, and the comparison of the severity scores before and after training R^1 is shown in Table 6-8.

For the BRB model, values of the antecedent attribute weights, rule weights, and consequent belief degrees before and after each training round are shown in Table 6-9, Table 6-10 and Table 6-11 respectively. In Table 6-8, Table 6-9, Table 6-10 and Table 6-11, values in column 'Initial' represent the initial values of associated parameters before training. Values in columns 'R^1', 'R^2', 'R^3', 'R^4' and 'R^5' represent the trained values of associated parameters after training R^1, R^2, R^3, R^4 and R^5 respectively.

Table 6-8: Severity Scores of Consequent Risk Levels Before and After Training R^1

Risk Level	Severity Score	
	Initial	R^1
Very High	1.0000	1.0000
High	0.6700	0.5977
Low	0.3300	0.4006
No	0.0000	0.0038

Table 6-9: Antecedent Attribute Weights Before and After Training R^1, R^2, and R^3

Antecedent Attribute	Attribute Weight			
	Initial	R^1	R^2	R^3
A^1	1	0.5	0.5	0.5
A^2	1	1	1	1
A^3	1	1	1	1
A^4	1	1	1	1
A^5	1	1	1	1

Table 6-10: Rule Weights Before and After Training R^1, R^2, and R^4

No.	Rule Weight				No.	Rule Weight			
	Initial	R^1	R^2	R^4		Initial	R^1	R^2	R^4
1	1	1.0000	1.0000	1.0000	25	1	1.0000	1.0000	1.0000
2	1	1.0000	1.0000	1.0000	26	1	1.0000	1.0000	1.0000
3	1	1.0000	1.0000	1.0000	27	1	1.0000	1.0000	1.0000
4	1	1.0000	1.0000	1.0000	28	1	1.0000	1.0000	1.0000
5	1	1.0000	1.0000	1.0000	29	1	1.0000	1.0000	1.0000
6	1	1.0000	1.0000	1.0000	30	1	1.0000	1.0000	1.0000
7	1	1.0000	1.0000	1.0000	31	1	1.0000	1.0000	1.0000
8	1	1.0000	1.0000	1.0000	32	1	1.0000	1.0000	1.0000
9	1	1.0000	1.0000	1.0000	33	1	0.5829	0.6754	0.7614
10	1	1.0000	1.0000	1.0000	34	1	0.8170	0.8258	0.6635
11	1	0.4153	0.4097	0.2498	35	1	0.9472	0.9396	0.9886
12	1	0.0100	0.0100	0.0100	36	1	1.0000	1.0000	0.9997
13	1	0.9986	0.9973	0.9992	37	1	0.0386	0.0119	0.0175
14	1	0.7337	0.7586	0.5464	38	1	1.0000	1.0000	0.9999
15	1	0.0100	0.0100	0.0100	39	1	0.9998	0.9995	0.9988
16	1	0.7860	0.8113	0.9173	40	1	0.9765	0.9981	0.6959
17	1	1.0000	1.0000	1.0000	41	1	0.8759	0.8887	0.8321
18	1	1.0000	1.0000	1.0000	42	1	0.0343	0.0778	0.0100
19	1	0.8134	0.7930	1.0000	43	1	1.0000	0.9999	0.9514
20	1	0.8789	0.8662	1.0000	44	1	0.9999	0.9982	0.9031
21	1	0.7496	0.7264	0.9666	45	1	1.0000	1.0000	1.0000
22	1	0.7571	0.7316	1.0000	46	1	1.0000	1.0000	0.9893
23	1	0.6853	0.6488	0.9857	47	1	1.0000	0.9994	0.9994
24	1	0.6206	0.5783	0.9969	48	1	0.9983	0.9893	0.9567

Table 6-11: Consequent Belief Degrees Before and After Training R^1, R^2, R^3, R^4, and R^5

No.	Consequent Belief Degree																							
	Very High Clinical Risk						High Clinical Risk						Low Clinical Risk						No Clinical Risk					
	Initial	R^1	R^2	R^3	R^4	R^5	Initial	R^1	R^2	R^3	R^4	R^5	Initial	R^1	R^2	R^3	R^4	R^5	Initial	R^1	R^2	R^3	R^4	R^5
1	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
2	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
3	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
4	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
5	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
6	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
7	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
8	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
9	1.0000	0.9496	0.9566	0.9771	0.9367	0.9444	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0029	0.0000	0.0034	0.0000	0.0000	0.0504	0.0405	0.0229	0.0599	0.0556
10	1.0000	0.8402	0.8393	0.8637	0.7488	0.7603	0.0000	0.0107	0.0091	0.0055	0.0516	0.0121	0.0000	0.0000	0.0012	0.0000	0.0137	0.0314	0.0000	0.1491	0.1504	0.1309	0.1859	0.1961
11	0.8889	0.8732	0.8745	0.7762	0.8731	0.7566	0.0000	0.0010	0.0149	0.0000	0.0178	0.0000	0.0000	0.0266	0.0396	0.0000	0.0127	0.0000	0.1111	0.0991	0.0709	0.2238	0.0964	0.2434
12	1.0000	0.9550	0.8969	0.5010	0.9092	0.4126	0.0000	0.0450	0.1031	0.0000	0.0696	0.0001	0.0000	0.0000	0.0000	0.0000	0.0013	0.0000	0.0000	0.0000	0.0000	0.4990	0.0200	0.5873
13	1.0000	0.8242	0.8322	0.9765	0.8161	0.9183	0.0000	0.0059	0.0076	0.0000	0.0022	0.0000	0.0000	0.0280	0.0092	0.0117	0.0471	0.0002	0.0000	0.1419	0.1511	0.0117	0.1345	0.0815
14	1.0000	0.9945	0.9891	0.9238	0.9945	0.8731	0.0000	0.0055	0.0069	0.0000	0.0007	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0040	0.0762	0.0048	0.1269
15	1.0000	0.9879	0.9850	0.6819	0.9985	0.6274	0.0000	0.0029	0.0073	0.0000	0.0015	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0092	0.0077	0.3181	0.0000	0.3726
16	1.0000	0.8258	0.8282	0.7665	0.6706	0.7095	0.0000	0.0011	0.0008	0.0000	0.0224	0.0000	0.0000	0.0000	0.0000	0.0000	0.0070	0.0000	0.0000	0.1731	0.1710	0.2335	0.3000	0.2905
17	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
18	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
19	1.0000	0.8877	0.8742	0.7538	1.0000	1.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0001	0.0000	0.0000	0.0000	0.0000	0.1123	0.1257	0.2462	0.0000	0.0000
20	1.0000	0.9218	0.9133	0.8344	1.0000	1.0000	0.0000	0.0000	0.0001	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0782	0.0866	0.1656	0.0000	0.0000
21	1.0000	0.8488	0.8328	0.6815	0.9799	0.9826	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0001	0.0000	0.0000	0.0000	0.1512	0.1672	0.3185	0.0201	0.0173
22	1.0000	0.8744	0.8578	0.7144	1.0000	1.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.1256	0.1422	0.2856	0.0000	0.0000
23	1.0000	0.8470	0.8192	0.6224	0.9927	0.9928	0.0000	0.0000	0.0002	0.0000	0.0001	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0001	0.0000	0.1530	0.1806	0.3776	0.0071	0.0071
24	1.0000	0.8417	0.8139	0.5564	0.9985	1.0000	0.0000	0.0002	0.0018	0.0000	0.0015	0.0000	0.0000	0.0004	0.0000	0.0020	0.0000	0.0000	0.0000	0.1577	0.1843	0.4417	0.0000	0.0000

Table 6-11 (Cont.): Consequent Belief Degrees Before and After Training R1, R2, R3, R4, and R5

No.	Consequent Belief Degree																							
	Very High Clinical Risk						High Clinical Risk						Low Clinical Risk						No Clinical Risk					
	Initial	R^1	R^2	R^3	R^4	R^5	Initial	R^1	R^2	R^3	R^4	R^5	Initial	R^1	R^2	R^3	R^4	R^5	Initial	R^1	R^2	R^3	R^4	R^5
25	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
26	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
27	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
28	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
29	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
30	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
31	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
32	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
33	0.2857	0.3250	0.3465	0.3877	0.4252	0.4453	0.0000	0.0000	0.0000	0.0000	0.0000	0.0001	0.0000	0.0020	0.0036	0.0000	0.0166	0.0012	0.7143	0.6730	0.6499	0.6123	0.5583	0.5534
34	0.0000	0.1221	0.1287	0.1680	0.1735	0.2659	0.0000	0.0000	0.0000	0.0000	0.0004	0.0000	0.0000	0.0030	0.0003	0.0000	0.0259	0.0012	1.0000	0.8749	0.8710	0.8320	0.8001	0.7328
35	0.3077	0.1150	0.0968	0.0000	0.1042	0.0000	0.0000	0.2191	0.2280	0.3237	0.2215	0.3487	0.0000	0.0367	0.0259	0.0061	0.0743	0.0044	0.6923	0.6292	0.6492	0.6702	0.6000	0.6469
36	0.1667	0.0676	0.0819	0.0000	0.0674	0.1133	0.0000	0.1196	0.1132	0.1907	0.1195	0.1126	0.0000	0.0813	0.0773	0.0331	0.1142	0.0007	0.8333	0.7314	0.7276	0.7762	0.6989	0.7735
37	0.2143	0.1899	0.1844	0.3329	0.1955	0.3911	0.0000	0.0200	0.0229	0.0000	0.0184	0.0000	0.0000	0.0169	0.0553	0.0000	0.0281	0.0000	0.7857	0.7732	0.7374	0.6671	0.7580	0.6089
38	0.3333	0.1307	0.1454	0.1021	0.1594	0.1281	0.0000	0.0307	0.0072	0.0523	0.0171	0.0458	0.0000	0.0276	0.0420	0.0446	0.0339	0.0287	0.6667	0.8110	0.8054	0.8010	0.7896	0.7974
39	0.1667	0.0538	0.0555	0.0000	0.0727	0.0314	0.0000	0.1897	0.1795	0.2134	0.1494	0.1810	0.0000	0.0358	0.0161	0.0031	0.0286	0.0131	0.8333	0.7206	0.7488	0.7836	0.7493	0.7744
40	0.0323	0.0358	0.0290	0.0000	0.0323	0.0520	0.0000	0.0000	0.0114	0.0369	0.0039	0.0000	0.0000	0.0000	0.0021	0.0209	0.0083	0.0001	0.9677	0.9642	0.9575	0.9422	0.9555	0.9479
41	0.4000	0.3983	0.4188	0.4277	0.4531	0.4676	0.0000	0.0385	0.0015	0.0000	0.0012	0.0000	0.0000	0.0001	0.0006	0.0000	0.0041	0.0018	0.6000	0.5631	0.5790	0.5723	0.5416	0.5306
42	0.1667	0.1438	0.1499	0.3153	0.1774	0.3581	0.0000	0.0162	0.0100	0.0000	0.0001	0.0000	0.0000	0.0014	0.0057	0.0000	0.0179	0.0000	0.8333	0.8385	0.8344	0.6847	0.8046	0.6419
43	0.1818	0.1123	0.1127	0.0663	0.1893	0.1915	0.0000	0.0857	0.0877	0.1478	0.0000	0.0000	0.0000	0.0136	0.0169	0.0335	0.0000	0.0000	0.8182	0.7884	0.7827	0.7524	0.8107	0.8085
44	0.2000	0.1552	0.1468	0.1254	0.2150	0.2197	0.0000	0.0440	0.0658	0.0976	0.0008	0.0000	0.0000	0.0173	0.0109	0.0231	0.0000	0.0000	0.8000	0.7835	0.7765	0.7539	0.7843	0.7803
45	0.1923	0.1178	0.1139	0.0565	0.1773	0.1759	0.0000	0.0981	0.0970	0.1705	0.0206	0.0173	0.0000	0.0237	0.0239	0.0381	0.0045	0.0047	0.8077	0.7604	0.7652	0.7349	0.7977	0.8021
46	0.1200	0.0480	0.0381	0.0000	0.1201	0.1211	0.0000	0.0988	0.1049	0.1707	0.0001	0.0000	0.0000	0.0227	0.0244	0.0346	0.0005	0.0000	0.8800	0.8305	0.8327	0.7946	0.8793	0.8789
47	0.1017	0.0445	0.0021	0.0010	0.0771	0.0917	0.0000	0.0456	0.1351	0.1348	0.0336	0.0091	0.0000	0.0623	0.0318	0.0260	0.0094	0.0051	0.8983	0.8476	0.8309	0.8382	0.8799	0.8941
48	0.0364	0.0161	0.0064	0.0014	0.0110	0.0352	0.0000	0.0130	0.0225	0.0482	0.0021	0.0011	0.0000	0.0251	0.0499	0.0027	0.0712	0.0013	0.9636	0.9457	0.9212	0.9477	0.9156	0.9623

From Table 6-10 and Table 6-11, we can see that the rule weights and belief degrees attached to rules numbering from 1 to 8 and numbering from 25 to 32 kept untouched in the training process. The reason is that as described in Section 6.6.1.1, there are no enough clinical cases in the simulated dataset fall in the regions that above rules were designed to operate, and then we put those scarce cases aside before drawing data into training set and test set.

Therefore, in the following system performance test over both the test set, above rules that kept untouched in the training process will not make contribution, and changes of the system's performance before and after BRB training over the test set are made by the other rules, which were trained by the training data.

6.6.2.2. System Diagnostic Performance over Test Set

This Section discusses changes of the system's diagnostic performance before and after BRB training over simulated patients in test set. Here, BRB training was conducted with R^1, R^2, R^3, R^4, and R^5. Based on the system generated overall severity scores (as discussed in Section 6.5.1) for patients in the test set before and after each BRB training, we employed SPSS to plot six different ROC curves as in Figure 6-5 to illustrate the system's performance before and after each BRB training. In Figure 6-5, the source of each curve is annotated, where *ScoreWithPreTrainedBRB* represents severity scores of patients generated by the system running with a BRB model before any training; $ScoreAfterR^1$ represents severity scores generated by the system running with a BRB model after training R^1; $ScoreAfterR^2$ represents severity scores generated by the system running with a BRB model after training R^2; $ScoreAfterR^3$ represents severity scores generated by the system running with a BRB model after training R^3; $ScoreAfterR^4$ represents severity scores generated by the system running

with a BRB model after training R^4; *ScoreAfterR*^{5} represents severity scores generated by the system running with a BRB model after training R^5. Take the blue curve in Figure 6-5 for example, from the annotated source of the curve in the figure, we know that the blue curve is based on patients' severity scores generated by the system before BRB training, and the curve can be used to illustrate the system's performance over the test set before BRB training.

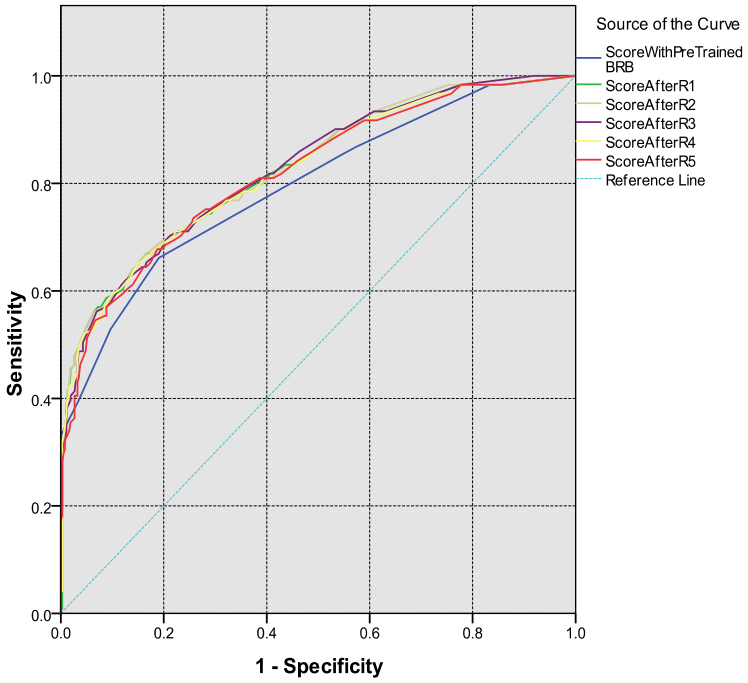


Figure 6-5: Six ROC Curves Demonstrating the Diagnostic Performance of the System before and after BRB Training over Test Set

We then required SPSS to estimate corresponding AUC values together with their 95% confidence intervals for the six ROC curves. The AUC values and their 95% confidence intervals estimated by SPSS are shown in Table 6-12, where each ROC curve is denoted by its source severity scores.

Table 6-12: AUC Values and 95% Confidence Intervals of the Six ROC Curves in Figure 6-5

ROC Curves (denoted by source scores)	AUC	95% Confidence Interval	
		Lower Bound	Upper Bound
ScoreWithPreTrainedBRB	0.7956	0.7464	0.8449
ScoreAfterR^1	0.8280	0.7830	0.8729
ScoreAfterR^2	0.8299	0.7860	0.8739
ScoreAfterR^3	0.8290	0.7851	0.8729
ScoreAfterR^4	0.8245	0.7791	0.8699
ScoreAfterR^5	0.8208	0.7752	0.8664

As seen from Table 6-12, AUC of the ROC curves representing system's diagnostic performance after BRB training are all larger than the AUC of the ROC curve representing system's diagnostic performance before BRB training.

Based on results as shown in Table 6-12, we can draw a conclusion that the training module implemented with different sets of training parameters can invariably help the system to improve the diagnostic performance, and training rounds R^1, R^2, and R^3 helped to bring better system performance than training rounds of R^4 and R^5 .

