

# 헬스 케어를 위한 RDMS 설계

이병관\*, 정은희<sup>o</sup>

## Design of Rough Set Theory Based Disease Monitoring System for Healthcare

Byung-Kwan Lee\*, Eun-Hee Jeong<sup>o</sup>

요 약

본 논문에서는 헬스 케어 시스템에서 효율적으로 질병을 관리할 수 있는 RDMS(Rough Set Theory based Disease Monitoring System)을 제안한다. RDMS는 DCM(Data Collection Module), RDRGM(RST based Disease Rule Generation Module), HMM(Healthcare Monitoring Module)로 구성된다. DCM은 바이오센서로부터 환자의 생체 정보를 수집하고, 데이터 처리 절차에 따라 RDMS DB에 저장한다. RDRGM은 RST의 코어와 속성의 지지율을 이용하여 질병 규칙을 생성한다. HMM은 DCM에 의해 수집된 환자 정보를 이용하여 환자의 질병에 대한 위험지수뿐만 아니라 질병에 대한 합병증에 관한 위험지수까지 분석함으로써 환자의 질병을 예측하고, 환자의 위험지수에 따라 환자, 주치의 등에 시각화된 환자의 정보를 전달한다. 또한, RDRGM에 의해 생성된 규칙들에 따라 환자의 의료정보, 현재의 환자건강상태, 환자 가족력 등을 비교분석하여 환자의 질병을 예측하고, 예측결과에 따라 환자 맞춤형 의료 서비스와 의료 정보를 신속하고 신뢰성 있게 제공할 수 있다.

**Key Words** : Risk quotient, Rough Set Theory, Disease Rule, Healthcare Monitoring, Healthcare System

### ABSTRACT

This paper proposes the RDMS(Rough Set Theory based Disease Monitoring System) which efficiently manages diseases in Healthcare System. The RDMS is made up of DCM(Data Collection Module), RDRGM(RST based Disease Rules Generation Module), and HMM(Healthcare Monitoring Module). The DCM collects bio-metric informations from bio sensor of patient and stores it in RDMS DB according to the processing procedure of data. The RDRGM generates disease rules using the core of RST and the support of attributes. The HMM predicts a patient's disease by analyzing not only the risk quotient but also that of complications on the patient's disease by using the collected patient's information by DCM and transfers a visualized patient's information to a patient, a family doctor, etc according to a patient's risk quotient. Also the HMM predicts the patient's disease by comparing and analyzing a patient's medical information, a current patient's health condition, and a patient's family history according to the rules generated by RDRGM and can provide the Patient-Customized Medical Service and the medical information with the prediction result rapidly and reliably.

※ 이 논문은 2013년도 정부(교육과학기술부)의 재원으로 한국연구재단의 기초연구사업 지원을 받아 수행된 것임 (NRF-2012R1A1A4A01012039)

♦ First Author