However, since the difference between the AUC of ROC curves after BRB training R^1, R^2, R^3, R^4 and R^5 is slight, we cannot tell which training has brought most significant performance improvement to the system. To test which BRB training can help the system to achieve most significant performance improvement among all conducted BRB training, we need to statistically analyze the statistical significance of the performance improvement after each BRB training round.

Thereafter, to measure the statistical significance of the system's diagnostic performance improvement, we used StAR to compare the AUC values of RUC curves representing system's performance before BRB training and after each BRB training with different training parameters, and the AUC comparison results are shown in Table 6-13.

Table 6-13: *p*-values for AUC Comparison between ROC Curves before and after BRB Training

	ScoreAfterR[1]	ScoreAfterR[2]	ScoreAfterR[3]	ScoreAfterR[4]	ScoreAfterR[5]
ScoreAfterR[2]	0.4530				
ScoreAfterR[3]	0.8863	0.8881			
ScoreAfterR[4]	0.4984	0.2038	0.5639		
ScoreAfterR[5]	0.3890	0.2326	0.0984	0.5596	
ScorePre TrainedBRB	0.0323	0.0198	**0.0076**	0.0764	0.0836

Based on the comparison results as shown in Table 6-13, we know that training R^3 (with *p*-value 0.0076) brought the most statistically significant performance improvement for the system though there is no significant difference between system performances after training rounds of R^1, R^2, R^3, R^4, and R^5 . Thus we considered the parameter set that was used in training R^3 as the most suitable training parameter set for the training module, and it is composed of antecedent attribute weights $\delta_i \left(i = 1, \cdots, 5\right)$ and consequent belief degrees $\beta_{kj} \left(k = 1, \cdots, 48; j = 1, \cdots, 4\right)$.

6.7. Summary

This chapter presents a validation study of the developed CDSS prototype. The system validation was conducted using a set of 1000 simulated patients in CCP. The BRB constructed in the system for validation is based on rules for assessing clinical risk of CCP provided by one of our research collaborators – Dr. Richard Body at MRI, and the statistical method we used for diagnostic performance evaluation is ROC analysis. There are two main conclusions that can be drawn from the validation study.

- Firstly, based on the RIMER methodology, the developed prototype CDSS can handle different uncertainties in both clinical domain knowledge and clinical data, and the system can provide reliable diagnosis recommendations.

- Secondly, based on the BRB optimization model implemented in the system, the system's performance can be statistically significantly improved after training BRB with available accumulated cases. While the most suitable training parameters for the training module contain antecedent attribute weights and belief degrees.

Based on the developed CDSS prototype and the system validation study presented in this chapter, we can conclude that it is feasible, viable, and reliable to use RIMER for implementing a CDSS.

Chapter 7

Conclusions and Potential Future Research

7.1. Introduction

This chapter first summarises the whole study, pointing out what has been done in the study in Section 7.2. Then Section 7.3 recapitulate the whole thesis and highlights main findings and contributions of the research, where the main findings from previous chapters are discussed in Section 7.3.1 and main contributions of the research are discussed in Section 7.3.2. Finally, limitations of the research and possible future research are discussed in Section 7.4

7.2. Summary of the Study

Motivated by the strong need in CDSSs research for a competent CDSS, which can (a) represent and reason with clinical domain knowledge under uncertainties; (b) update knowledge base automatically based on accumulated clinical cases; and (c) provide online group clinical decision support, the study aimed to use a newly developed belief rule-base inference methodology – RIMER (Yang et al., 2006) for the design and development of an online intelligent group CDSS. Main research questions that the study tries to answer include: (a) is it feasible to employ RIMER for developing a CDSS? (b) how to facilitate online group clinical decision making and arrive at a group combined clinical recommendation for target patient in a belief rule-based CDSS? and (c) how to train belief rule-based CDSS and make its knowledge base be adaptive to clinical practice? Based on theses research questions, the measurable objectives of the research include: (a) investigate existent CDSSs, and identify system

features of existing CDSSs; (b) acquire target clinical domain knowledge; (c) investigate the feasibility of employing belief rule base (BRB) to model clinical domain knowledge and using the evidential reasoning (ER) approach to do clinical inference in a CDSS; (d) design and develop an online belief rule-based group CDSS prototype; and (e) validate the online intelligent CDSS prototype using clinical cases in target clinical areas. Clinical areas being investigated in the study includes upper Gastrointestinal (GI) bleed and CCP. Finally, a belief rule-based online group CDSS, which provides guideline-based individual diagnosis support, group consultation support, and automatic knowledge base updating via learning through accumulated clinical cases, was developed in the research. The prototype CDSS has been validated using a set of simulated clinical cases in CCP.

The research methodology used in the study is a multiple-methodology approach. Modelling and prototyping are the two main research methods used for prototype design and development. Field study is used in the study for gaining deep understanding of domain knowledge and daily clinical work flow in NHS hospitals, and acquisition of users' requirements of a CDSS. Statistical techniques including the receiver operating characteristic (ROC) curve analysis and the area under the ROC curve (AUC) comparison are used in the prototype validation for system performance analysis.

In the research, various research gaps in the CDSSs literature that impedes successfully application of existent CDSSs in clinical practice were identified first. Then a preliminary study on the feasibility of using RIMER for developing a CDSS was conducted. It is followed by the design and development of a belief rule-based online group CDSS which can help address the identified research gaps, and finally

the validation of the developed prototype was conducted using a set of simulated patients' data.

The research developed an online belief rule-based group CDSS and proved that (a) it is feasible and viable to use RIMER for developing a CDSS; (b) the developed CDSS can handle uncertainties in both clinical domain knowledge and clinical data, and the system can provide reliable diagnosis recommendations; and (c) the BRB in the system can be updated automatically by learning through available cases accumulated in clinical practice, and the BRB training can help to improve the system's diagnostic performance statistically significantly.

In a word, all the research questions presented in Section 1.2 of Chapter 1 have been addressed comprehensively by the study.

7.3. Major Conclusions

7.3.1. Findings

This thesis has proposed using RIMER for developing an intelligent CDSS that can make use of the uncertainty-handling capability of RIMER for representing and reasoning with clinical domain knowledge under uncertainties. A preliminary feasibility study proved that it is logically feasible to employ RIMER for developing a CDSS. In order to demonstrate the technical feasibility of the proposed belief rule-based CDSS, an online CDSS prototype was developed using Visual Studio 2003 .NET and MATLAB. A set of patients in CCP was simulated by an expert clinician in MRI for validating the developed prototype. The major findings addressing the research questions are outlines as follows.

- The developed prototype CDSS shows the feasibility and viability of using RIMER for developing an online intelligent CDSS.

- Representing clinical domain knowledge with belief rules and inference with BRB using the ER approach enables the system to handle uncertainties in both clinical domain knowledge and clinical data.

- The group decision supporting module implemented in the prototype enables clinicians to hold online group meetings, discussions or consultations via the system. The ER-based group preferences aggregation mechanism in the module can help to arrive at a group combined diagnosis recommendation.

- The BRB training module implemented in the system helps the system to update the knowledge base automatically by learning through available accumulated cases, and it helps to improve the system's performance after learning from accumulated cases. Therefore, the training module enables the system to be adaptive to clinical practice and provide an evidence-based clinical decision support.

- Structuring and storing BRB in relational database facilitates the interactions between knowledge base and other system components. It also facilitates the sharing of domain specific knowledge due to the mature database technology and networking technology.

- The system validation offers encouraging outcomes for the system. Firstly the system can provide reliable diagnosis recommendations under clinical uncertainties. Secondly automatic BRB training using accumulated clinical cases can help the system improve diagnostic performance statistically significantly.

7.3.2. Contributions

The research questions pursued in this study are new, creative, and important in CDSSs research fields. The research is quite complex and demanding, as it is interdisciplinary and mainly involves (a) investigation of the existent CDSSs; (b) investigation of clinical domain knowledge; (c) investigation of advanced models for representing and reasoning with clinical domain knowledge under uncertainties, group preferences aggregation, and knowledge base training; (d) system design and development; and (e) system validation. The research deals with theoretical investigation, field study, software development, and system validation. It bridges the gaps in the CDSS literature. Major contributions of the research are listed as follows.

(1) From the CDSS research perspective:

- The research develops a new CDSS framework which integrates automatic knowledge learning functionality and online group decision supporting functionality into a knowledge-based CDSS.

- The research proposes and uses relational database to uniquely store and manage BRB model, and this makes physical knowledge base construction flexible and portable. It also makes the knowledge sharing between different clinical systems free of technology barriers thanks to mature relational database technologies.

(2) From a practical domain application perspective:

- The research develops a target clinical domain BRB for modelling domain specific knowledge under uncertainty. The BRB can be used not only for

generating automatic diagnosis recommendations but also for clinicians' future domain knowledge reference in practice.

- The research develops an ER based inference engine to infer with input uncertain clinical data and back-end uncertain domain knowledge in the BRB. The inference engine infers with different clinical uncertainties in a rational way, and can generate prioritised and informative diagnosis recommendations.

- The research develops an ER based group clinical decision supporting module. The group decision support module provides not only a group diagnosis preferences aggregation mechanism but also a discussion forum for group consultants to hold online group discussions or consultations.

- The research develops a BRB training module that can help update the embedded clinical rules automatically and routinely and keep the knowledge base being adaptive to clinical practice.

- The research implements guideline-based user interfaces which not only facilitates clinicians complying with the practice guidelines, but also makes the integration of CDSS with clinical work flow implemented easily.

7.4. Potential Future Research

Although the research shows positive and encouraging results about employing RIMER methodology for developing a CDSS, there are limitations in the research. Firstly, from technical perspective, the system developed in the research is a prototype with preliminary functionalities, and the system has not been tested with real clinical scenarios. Secondly, from system validation perspective, only inference engine and

BRB training module were validated in the research with simulated data, the group decision supporting module was not validated due to lack of data. Thirdly, about the simulated data used in system validation, though the data is close to reality, the recorded outcome of patients in the dataset is a composite one, and the dataset can not tell exactly at which level that one patient's clinical risk is. Fourthly, in clinical environment, one patient's clinical data that are necessary for clinical risk assessment may not be available at the point of risk assessment, and the risk status of one patient may keep changing, but these two situations in clinical risk assessment have not been considered in the research. Finally, through the ROC curve analysis in system validation, we know that the system can provide reliable diagnosis recommendations and the system's performance can be statistically significantly improved after BRB training, but it is difficult for doctors to accept the system and use it in clinical practice just based on current research outcomes. Then we need further research to convince the clinicians to accept the system in their clinical practice.

We know that more work needs to be done to deploy the system in a real clinical environment. Future research on the belief rule-based online group CDSS could be promising in many areas. Some of them are listed as follows.

- Apart from the uncertainty in domain knowledge and clinical data, which are represented as incomplete clinical rules in the system's BRB and uncertain subjective judgements about patients' clinical symptoms, there are other various types of uncertainties in clinical domain specific knowledge and patients' data. For example, belief degrees assigned to possible consequents of one clinical rule may be an interval other than a numerical value. Thus the BRB model in the system can be further developed to embed other possible types and degrees of

uncertainties in domain knowledge, and the ER based inference engine can be extended to accommodate these uncertainties in the inference process.

- A dynamic belief rule-based clinical inference model can be developed and integrated into the system for providing continuous and dynamic clinical decision support, considering the fact that availability of patients' clinical data is unstable and the clinical status of a specific patient may keep changing in the diagnosis process.

- The prototype can be tested by various real clinical scenarios, and real patients' clinical data in target clinical areas can be used to train and to validate the system to get more convincible results about the system.

- As for knowledge base training with accumulated clinical cases, study about more advanced BRB training techniques can be conducted for learning situations where there are no preliminary clinical rules but a large set of patients' data is available in one clinical area.

- The system can be deployed in a real clinical environment after various technical tests, and then the study about real clinical benefits that the system can bring to hospitals, doctors and patients can be conducted.

- Research about user acceptance of the system can be conducted after the system is deployed in real clinical environment, considering the fact that doctors may still have doubts about the adoption of a CDSS in their clinical practice even if the system is proven to technically facilitate clinical decision making and help reduce medical errors.

Many fruitful research can be conducted by using and enhancing the belief rule-based

online group CDSS although many challenges remain ahead.

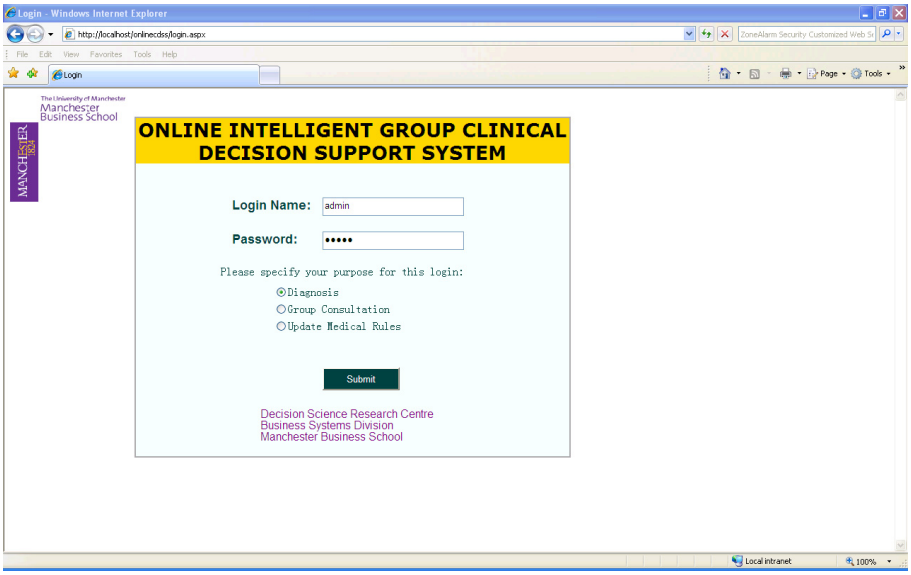
Appendix A Brief User Manual

A1. Introduction

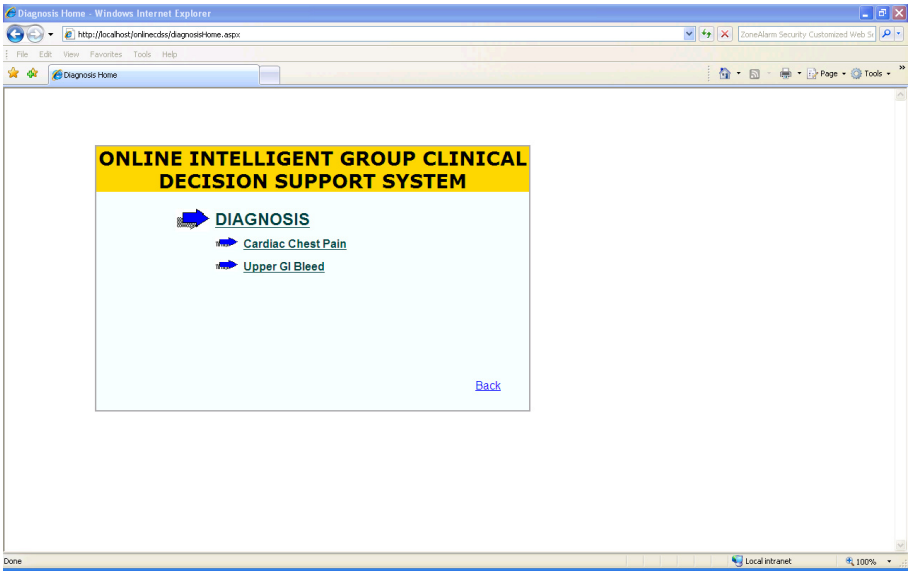
This user manual briefly describes guidance on the use of the system for various specific purposes: support of individual diagnosis of one patient; support of group consultation about one patient; and support of automatic knowledge base updating. This brief user manual uses screenshots to guide system users to main features of the system.

A2. Individual Diagnosis Support

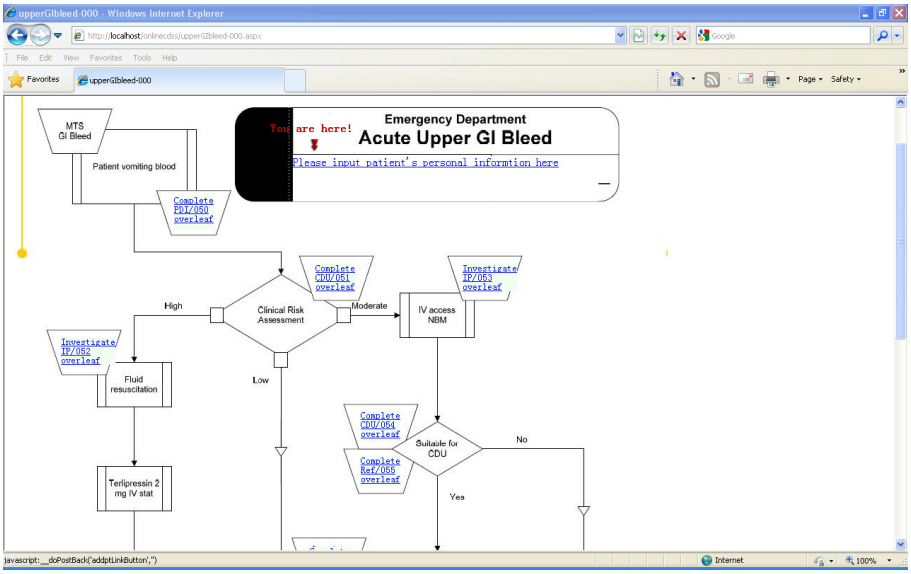
(1) Login as a clinician for the purpose of individual diagnosis.



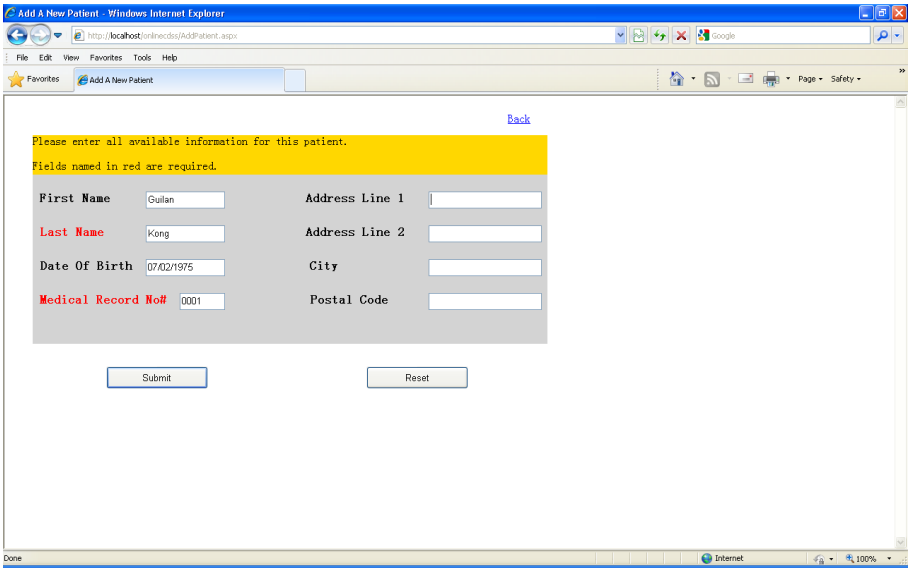
(2) Click on the main symptom of the patient. Note that we use upper gastrointestinal (GI) bleed as an example diagnosis for illustration in this manual.



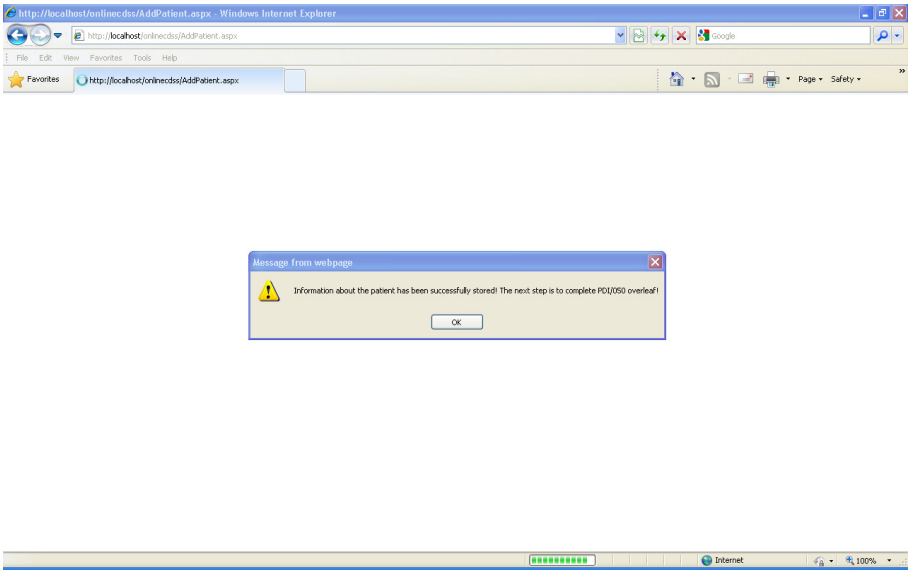
(3) Click on 'Please input patient's personal information here'.



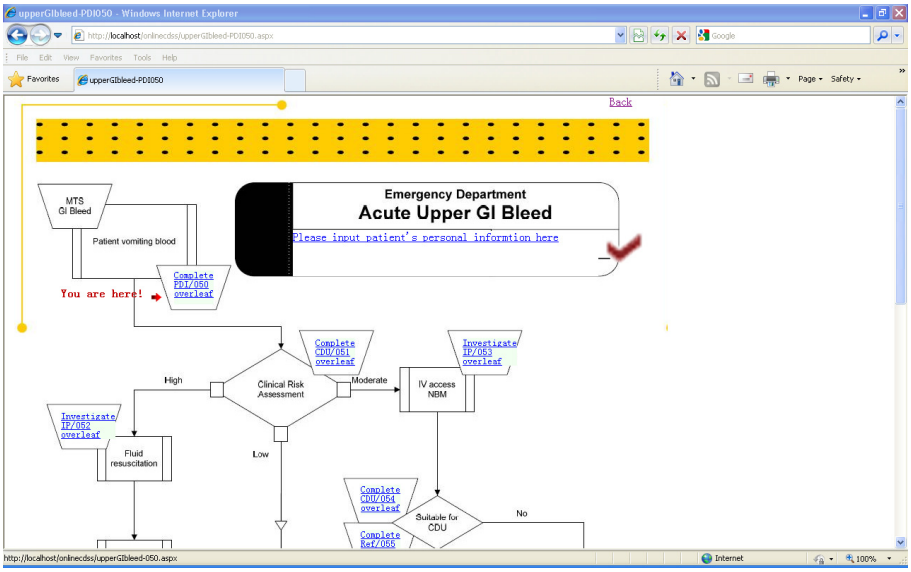
(4) Input the patient's personal information.



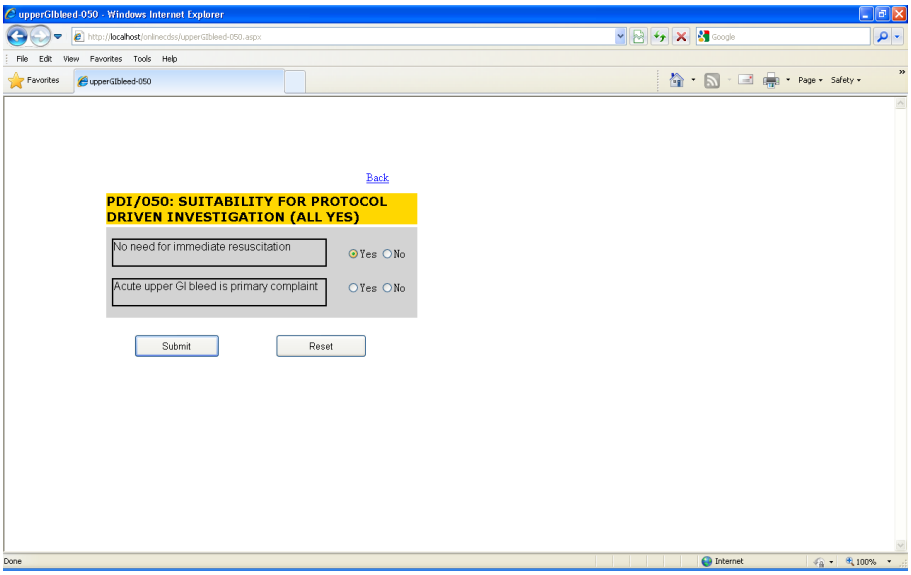
(5) Click 'OK' to go back to the main interface for diagnosis of upper GI bleed.



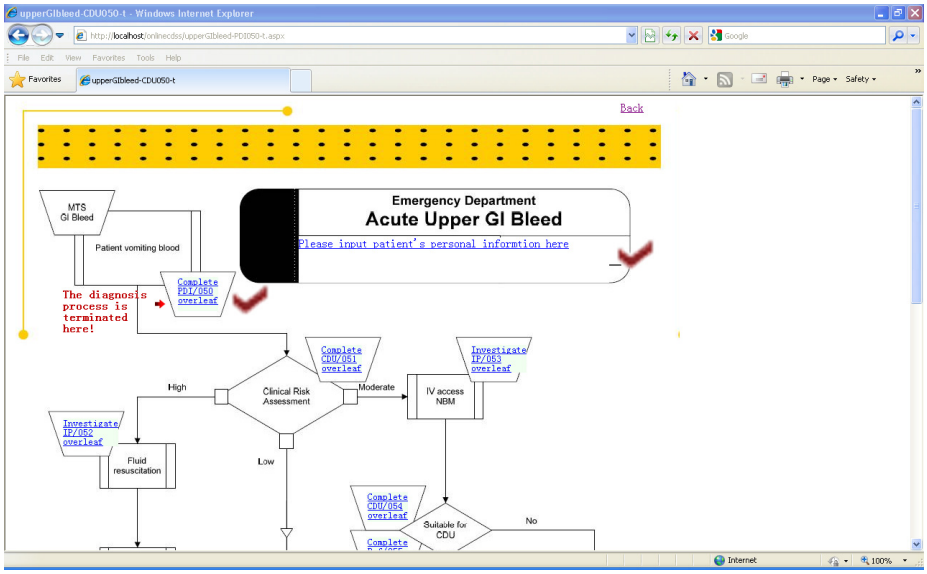
(6) Click on 'Complete PDI/050 overleaf".



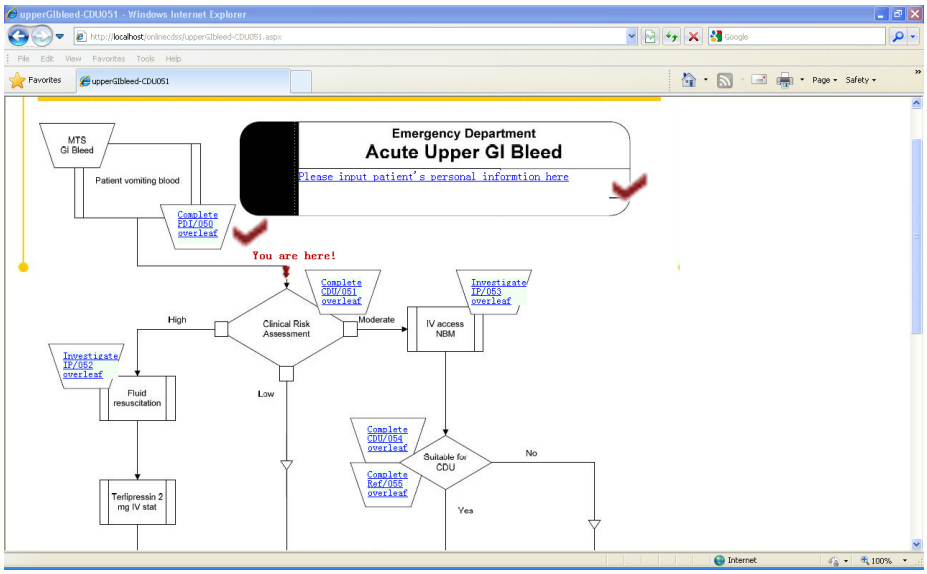
(7) Input your judgment about the patient.



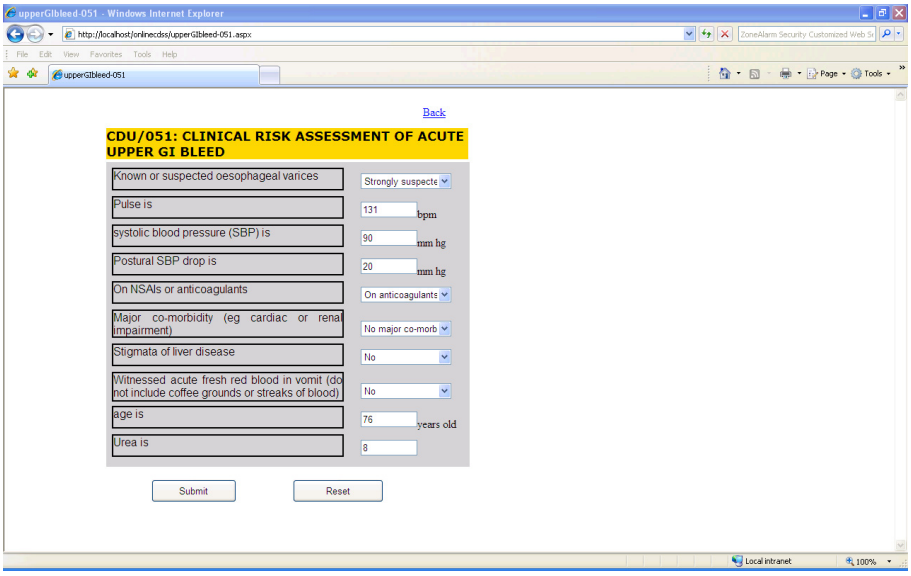
(8) Clinical decision support terminates here if the patient is not suitable for the guideline-driven investigation.



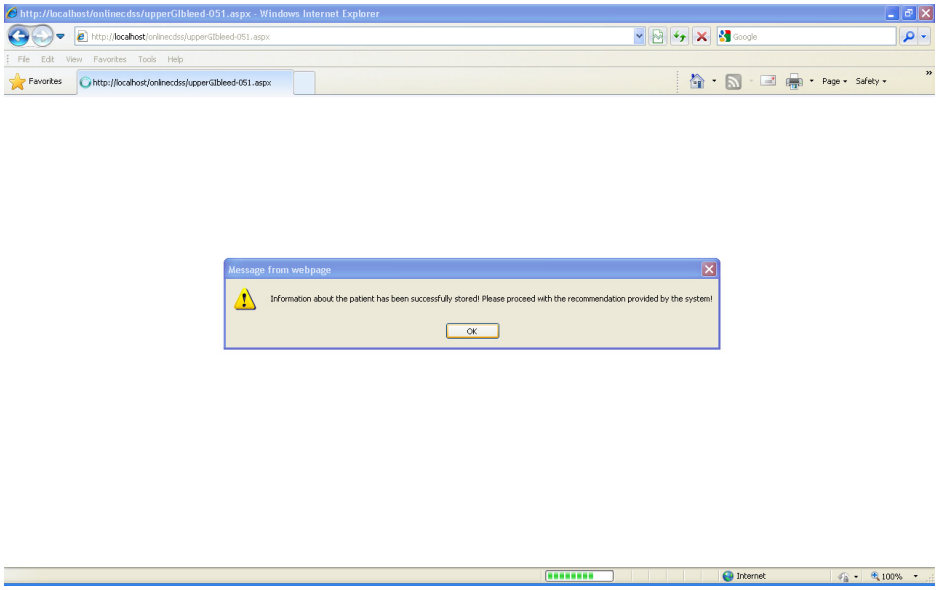
(9) Or, clinical decision support continues if the patient is suitable for the guideline-driven investigation. Click on 'Complete CDU/051 overleaf'.



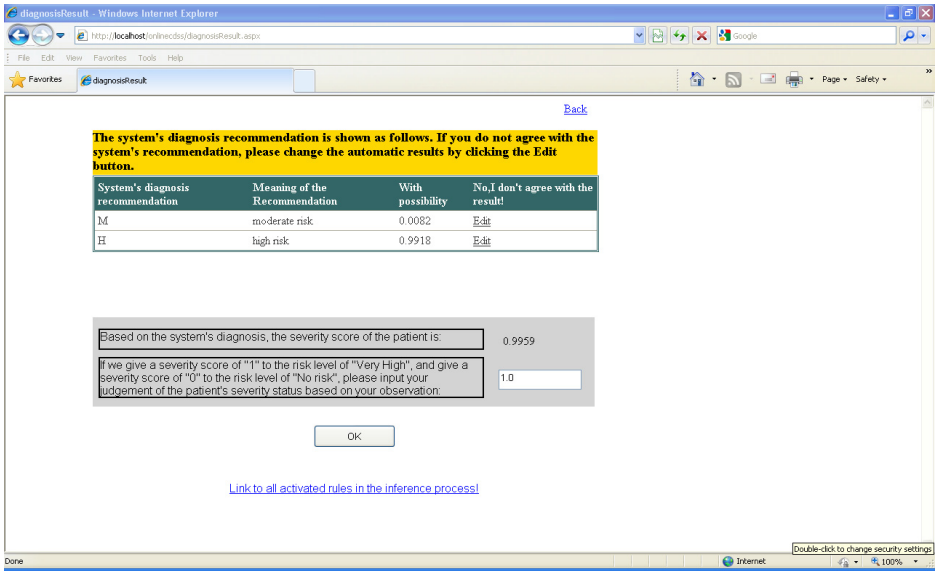
(10) Input the patient's clinical information for risk assessment.



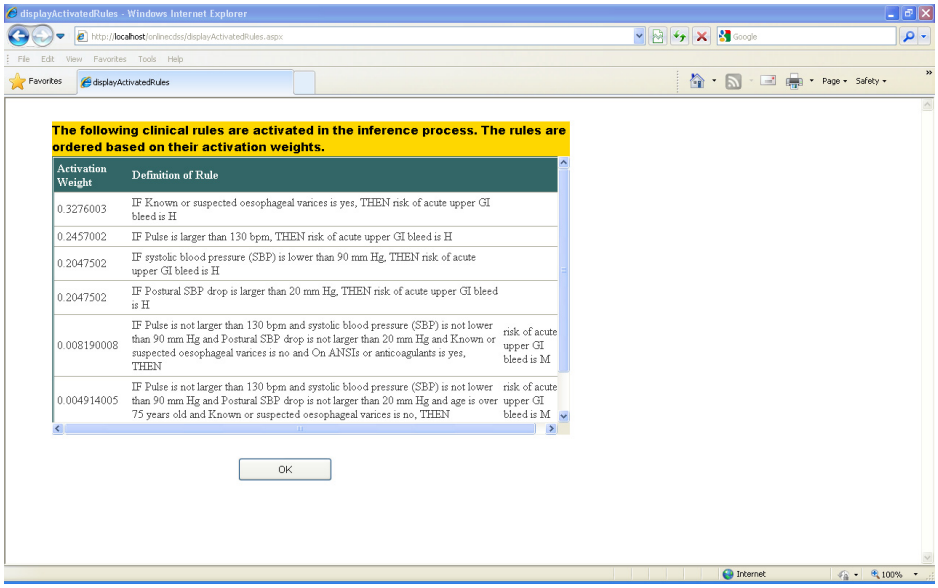
(11) Click on 'OK' to check the risk assessment result for the patient.



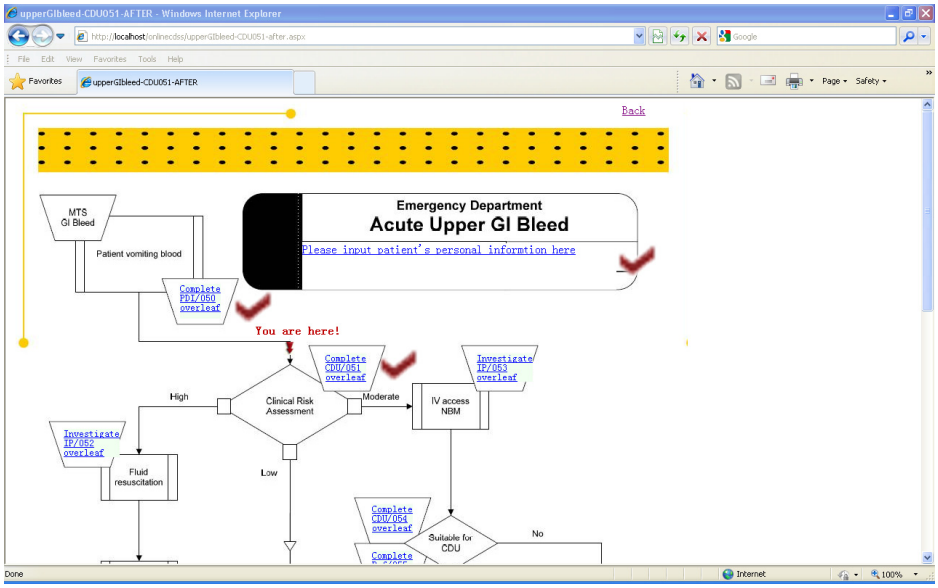
(12) Click on 'Check it!' to get the patient's severity score. Input your judgment about the patient's severity score. Click on 'Link to all activated rules in the inference process!' to check all rules activated in the process of assessing the patient's risk.



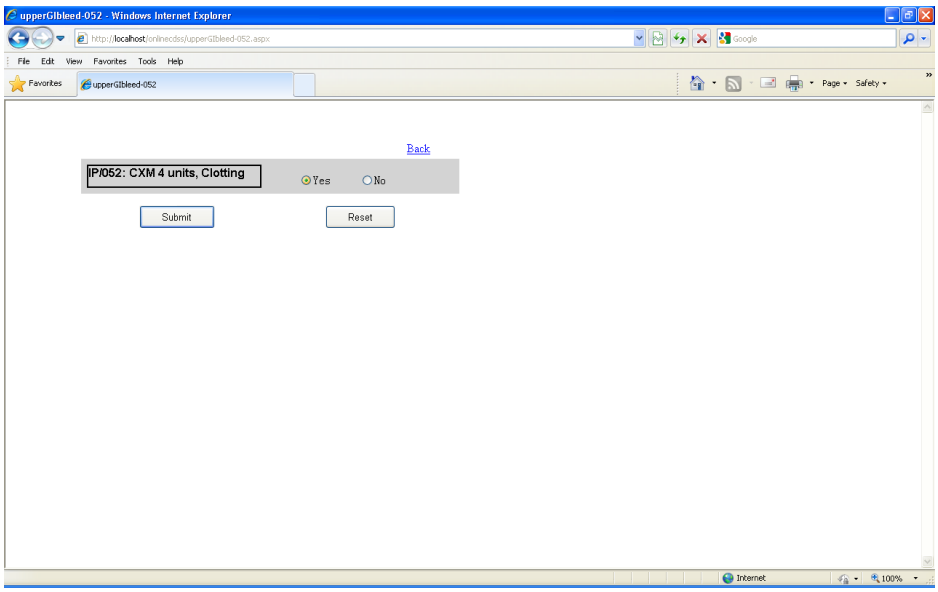
(13) All activated rules are displayed.



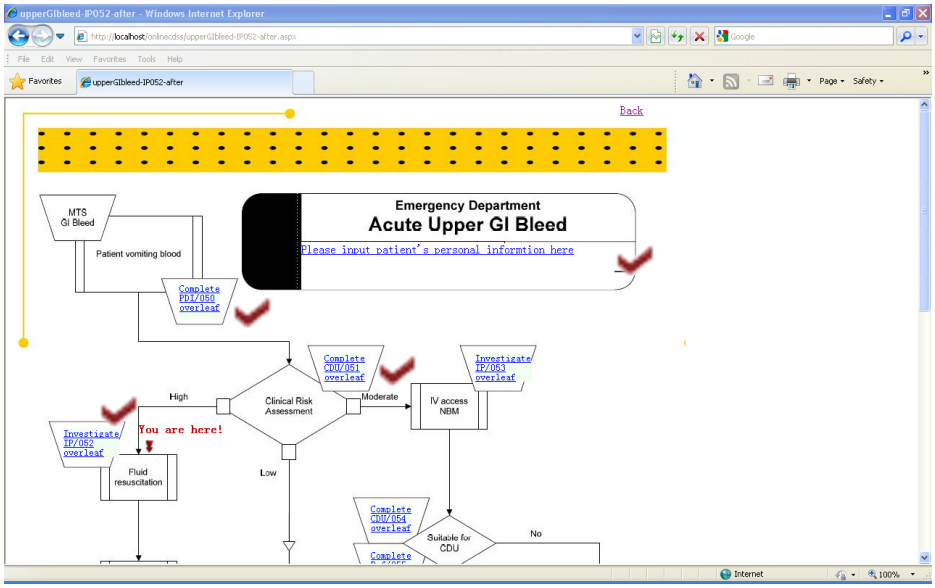
(14) Go back to the main diagnosis interface. Click on ‘Investigate IP/052 overleaf’. Here we use clinical management of patient in ‘High risk’ for illustration in this manual.



(15) Select the clinical decision about the patient.



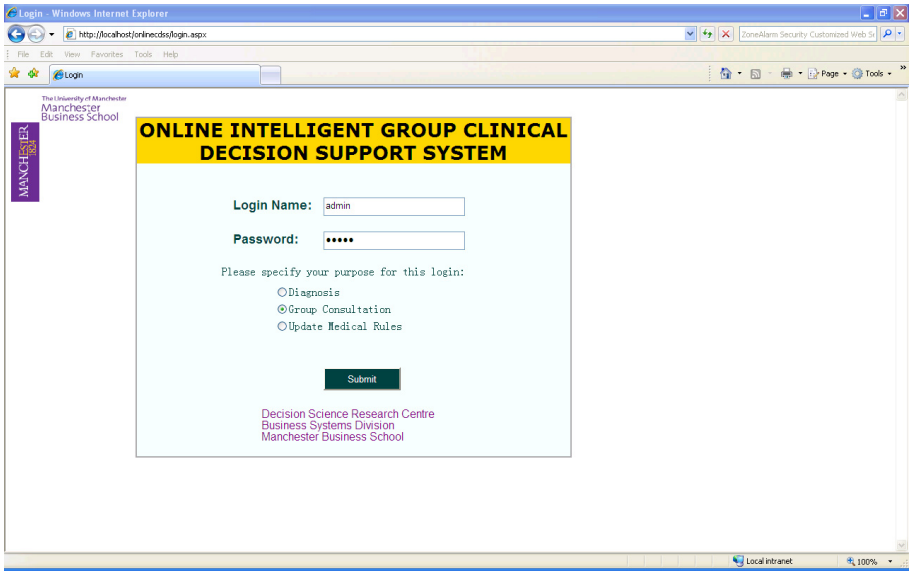
(16) What to do next is to follow the 'High risk' branch in the guideline and manage the patient as the guideline indicated step by step.



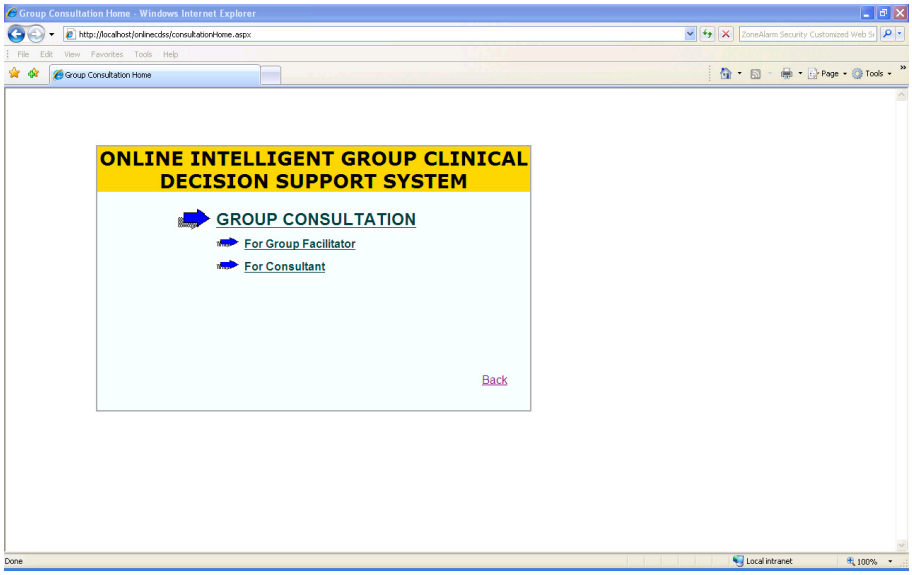
A3. Group Consultation Support

(1) For group facilitator

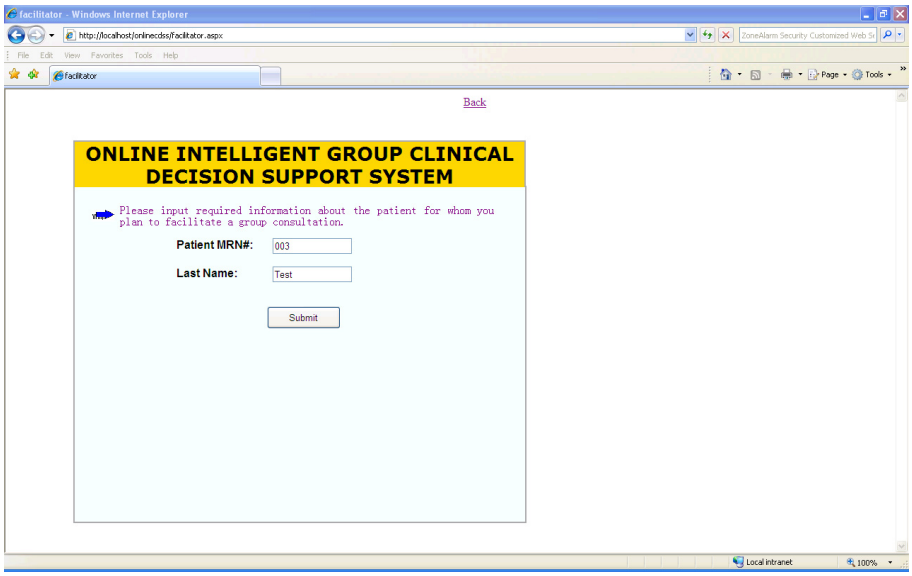
a. Login as a clinician for the purpose of group consultation.



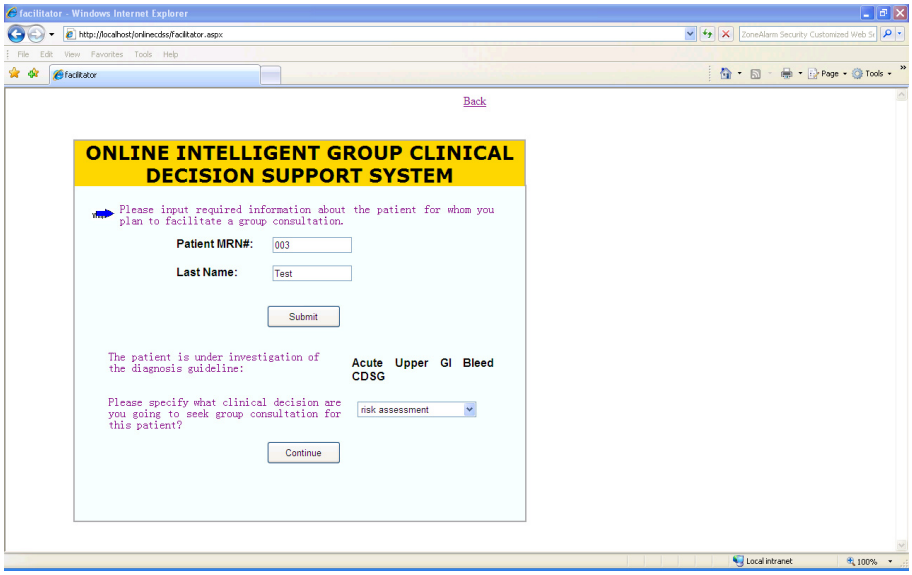
b. Click on 'For Group Facilitator'.



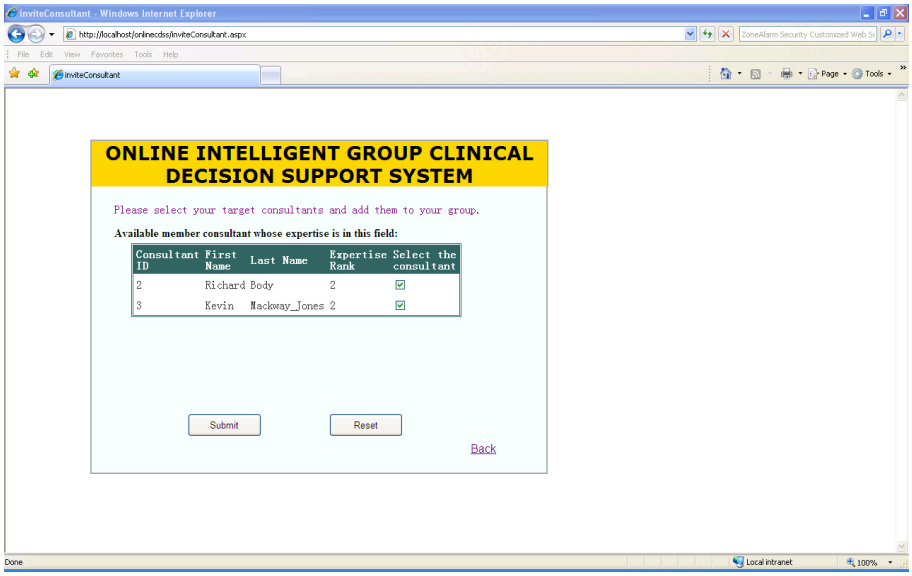
c. Input necessary information about the patient who is the target of the group consultation.



d. Select the purpose of the group consultation from the listbox.

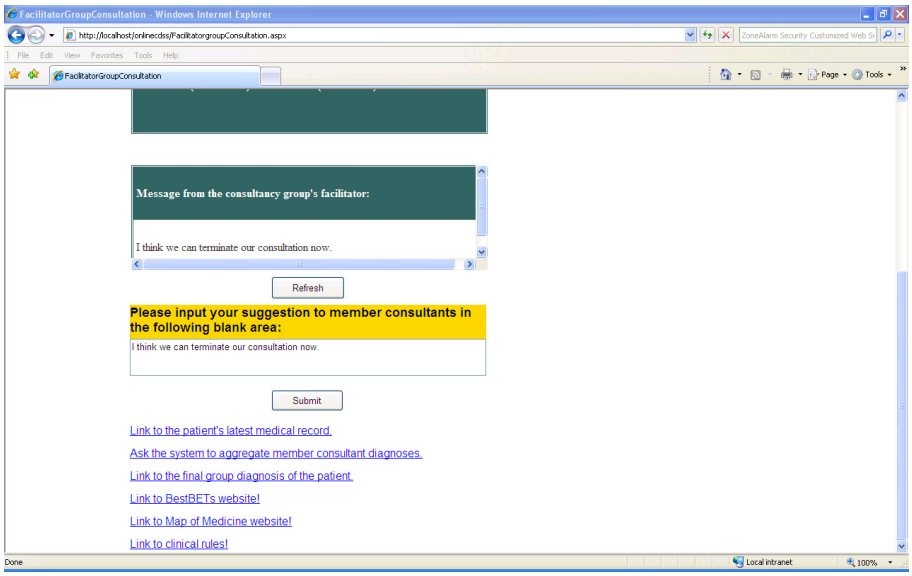


e. Select available consultants whose expertise is in target area.

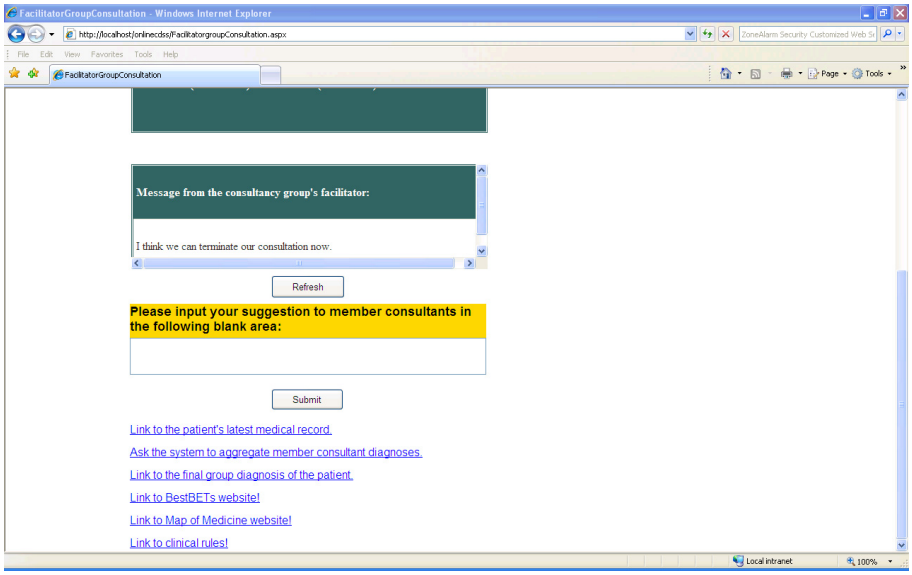


f. Wait online for group consultants' discussion and diagnosis preferences.

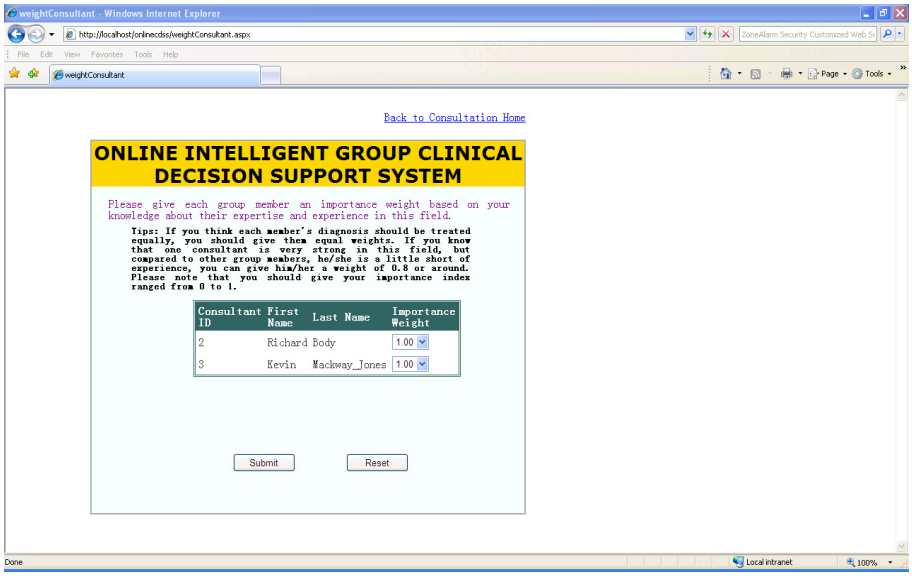
g. Facilitate the group consultation by inputting your judgment in the textbox.



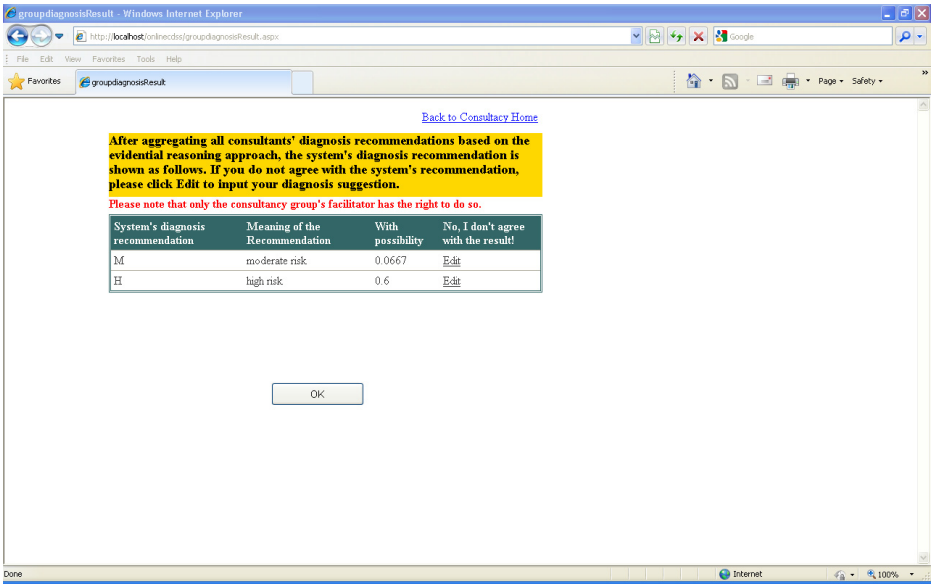
h. Click on 'Ask the system to aggregate member consultants' diagnoses' after you have terminated the group consultation.



i. Assign weight to group consultants in the group consultation.

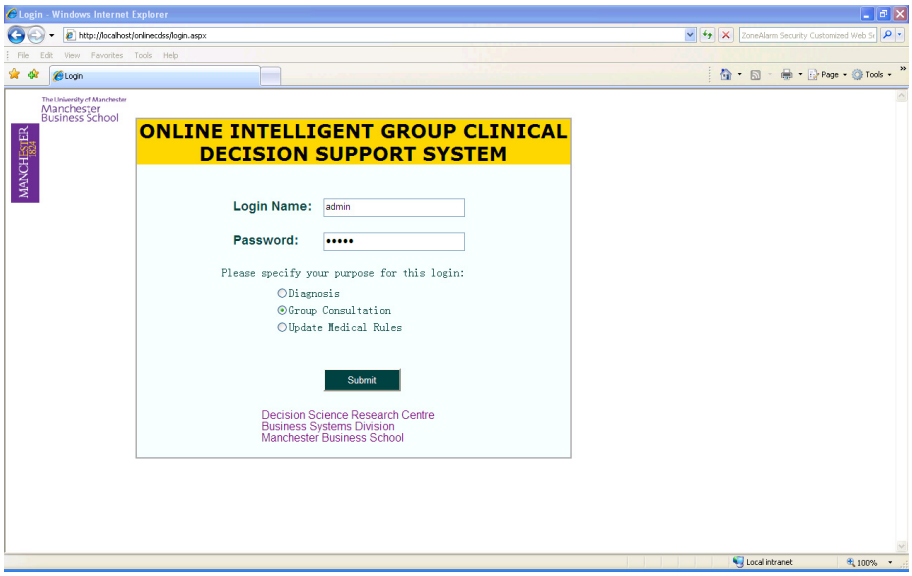


j. Check the aggregated risk result for the patient. Click on 'Edit' to input final judgment for record purpose.

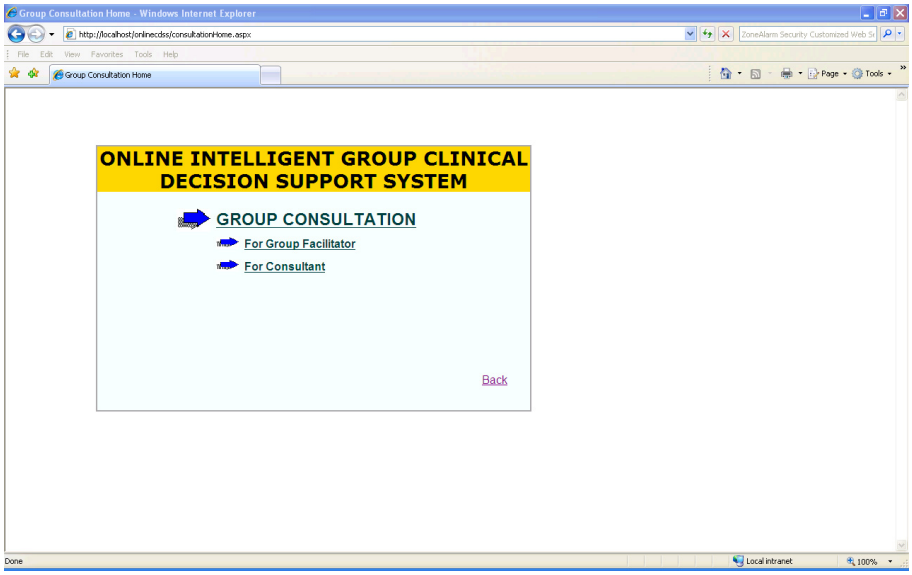


(2) For consultant

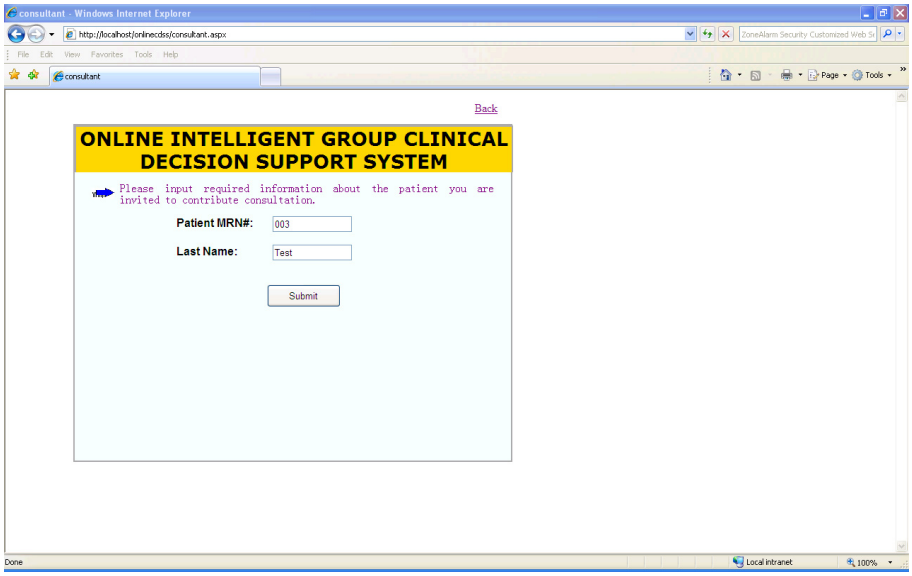
a. Login as a clinician to do group consultation.



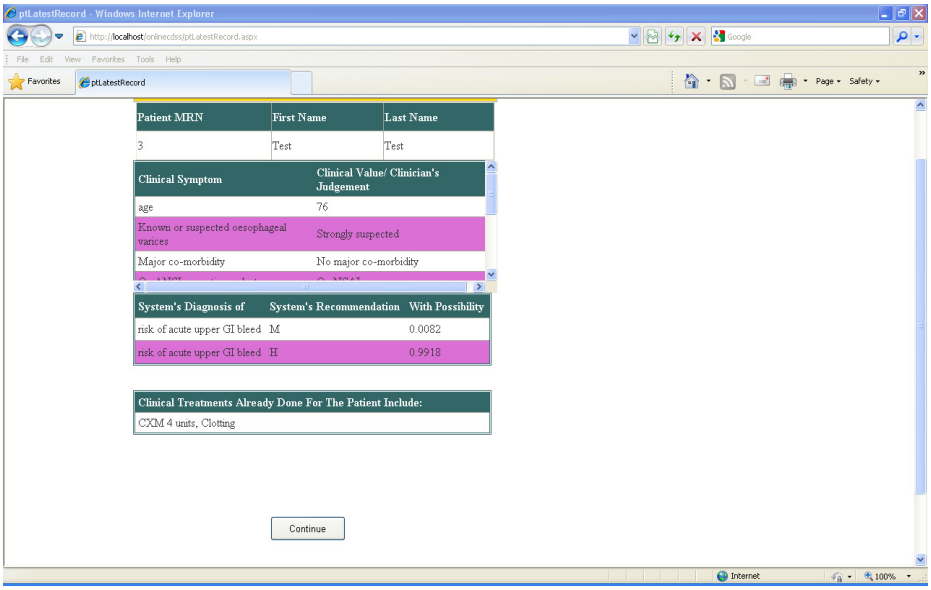
b. Click on ‘For Consultant’.



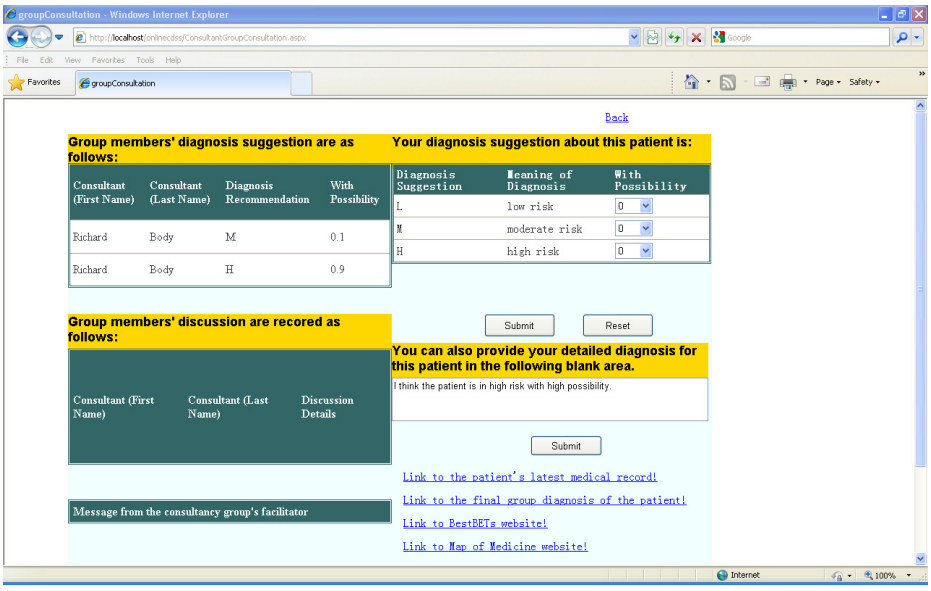
c. Input necessary information about the target patient.



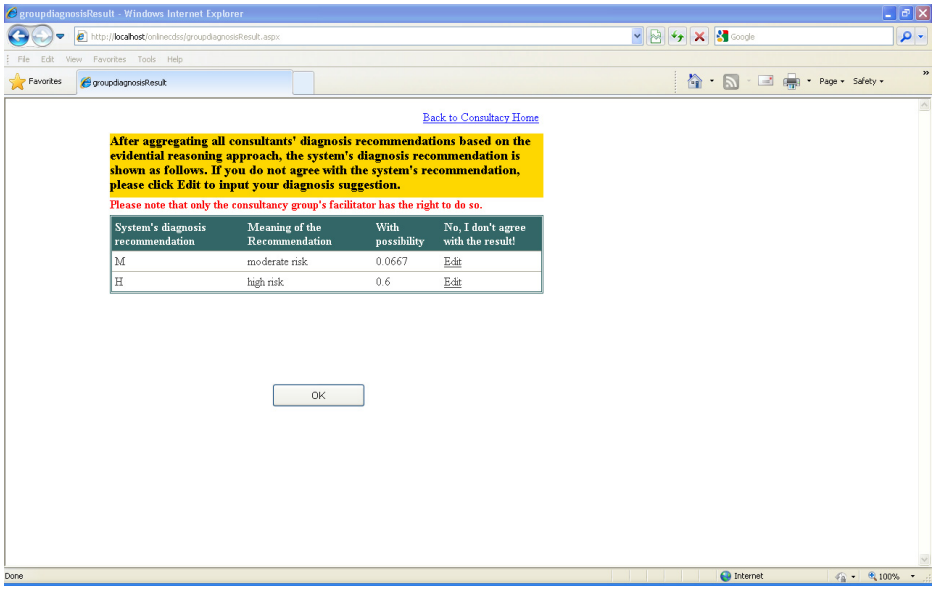
d. Check the recorded information about the patient.



e. Select your preference from the listbox about the patient's risk level. Enter your judgment about the patient's in the textbox. Click on 'Link to the final group diagnosis of the patient!' after receiving the message from the group facilitator for terminating the group consultation.

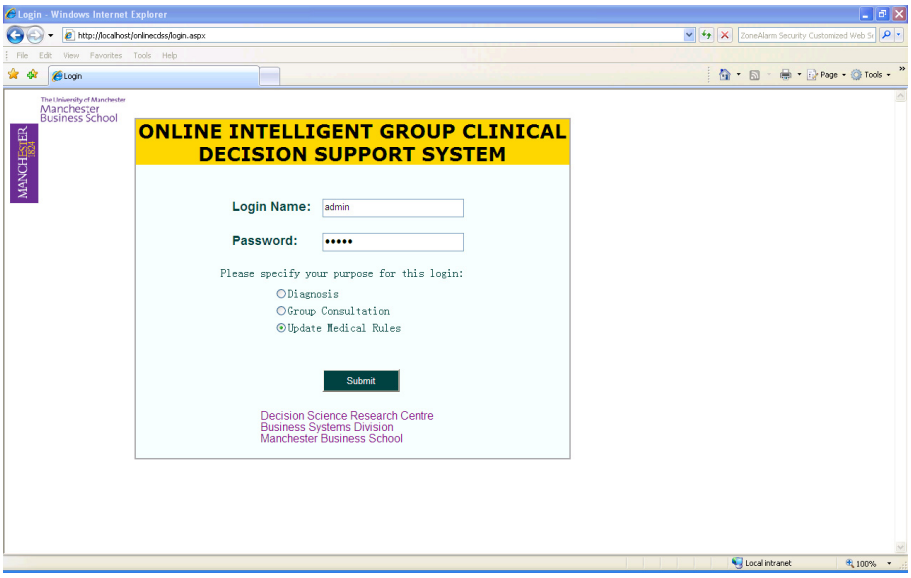


f. Check the final aggregated risk result for the patient.

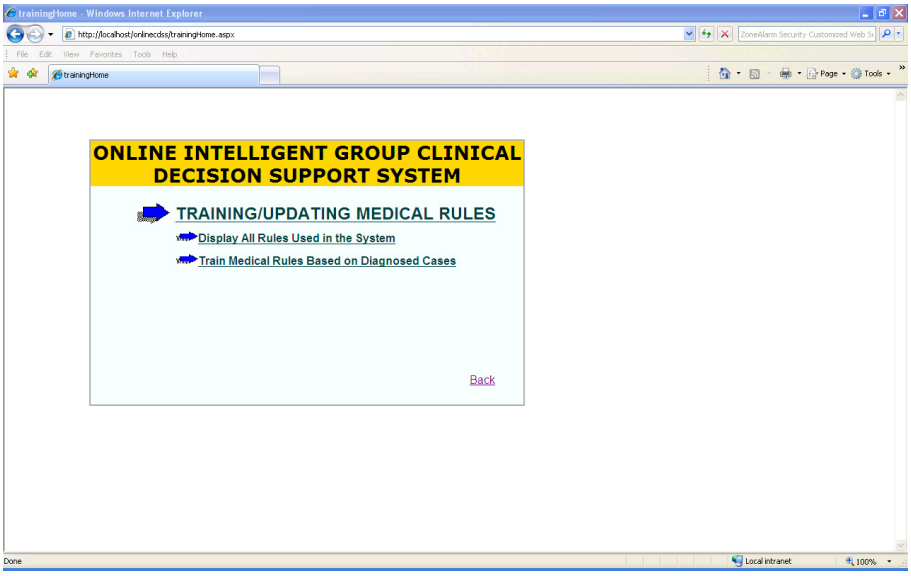


A4. BRB Updating Support

(1) Login as a clinician for the purpose of updating medical rules.

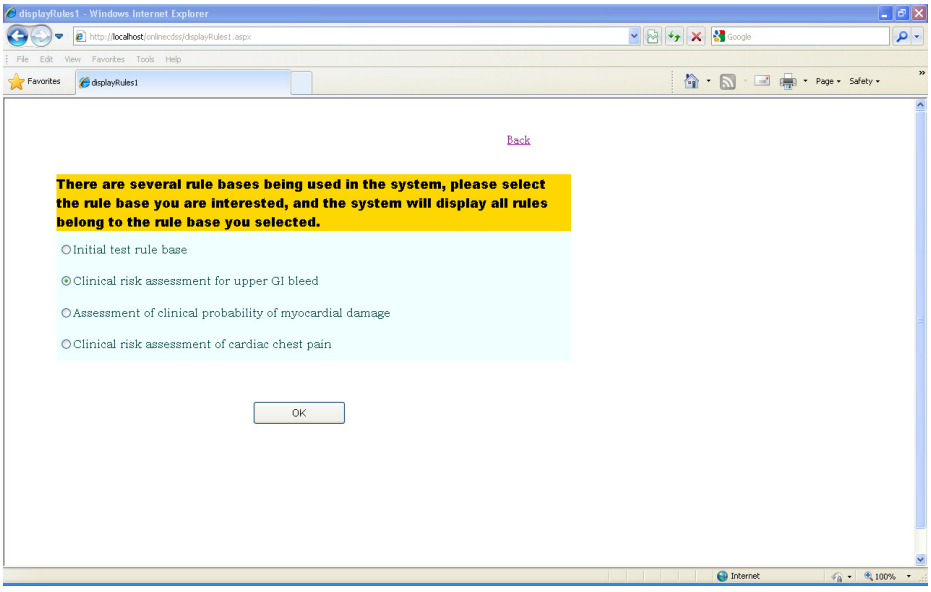


(2) Click on 'Display All Rules Used in The System' for displaying all medical rules stored in the system. Click on 'Train Medical Rules based on Diagnosed Cases' for automatically training medical rules using accumulated cases in the system.

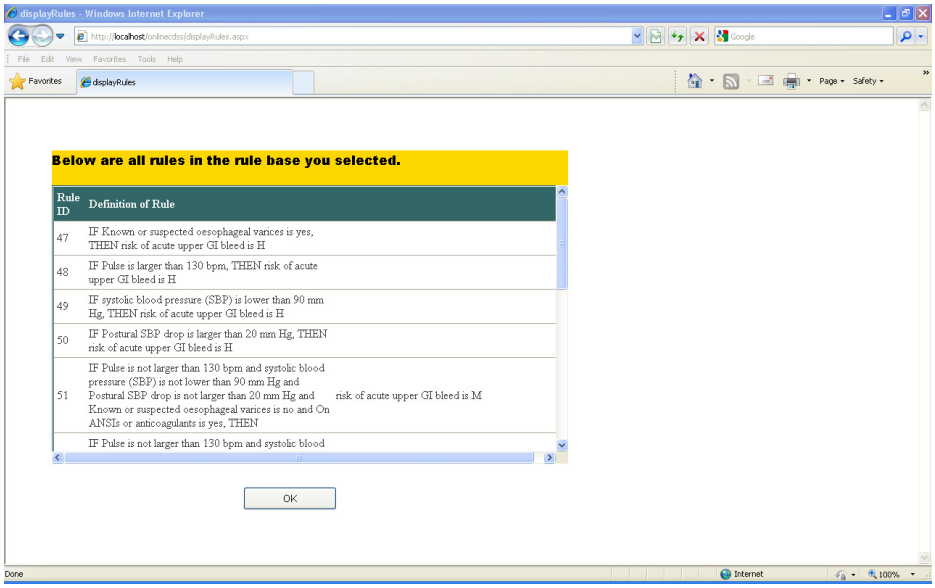


(3) For displaying medical rules

a. Select the target rule base for displaying.

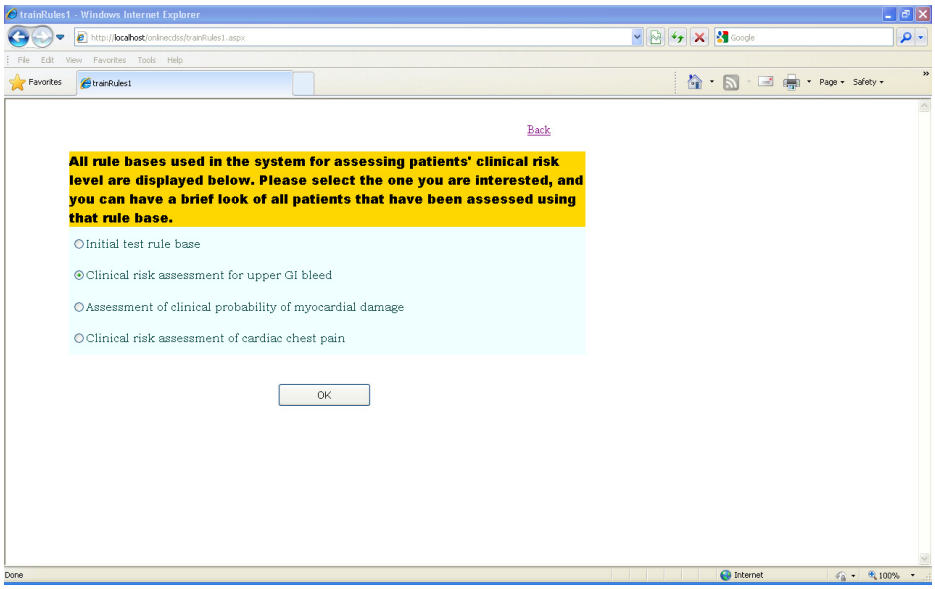


b. Check all rules stored in the rule base.

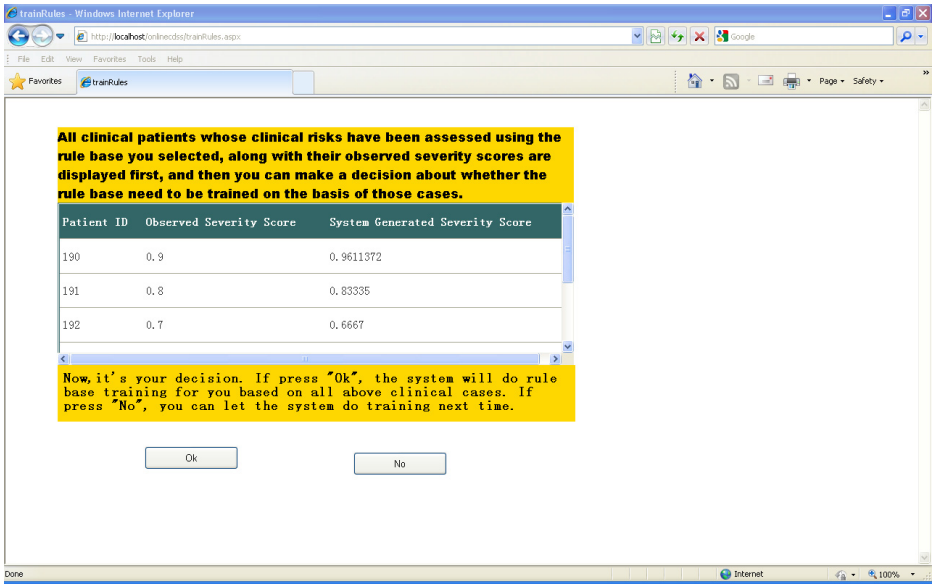


(4) For automatically updating medical rules

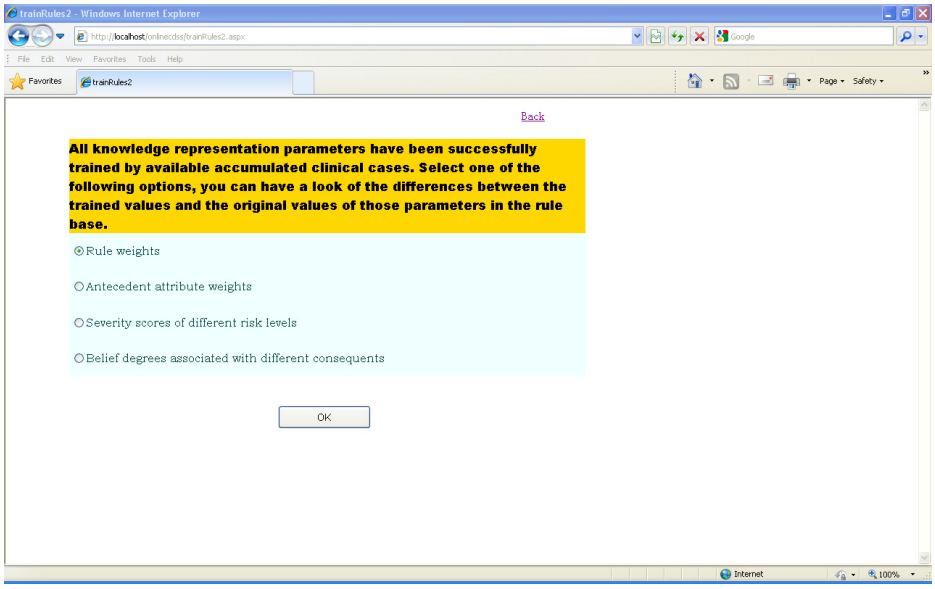
a. Select target rule base for training.



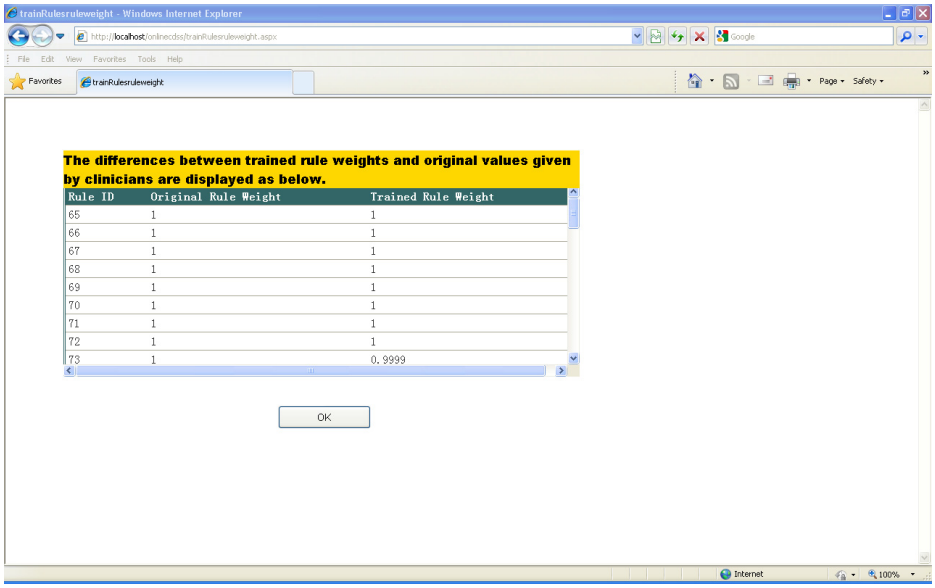
b. Click on 'OK' to train selected rule base using accumulated clinical cases.



c. Select the training parameters you want to check after the training. We use rule weights as an example for illustration in this manual.



d. Trained rule weights are displayed.


trainRulesruleweight - Windows Internet Explorer
http://localhost/onlinecdss/trainRulesruleweight.aspx
File Edit View Favorites Tools Help
Favorites
trainRulesruleweight
Page Safety
The differences between trained rule weights and original values given by clinicians are displayed as below.
Rule ID
Original Rule Weight
Trained Rule Weight
65 1 1
66 1 1
67 1 1
68 1 1
69 1 1
70 1 1
71 1 1
72 1 1
73 1 0.9999
OK
Internet
100%

Appendix B BRB Inference Using the ER Approach in Recursive Format (Yang et al., 2006)

Suppose we have a BRB that has L belief rules with N possible consequents, and rule activation weights $\omega_k (k = 1, \cdots, L)$ and belief degrees in consequents $\beta_{jk} (j = 1, \cdots, N; k = 1, \cdots, L)$ of the BRB have been calculated based on observed facts. The ER approach in recursive format can be directly applied to inference with the BRB as follows.

First, transform the degrees of belief β_{jk} for all $j = 1, \cdots, N$, $k = 1, \cdots, L$ into basic probability masses using the following ER algorithm:

$$m_{j,k} = \omega_k \beta_{jk}, j = 1, \cdots, N$$

$$m_{D,k} = 1 - \sum_{j=1}^{N} m_{j,k} = 1 - \omega_k \sum_{j=1}^{N} \beta_{jk}$$

$$\overline{m}_{D,k} = 1 - \omega_k$$

$$\widetilde{m}_{D,k} = \omega_k \left(1 - \sum_{j=1}^{N} \beta_{jk} \right)$$

for all $k = 1, \cdots, L$, where $m_{D,k} = \overline{m}_{D,k} + \widetilde{m}_{D,k}$ for all $k = 1, \cdots, L$ and $\sum_{j=1}^{L} \omega_j = 1$. The probability mass assigned to the consequent set D, which is unassigned to any individual consequent, is split into two parts: one caused by the relative importance of the kth packet antecedent A^k (or $\overline{m}_{D,k}$) and the other by the incompleteness of the kth packet antecedent A^k (or $\widetilde{m}_{D,k}$).

Then, aggregate all the packet antecedents of the L rules to generate the combined

degree of belief in each possible consequent D_j in D. Suppose $m_{j,I(k)}$ is the combined degree of belief in D_j by aggregating the first k packet antecedents $\left(A^1,\cdots,A^k\right)$, and $m_{D,I(k)}$ is the remaining degree of belief unassigned to any consequent. Let $m_{j,I(1)}=m_{j,1}$ and $m_{D,I(1)}=m_{D,1}$. Then, the overall combined degree of belief β_j in D_j is generated as:

$$\left\{D_j\right\}: m_{j,I(k+1)}=K_{I(k+1)}\left[m_{j,I(k)}m_{j,k+1}+m_{j,I(k)}\times m_{D,k+1}+m_{D,I(k)}m_{j,k+1}\right], k=1,\cdots,L-1$$

$$m_{D,I(k)}=\overline{m}_{D,I(k)}+\widetilde{m}_{D,I(k)}, k=1,\cdots,L$$

$$\left\{D\right\}: \widetilde{m}_{D,I(k+1)}=K_{I(k+1)}\left[\widetilde{m}_{D,I(k)}\widetilde{m}_{D,k+1}+\widetilde{m}_{D,I(k)}\times\overline{m}_{D,k+1}+\overline{m}_{D,I(k)}\widetilde{m}_{D,k+1}\right], k=1,\cdots,L-1$$

$$\left\{D\right\}: \overline{m}_{D,I(k+1)}=K_{I(k+1)}\left[\overline{m}_{D,I(k)}\overline{m}_{D,k+1}\right], k=1,\cdots,L-1$$

$$K_{I(k+1)}=\left[1-\sum_{j=1}^{N}\sum_{\substack{t=1\\ t\neq j}}^{N}m_{j,I(k)}m_{t,k+1}\right]^{-1}, k=1,\cdots,L-1$$

$$\left\{D_j\right\}: \beta_j=\frac{m_{j,I(L)}}{1-\overline{m}_{D,I(L)}}, j=1,\cdots,N$$

$$\left\{D\right\}: \beta_D=\frac{\widetilde{m}_{D,I(L)}}{1-\widetilde{m}_{D,I(L)}}.$$

β_D represents the remaining belief degrees unassigned to any D_j. It has been proven that $\sum_{j=1}^{N}\beta_j+\beta_D=1$. The final conclusion generated by aggregating the L rules, which are activated by the actual input vector $A^*=\left\{A^k, k=1,\cdots,L\right\}$, can be represented as

$$S\left(A^*\right)=\left\{\left(D_j,\beta_j\right), j=1,\cdots,N\right\}.$$

References

Achour, S. L., Dojat, M., Rieux, C., Bierling, P. & Lepage, E. (2001) A UMLS-based knowledge acquisition tool for rule-based clinical decision support system development. *J Am Med Inform Assoc,* 8**,** 351-360.

Aczel, A. D. & Sounderpandian, J. (2005) *Business Statistics,* London, McGraw Hill.

Agarwal, S., Dugar, D. & Sengupta, S. (2010) Ranking chemical structures for drug discovery: a new machine learning approach. *Journal of Chemical Information and Modeling,* 50**,** 716-731.

Aikins, J. S. (1980) Prototypes and Production Rules- A Knowledge Representation for Computer Consultations. *Department of Computer Science.* California, Stanford University

Aikins, J. S., Kunz, J. C., Shortliffe, E. H. & Fallat, R. J. (1983) PUFF: An expert system for interpretation of pulmonary function data. *Computers and Biomedical Research,* 16**,** 199-208.

Althoff, K.-D., Bergmann, R., Wess, S., Manago, M., Auriol, E., Larichev, O. I., Bolotov, A., Zhuravlev, Y. I. & Gurov, S. I. (1998) Case-based reasoning for medical decision support tasks: The Inreca approach. *Artificial Intelligence in Medicine,* 12**,** 25-41.

Bates, D. W., Cohen, M., Leape, L. L., Overhage, J. M., Shabot, M. M. & Sheridan, T. (2001) Reducing the frequency of errors in medicine using information technology. *J Am Med Inform Assoc,* 8**,** 299-308.

Bates, D. W., Leape, L. L., Cullen, D. J., Laird, N., Petersen, L. A., Teich, J. M., Burdick, E., Hickey, M., Kleefield, S., Shea, B., Vander Vliet, M. & Seger, D. L. (1998) Effect of computerized physician order entry and a team intervention on prevention of serious medication errors. *JAMA,* 280**,** 1311-1316.

Becker, K., Thull, B., Kasmacher-Leidinger, H., Johannes, S., Rau, G., Kalff, G. & Zimmermann, H.-J. (1997) Design and validation of an intelligent patient monitoring and alarm system based on a fuzzy logic process model. *Artificial Intelligence in Medicine,* 11**,** 33-53.

Beres, J. (2003) *Sams Teach Yourself Visual Studio .NET 2003 in 21 Days,* Indianapolis, Sams Publishing.

Berlin, A., Sorani, M. & Sim, I. (2006) A taxonomic description of computer-based clinical decision support systems. *Journal of Biomedical Informatics,* 39**,** 656-667.

Berner, E. S. (1999) Testing System Accuracy. IN BERNER, E. S. (Ed.) *Clinical Decision Support Systems: Theory and Practice.* New York, Springer-Verlag.

Berner, E. S. & La Lande, T. J. (2007) Overview of Clinical Decision Support Systems. IN BERNER, E. S. (Ed.) *Clinical Decision Support Systems: Theory and Practice.* 2nd ed. New York, Springer-Verlag.

Berner, E. S. & Moss, J. (2005) Informatics challenges for the impending patient information explosion. *Journal of the American Medical Informatics Association,* 12**,** 614-617.

Berners-Lee, T., Cailliau, R., Luotonen, A., Nielsen, H. F. & Secret, A. (1994) The world wide web. *Communicaltions of the ACM,* 37**,** 76-82.

Bhargava, H. K., Power, D. J. & Sun, D. (2007) Progress in Web-based decision support technologies. *Decision Support Systems,* 43**,** 1083-1095.

Body, R. (2009) Clinical Decision Rules to Enable Exclusion of Acute Coronary Syndromes in the Emergency Department. *Faculty of Health, Psychology and Social Care.* Manchester, Manchester Metropolitan University.

Bradley, A. P. (1997) The use of the area under the ROC curve in the evaluation of machine learning algorithms. *Pattern Recognition,* 30**,** 1145-1159.

Bryson, N. (1996) Group decision-making and the analytic hierarchy process: Exploring the consensus-relevant information content. *Computers & Operations Research,* 23**,** 27-35.

Burnside, E. S., Rubin, D. L., Fine, J. P., Shachter, R. D., Sisney, G. A. & Leung, W. K. (2006) Bayesian network to predict breast cancer risk of mammographic microcalcifications and reduce number of benign biopsy results: initial experience. *Radiology,* 240**,** 666-673.

Carruthers, K. F., Dabbous, O. H., Flather, M. D., Starkey, I., Jacob, A., Macleod, D., Fox, K. A. A. & On Behalf of the, G. I. (2005) Contemporary management of acute coronary syndromes: does the practice match the evidence? The global registry of acute coronary events (GRACE). *Heart,* 91**,** 290-298.

Carter, J. H. (1999) Design and Implementation Issues. IN BERNER, E. S. (Ed.) *Clinical Decision Support Systems: Theory and Practice.* New York, Springer-Verlag.

Carter, J. H. (2007) Design and Implementation Issues. IN BERNER, E. S. (Ed.) *Clinical Decision Support Systems: Theory and Practice.* 2nd ed. New York, Springer-Verlag.

Central Manchester and Manchester Children's University Hospitals Nhs Trust (2003a) Clinical decision support guideline (CDSG) 2003-01 for “Cardiac Chest Pain”. Emergency Care Specialist Library, Central Manchester and Manchester Children's University Hospitals NHS Trust.

Central Manchester and Manchester Children's University Hospitals Nhs Trust (2003b) Clinical decision support guideline (CDSG) 2003-05 for “Acute Upper GI Bleed”. Emergency Care Specialist Library, Central Manchester and Manchester Children's University Hospitals NHS Trust.

Christensen, C. & Larson, J. R. J. R. (1993) Collaborative medical decision making. *Med Decis Making,* 13**,** 339-346.

Clancey, W. J. & Cooper, G. (1984) Uncertainty and Evidential Support. IN BUCHANAN, B. G. & SHORTLIFFE, E. H. (Eds.) *Rule-Based Expert Systems: The MYCIN Experiments of the Standford Heuristic Programming Project.* Reading, MA, Addison-Wesley.

Clayton, P. D., Pryor, T. A., Wigertz, O. B. & Hripcsak, G. (1989) Issues and Structures for Sharing Knowledge among Decision-making Systems: The 1989 Arden Homestead Retreat. IN KINGSLAND, L. C. (Ed.) *Thirteenth Annual Symposium on Computer Applications in Medical Care: Proceedings.* New York, IEEE Computer Society Press

Collins English Electronic Dictionary (2008) London, HarperCollins

Colombet, I., Aguirre-Junco, A.-R., Zunino, S., Jaulent, M.-C., Leneveut, L. & Chatellier, G. (2005) Electronic implementation of guidelines in the EsPeR system: A knowledge specification method. *International Journal of Medical Informatics,* 74**,** 597-604.

Dale, J., Popescu, L. & Karp, P. (2010) Machine learning methods for metabolic pathway prediction. *BMC Bioinformatics,* 11**,** 15.

De Clercq, P. A., Hasman, A., Blom, J. A. & Korsten, H. H. M. (2001) Design and implementation of a framework to support the development of clinical guidelines. *International Journal of Medical Informatics,* 64**,** 285-318.

De Dombal, F. T., Leaper, D. J., Staniland, J. R., Mccann, A. P. & Horrocks, J. C. (1972) Computer-aided diagnosis of acute abdominal pain. *British Medical Journal,* 2(5804) 5-9.

Delong, E. R., Delong, D. M. & Clarke-Pearson, D. L. (1988) Comparing the areas under two or more correlated receiver operating characteristic curves: A nonparametric approach. *Biometrics,* 44**,** 837-845.

Department of Health (1998) A First Class Service: Quality in the New NHS. IN HEALTH, D. O. (Ed.).

Department of Health (2001) Building a Safer NHS for Patients - Implementing an Organisation with a Memory. IN HEALTH, D. O. (Ed.).

Department of Health (2004) Seven Steps to Patient Safety. IN NATIONAL PATIENT SAFETY AGENCY, D. O. H. (Ed.).

Department of Health Expert Group (2000) An Organisation with a Memory. IN HEALTH, D. O. (Ed.), The Stationery Office.

Driscoll, P., Thomas, M., Touquet, R. & Fothergill, J. (2001) Risk Management in Accident and Emergency Medicine. IN CHARLES, V. (Ed.) *Clinical Risk Management: Enhancing Patient Safety.* Second ed. London, BMJ publishing house.

Fearn, P., Regan, K., Sculli, F., Fajardo, J., Smith, B. & Alli, P. (2007) Lessons Learned from Caisis: An Open Source, Web-Based System for Integrating Clinical Practice and Research. *Computer-Based Medical Systems, 2007. CBMS '07. Twentieth IEEE International Symposium on.*

Field, M. J. & Lohr, K. N. (1992) *Guidelines for Clinical Practice: From Development to Use.,* Washington, D.C.: Institute of Medicine, National Academy Press.

Fox, J., Johns, N. & Rahmanzadeh, A. (1998) Disseminating medical knowledge: the PROforma approach. *Artificial Intelligence in Medicine,* 14**,** 157-182.

Freund, Y., Seung, H. S., Shamir, E. & Tishby, N. (1997) Selective sampling using the query by committee algorithm. *Machine Learning,* 28**,** 133-168.

Gago, P., Silva, Á. & Santos, M. (2007) Adaptive Decision Support for Intensive Care. IN NEVES, J. M., SANTOS, M. F. & MACHADO, J. M. (Eds.) *Progress in Artificial Intelligence.* Berlin, Springer.

Gerald, L. B., Tang, S., Bruce, F., Redden, D., Kimerling, M. E., Brook, N., Dunlap, N. & Bailey, W. C. (2002) A decision tree for tuberculosis contact investigation. *Am. J. Respir. Crit. Care Med.,* 166**,** 1122-1127.

Graham, T., Bullard, M., Kushniruk, A., Holroyd, B. & Rowe, B. (2008) Assessing the sensibility of two clinical decision support systems. *Journal of Medical Systems,* 32**,** 361-368.

Grams, R. R. (1993) Clinical laboratory test reference (CLTR). *Journal of Medical Systems,* 17**,** 59-67.

Grigorova, D. & Nikolov, N. (2007) Knowledge representation in systems with natural language interface. IN RACHEV, B., SMRIKAROV, A. & DIMOV, D. (Eds.) *2007 International Conference on Computer Systems and Technologies* Bulgaria ACM Press.

Grimshaw, J. M. & Russell, I. T. (1993) Effect of clinical guidelines on medical practice: a systematic review of rigorous evaluations. *The Lancet* 342**,** 1317-22.

Grzymala-Busse, J. W. & Woolery, L. K. (1994) Improving prediction of preterm birth using a new classification scheme and rule induction. *Proceedings / the ... Annual Symposium on Computer Application [sic] in Medical Care.* American Medical Informatics Association.

Hamilton, B. & Macdonald, M. (2003) *ADO.NET in a Nutshell,* California, O'Reilly & Associates.

Hanley, J. A. (1988) The robustness of the "Binormal" assumptions used in fitting ROC curves. *Medical decision making,* 8**,** 197-203.

Hardin, J. M. & Chhieng, D. C. (2007) Data Mining and Clinical Decision Support Systems. IN BERNER, E. S. (Ed.) *Clinical Decision Support Systems: Theory and Practice.* 2nd ed. New York, Springer-Verlag.

Hastie, T., Tibshirani, R. & Friedman, J. (2001) *The Elements of Statistical Learning; Data mining, Inference, and Prediction,* New York, Springer.

Hatcher, M. (1990) Uniqueness of Group Decision Support Systems (GDSS) in medical and health applications. *Journal of Medical Systems,* 14**,** 351-364.

Hatcher, M. (1994) Voting and priorities in health care decision making, portrayed through a group decision support system, using analytic hierarchy process. *Journal of Medical Systems,* 18**,** 267-288.

Heckerman, D. E. & Shortliffe, E. H. (1992) From certainty factors to belief networks. *Artificial Intelligence in Medicine,* 4**,** 35-52.

Hernandez, M. J. (2003) *Database Design for Mere Mortals: A Hands-On Guide to Relational Database Design,* Boston, Addsion-Wesley.

Http://Protein.Bio.Puc.Cl/Cardex/Servers/Roc/Home.Php.

Http://Ruleml.Org/.

Http://Www-Radiology.Uchicago.Edu/Krl/Krl_Roc/Software_Index6.Htm.

Http://Www.Bestbets.Org/.

Http://Www.Mathworks.Com/Products/Matlab/.

Http://Www.Medcalc.Be/.

Http://Www.Spss.Com/.

Http://Www.W3.Org/Xml/.

Huang, M.-J. & Chen, M.-Y. (2007) Integrated design of the intelligent web-based Chinese Medical Diagnostic System (CMDS) - Systematic development for digestive health. *Expert Systems with Applications,* 32**,** 658-673.

Hudson, D. & Cohen, M. (1987) Management of uncertainty in a medical expert system. IN BOUCHON, B. & YAGER, R. R. (Eds.) *Uncertainty in Knowledge-Based Systems: International Conference on Information Processing and Management of Uncertainty in Knowledge-Based Systems.* Berlin, Springer.

Hudson, D. L. & Cohen, M. E. (1988) Approaches to management of uncertainty in an expert system. *International Journal of Intelligent Systems,* 3**,** 45-58.

Hudson, D. L. & Cohen, M. E. (2002) Use of intelligent agents in the diagnosis of cardiac disorders. IN BENCIVELLI, W. & PORENTA, G. (Eds.) *Computers in Cardiology.* Tennessee IEEE.

Ivan, B. (2001) *Prolog Programming for Artificial Intelligence,* New York, Harlow Addison Wesley.

Jin, H. & Ling, C. X. (2005) Using AUC and accuracy in evaluating learning algorithms. *Knowledge and Data Engineering, IEEE Transactions on,* 17**,** 299-310.

Johnston, M. E., Langton, K. B., Haynes, R. B. & Mathieu, A. (1994) Effects of computer-based clinical decision support systems on clinician performance and patient outcome. *Annals of internal medicine,* 120**,** 135-42.

Jonsbu, J., Aase, O., Rollag, A., Liestol, K. & Erikssen, J. (1993) Prospective evaluation of an EDB-based diagnostic program to be used in patients admitted to hospital with acute chest pain. *Eur Heart J,* 14**,** 441-446.

Junghans, C. & Timmis, A. D. (2006) Risk assessment after acute coronary syndrome. *BMJ,* 333**,** 1079-1080.

Kawamoto, K., Houlihan, C. A., Balas, E. A. & Lobach, D. F. (2005) Improving clinical practice using clinical decision support systems: a systematic review of trials to identify features critical to success. *BMJ,* 330**,** 765-772.

Keith, R. D. F. & Greene, K. R. (1994) 4 Development, evaluation and validation of an intelligent system for the management of labour. *Bailliere's Clinical Obstetrics and Gynaecology,* 8**,** 583-605.

Klein, M. R. & Methlie, L. B. (1995) *Knowledge-based decision support systems: with applications in business,* Chichester, John Wiley & Sons.

Kohn, L. T., Corrigan, J. M. & Donaldson, M. S. (2000) *To Err is Human: Building a Safer Health System,* Washington, DC, National Academy Press.

Kong, G. L., Xu, D. L., Liu, X. B. & Yang, J. B. (2009) Applying a belief rule-base inference methodology to a guideline-based clinical decision support system. *Expert Systems,* 26**,** 391-408.

Kong, G. L., Xu, D. L. & Yang, J. B. (2008a) Applying a new rule-base inference methodology into clinical decision making IN RUAN, D., MONTERO, J., LU, J., MARTÍNEZ, L., P. D'HONDT & KERRE, E. E. (Eds.) *The 8th International FLINS Conference on Computational Intelligence in Decision and Control* Madrid, Spain, World Scientific

Kong, G. L., Xu, D. L. & Yang, J. B. (2008b) Clinical decision support systems: A review on knowledge representation and inference under uncertainties. *International Journal of Computational Intelligence Systems,* 1**,** 159-167.

Kumar, K. A., Singh, Y. & Sanyal, S. (2009) Hybrid approach using case-based reasoning and rule-based reasoning for domain independent clinical decision support in ICU. *Expert Systems with Applications,* 36**,** 65-71.

Levin, M. (1995) Use of genetic algorithms to solve biomedical problems. *M.D.Computing: computers in medical practice,* 12**,** 193-199.

Li, Y.-C., Liu, L., Chiu, W.-T. & Jian, W.-S. (2000) Neural network modeling for surgical decisions on traumatic brain injury patients. *International Journal of Medical Informatics,* 57**,** 1-9.

Li, Y. C., Haug, P. J. & Warner, H. R. (1994) Automated transformation of probabilistic knowledge for a medical diagnostic system. *Proceedings / the ... Annual Symposium on Computer Application [sic] in Medical Care.* American Medical Informatics Association.

Liberty, J. & Hurwitz, D. (2002) *Programming ASP.NET,* California, O'Reilly & Associates.

Lin, C., Lin, C. M., Lin, B. & Yang, M.-C. (2009) A decision support system for improving doctors' prescribing behavior. *Expert Systems with Applications,* 36**,** 7975-7984.

Lin, L., Hu, P. J.-H. & Liu Sheng, O. R. (2006) A decision support system for lower back pain diagnosis: Uncertainty management and clinical evaluations. *Decision Support Systems,* 42**,** 1152-1169.

Linstone, H. A. & Turoff, M. (1975) *The Delphi Method: Techniques and Applications* Reading, MA, Addison-Wesley.

Liu, Y.-H., Chang, F. & Lin, C.-C. (2005) Language identification of character images using machine learning techniques. *Eighth International Conference on Document Analysis and Recognition (ICDAR'05).* Seoul, Keroa.

Lobach, D. F. & Hammond, W. E. (1994) Development and evaluation of a Computer-Assisted Management Protocol (CAMP): improved compliance with care guidelines for diabetes mellitus. *Proceedings - The Annual Symposium on Computer Applications in Medical Care.*

Mack, E. H., Wheeler, D. S. & Embi, P. J. (2009) Clinical decision support systems in the pediatric intensive care unit. *Pediatric Critical Care Medicine,* 10**,** 23-28

Mackway-Jones, K., Marsden, J. & Windle, J. (2005) *Emergency Triage,* Oxford, Blackwell.

Mackway-Jones, K. R. H. (2001) Emergency management of cardiac chest pain: a review. *Emergency Medicine Journal,* 18**,** 6-10.

Mangalampalli, A., Mangalampalli, S. M., Chakravarthy, R. & Jain, A. K. (2006) A neural network based clinical decision-support system for efficient diagnosis and fuzzy-based prescription of gynecological diseases using homoeopathic medicinal system. *Expert Systems with Applications,* 30**,** 109-116.

Marakas, G. M. (2003) *Decision Support Systems in the 21st Century,* Upper Saddle River, NJ, Prentice Hall.

Marc, A. F., Chandra, P. & Steven, E. (2000) Acute gastrointestinal bleeding. *The Medical clinics of North America,* 84**,** 1183-1208.

Mcdonald, C. J. & Overhage, J. M. (1994) Guidelines you can follow and trust: an ideal and an example. *The Journal of the American Medical Association,* 271**,** 872-3.

Mei-Ling Ting, L. & Bernard, A. R. (2001) The average area under correlated receiver operating characteristic curves: a nonparametric approach based on generalized two-sample Wilcoxon statistics. *Journal of the Royal Statistical Society: Series C (Applied Statistics),* 50**,** 337-344.

Metz, C. E. (1978) Basic principles of ROC analysis. *Seminars in Nuclear Medicine,* 8**,** 283-298.

Metz, C. E., Herman, B. A. & Roe, C. A. (1998) Statistical comparison of two ROC-curve estimates from partially-paired datasets. *Medical decision making,* 18**,** 110-121.

Metzger, J. & Macdonald, K. (2002) Clinical Decision Support for the Independent Physician Practice. Oakland, California Healthcare Foundation.

Miller, P., Frawley, S., Sayward, F., Yasnoff, W., Duncan, L. & Fleming, D. (1996) IMM/Serve: a rule-based program for childhood immunization. *Proceedings : a conference of the American Medical Informatics Association / ... AMIA Annual Fall Symposium.* Hanley And Belfus.

Miller, R. A. (1996) Evaluating evaluations of medical diagnostic systems. *J Am Med Inform Assoc,* 3**,** 429-431.

Miller, R. A. & Geissbuhler, A. (1999) Clinical Diagnostic Decision Support Systems: An Overview. IN BERNER, E. S. (Ed.) *Clinical Decision Support Systems: Theory and Practice.* New York, Springer-Verlag.

Miller, R. A. & Masarie, F. E. J. (1989) Use of the Quick Medical Reference (QMR) program as a tool for medical education. *Methods of Information in Medicine* 28**,** 340-345.

Minsky, M. & Haugeland, J. (1975) A Framework for Representing Knowledge. IN WINSTON, P. H. (Ed.) *The Psychology of Computer Vision.* New York, McGraw-Hill.

Mitchell, T. M. (1997) *Machine learning,* New York, McGraw-Hill.

Musen, M., Shahar, Y. & Shortliffe, E. (2006) Clinical Decision-Support Systems. IN SHORTLIFFE, E. H. & CIMINO, J. J. (Eds.) *Biomedical Informatics.*

Musen, M. A. (1997) Modelling of Decision Support. IN BEMMEL, J. H. V. A. M., M. A. (Ed.) *Handbook of medical informatics.* Houten, Bohn Stafleu Van Loghum.

Nicholson, A. E., Twardy, C. R., Korb, K. B. & Hope, L. R. (2008) Decision Support for Clinical Cardiovascular Risk Assessment. IN POURRET, O., NAIM, P. & MARCOT, B. (Eds.) *Bayesian Networks.*

Nunamaker, J. F., Jr. & Chen, M. (1990) Systems development in information systems research. IN LEE, W. H. & VEUJKO, M. (Eds.) *The Twenty-Third Annual Hawaii International Conference on System Sciences.* Hawaii, IEEE.

Orfali, R. & Harkey, D. (1998) *Client/Server Programming with Java and CORBA,* New York, John Wiley & Sons.

Palma, J., Juarez, J. M., Campos, M. & Marin, R. (2006) Fuzzy theory approach for temporal model-based diagnosis: An application to medical domains. *Artificial Intelligence in Medicine,* 38**,** 197-218.

Park, S. H., Goo, J. M. & Jo, C.-H. (2004) Receiver Operating Characteristic (ROC) curve: practical review for radiologists. *Korean J Radiol,* 5**,** 11-18.

Peleg, M., Boxwala, A., Ogunyemi, O. & Al., E. (2000) GLIF3: The evolution of a guideline representation format. IN OVERHAGE, J. M. (Ed.) *Proceedings AMIA Annual Symposium.* AMIA.

Peleg, M., Tu, S., Bury, J., Ciccarese, P., Fox, J., Greenes, R. A., Hall, R., Johnson, P. D., Jones, N., Kumar, A., Miksch, S., Quaglini, S., Seyfang, A., Shortliffe, E. H. & Stefanelli, M. (2003) Comparing computer-interpretable guideline models: A case-study approach. *J Am Med Inform Assoc,* 10**,** 52-68.

Pinciroli, F., Combi, C. & Pozzi, G. (1992) Object-orientated DBMS techniques for time-oriented medical record. *Medical informatics* 17**,** 231-241.

Rangel, E. (2009) Clinical ethics and the dynamics of group decision-making: applying the psychological data to decisions made by ethics committees. *HEC Forum,* 21**,** 207-228.

Rao, G. R. & Suresh, B. A. (1995) Collaborative communication in medicine using EIES II (Electronic Information Exchange System II) and medicalware. *Engineering in Medicine and Biology Society, 1995 and 14th Conference of the Biomedical Engineering Society of India. An International Meeting, the First Regional Conference.* New Delhi IEEE.

Rao, G. R., Suresh, B. A. & Turoff, M. (1996) An analysis of the schools of medical decision making from a group decision support system perspective. *Proceedings of the 18th Annual International Conference of the IEEE.*

Rao, G. R., Suresh, B. A., Turoff, M. & Hiltz, S. R. (1994) Issues in the development of a computer mediated communication system framework for collaborative medical decision making. *Proceedings of the 16th Annual International Conference of the IEEE.*

Rao, G. R. & Turoff, M. (2000) A hypermedia-based group decision support system to support collaborative medical decision-making. *Decision Support Systems,* 30**,** 187-216.

Reason, J. (2001) Understanding Adverse Events: the Human Factor. IN CHARLES, V. (Ed.) *Clinical Risk Management: Enhancing Patient Safety.* Second ed. London, BMJ publishing house.

Rockall, T. A., Logan, R. F. A., Devlin, H. B. & Northfield, T. C. (1995) Incidence of and mortality from acute upper gastrointestinal haemorrhage in the United Kingdom. *BMJ,* 311**,** 222-226.

Roukema, J., Steyerberg, E. W., Van Der Lei, J. & Moll, H. A. (2008) Randomized trial of a clinical decision support system: impact on the management of children with fever without apparent source. *Journal of the American Medical Informatics Association,* 15**,** 107-113.

Sam, C., Fai, W. & Yiping, L. (2008) An Incremental and Interactive Decision Tree Learning Algorithm for a Practical Diagnostic Supporting Workbench. *Networked Computing and Advanced Information Management, 2008. NCM '08. Fourth International Conference on.*

Seufert, S. & O'brien, D. (2007) Machine learning for automatic defense against distributed denial- of-service attacks. *International Conference on Communications (ICC'07).*

Shafer, G. (1976) *A Mathematical Theory of Evidence,* Princeton, Princeton University Press.

Shiomi, S., Kuroki, T., Jomura, H., Ueda, T., Ikeoka, N., Kobayashi, K., Ikeda, H. & Ochi, H. (1995) Diagnosis of chronic liver disease from liver scintiscans by fuzzy reasoning. *J Nucl Med,* 36**,** 593-598.

Shortliffe, E. H. (1976) *Computer-Based Medical Consultations: MYCIN,* New York, Elsevier.

Shortliffe, E. H. & Perreault, L. E. (1990) *Medical Informatics,* Reading, MA, Addison Wesley Publishing.

Sim, I., Gorman, P., Greenes, R. A., Haynes, R. B., Kaplan, B., Lehmann, H. & Tang, P. C. (2001) Clinical decision support systems for the practice of evidence-based medicine. *J Am Med Inform Assoc,* 8**,** 527-534.

Sittig, D. F., Wright, A., Osheroff, J. A., Middleton, B., Teich, J. M., Ash, J. S., Campbell, E. & Bates, D. W. (2008) Grand challenges in clinical decision support. *Journal of Biomedical Informatics,* 41**,** 387-392.

Skalska, H. & Freylich, V. (2006) Web-bootstrap estimate of area under ROC curve. *Austraian journal of statistics,* 35**,** 325-330.

Skevofilakas, M. T., Nikita, K. S., Templaleksis, P. H., Birbas, K. N., Kaklamanos, I. G. & Bonatsos, G. N. (2005) A decision support system for breast cancer treatment based on data mining technologies and clinical practice guidelines. *Engineering in Medicine and Biology Society, 2005. IEEE-EMBS 2005. 27th Annual International Conference of the.*

Sommerville, I. (2007) *Software Engineering,* Essex, Person Education.

Spooner, S. A. (1999) Mathematical Foundations of Decision Support Systems. IN BERNER, E. S. (Ed.) *Clinical Decision Support Systems: Theory and Practice.* New York, Springer-Verlag.

Spooner, S. A. (2007) Mathematical Foundations of Decision Support Systems. IN BERNER, E. S. (Ed.) *Clinical Decision Support Systems: Theory and Practice.* 2nd ed. New York, Springer-Verlag.

Starren, J. B., Hripcsak, G., Jordan, D., Allen, B., Weissman, C. & Clayton, P. D. (1994) Encoding a post-operative coronary artery bypass surgery care plan in the Arden Syntax. *Computers in Biology and Medicine,* 24**,** 411-417.

Stefania Montani, P. T. (2006) Exploiting decision theory concepts within clinical guideline systems: Toward a general approach. *International Journal of Intelligent Systems,* 21**,** 585-599.

Suryanarayanan, S., Reddy, N. P. & Canilang, E. P. (1995) A fuzzy logic diagnosis system for classification of pharyngeal dysphagia. *International Journal of Bio-Medical Computing,* 38**,** 207-215.

Szolovits, P. (1995) Uncertainty and Decisions in Medical Informatics. *Methods of Information in Medicine,* 34**,** 111-121.

Tan, T. Z., Quek, C., Ng, G. S. & Razvi, K. (2008) Ovarian cancer diagnosis with complementary learning fuzzy neural network. *Artificial Intelligence in Medicine,* 43**,** 207-222.

Terenziani, P., Molino, G. & Torchio, M. (2001) A modular approach for representing and executing clinical guidelines. *Artificial Intelligence in Medicine,* 23**,** 249-276.

Terenziani, P., Montani, S., Bottrighi, A., Torchio, M., Molino, G., Anselma, L. & Correndo, G. (2003) Applying Artificial Intelligence to clinical guidelines: The GLARE approach. *Ai(Asterisk)Ia 2003: Advances in Artificial Intelligence, Proceedings.*

Thilo, B., Vasileios, A., Gregor, H., Friederike, L., Jens, K., Simon, P. W. S., Hamid, G., Angelika Zabel-Du, B., Stefanie, M.-Z., Jgen, D., Hubert, J. B. & Walter, E. H. (2009) Multidisciplinary pain management based on a computerized clinical decision support system in cancer pain patients. *Pain,* 147**,** 20-28.

Thomas, E. J. & Brennan, T. A. (2001) Errors and Adverse Events in Medicine: An Overview. IN CHARLES, V. (Ed.) *Clinical Risk Management: Enhancing Patient Safety.* Second ed. London, BMJ publishing house.

Thomas, K. W., Dayton, C. S. & Peterson, M. W. (1999) Evaluation of internet-based clinical decision support systems. *Journal of Medical Internet Research* 1**,** e6.

Tierney, W. M., Overhage, J. M., Takesue, B. Y., Harris, L. E., Murray, M. D. & Vargo, D. L. (1995) Computerizing guidelines to improve care and patient outcomes: the example of heart failure. *Journal of the American Medical Informatics Association,* 2**,** 316-22.

Tong, S. & Koller, D. (2001) Support vector machine active learning with applications to text classification. *Journal of machine learning research,* 2**,** 45-66.

Tsymbal, A., Zhou, S. K. & Huber, M. (2009) Neighborhood graph and learning discriminative distance functions for clinical decision support. *Annual International Conference of the IEEE Engineering in Medicine and Biology Society* Minnesota, IEEE

Tu, S. W., Eriksson, H., Gennari, J. H., Shahar, Y. & Musen, M. A. (1995) Ontology-based configuration of problem-solving methods and generation of knowledge-acquisition tools: application of PROTEGE-II to protocol-based decision support. *Artificial intelligence in medicine,* 7**,** 257-289.

Tu, S. W. & Musen, M. A. (1999) A Flexible Approach to Guideline Modeling. *Proceeding of American Medical Informatics Association Annual Symposium.* AMIA.

Turban, E. & Aronson, J. E. (2001) *Decision Support Systems and Intelligent Systems,* New Jersey, Prentice-Hall International.

Vergara, I., Norambuena, T., Ferrada, E., Slater, A. & Melo, F. (2008) StAR: a simple tool for the statistical comparison of ROC curves. *BMC Bioinformatics,* 9**,** 265.

Wang, Y. M., Yang, J. B. & Xu, D. L. (2006) Environmental impact assessment using the evidential reasoning approach. *European Journal of Operational Research,* 174**,** 1885-1913.

Warner, H. R. (1979) *Computer-Assisted Medical Decision Making,* New York, Academic Press.

Warner, H. R. J. (1989) Iliad: moving medical decision-making into new frontiers. *Methods of information in medicine,* 28**,** 370-372.

Waymire, R. & Sawtell, R. (2000) *Sams Teach Yourself SQL Server 2000 in 21 Days* Indianapolis, Sams Publishing.

Wright, D. B. (1997) *Understanding Statistics: An Introduction for the Social Sciences,* London, SAGE Publications.

Xu, D. L., Liu, J., Yang, J. B., Liu, G. P., Wang, J., Jenkinson, I. & Ren, J. (2007) Inference and learning methodology of belief-rule-based expert system for pipeline leak detection. *Expert Systems with Applications,* 32**,** 103-113.

Xu, D. L. & Yang, J. B. (2005) An intelligent decision system based on the evidential reasoning approach and its applications. *Journal of Telecommunications and Information Technology,* 3**,** 73-80.

Yan, H., Jiang, Y., Zheng, J., Peng, C. & Li, Q. (2006) A multilayer perceptron-based medical decision support system for heart disease diagnosis. *Expert Systems with Applications,* 30**,** 272-281.

Yang, J. B. (2001) Rule and utility based evidential reasoning approach for multiattribute decision analysis under uncertainties. *European Journal of Operational Research,* 131**,** 31-61.

Yang, J. B., Liu, J., Wang, J., Sii, H. S. & Wang, H. W. (2006) Belief rule-base inference methodology using the evidential reasoning approach - RIMER. *IEEE Transactions on Systems Man and Cybernetics Part A-Systems and Humans,* 36**,** 266-285.

Yang, J. B., Liu, J., Xu, D. L., Wang, J. & Wang, H. (2007) Optimization models for training belief-rule-based systems. *IEEE Transactions on Systems, Man and Cybernetics, Part A-Systems and Humans,* 37**,** 569-585.

Yang, J. B. & Sen, P. (1994) A general multi-level evaluation process for hybrid MADM with uncertainty. *Systems, Man and Cybernetics, IEEE Transactions on,* 24**,** 1458-1473.

Yang, J. B. & Singh, M. G. (1994) An evidential reasoning approach for multiple-attribute decision making with uncertainty. *Systems, Man and Cybernetics, IEEE Transactions on,* 24**,** 1-18.

Yang, J. B. & Xu, D. L. (2002) On the evidential reasoning algorithm for multiple attribute decision analysis under uncertainty. *IEEE Transactions on Systems Man and Cybernetics Part A-Systems and Humans,* 32**,** 289-304.

Zadeh, L. A. (1973) Outline of a new approach to the analysis of complex systems and decision processes. *Systems, Man and Cybernetics, IEEE Transactions on,* SMC-3**,** 28-44.

Zhou, Z. J., Hu, C. H., Yang, J. B., Xu, D. L., Chen, M. Y. & Zhou, D. H. (2010) A sequential learning algorithm for online constructing belief-rule-based systems. *Expert Systems with Applications,* 37**,** 1790-1799.

Zhou, Z. J., Hu, C. H., Yang, J. B., Xu, D. L. & Zhou, D. H. (2009) Online updating belief rule based system for pipeline leak detection under expert intervention. *Expert Systems with Applications,* 36**,** 7700-7709.

Zou, K. H., Hall, W. J. & Shapiro, D. E. (1997) Smooth non-parametric receiver operating characteristic (ROC) curves for continuous diagnostic tests. *Statistics in medicine,* 16**,** 2143-2156.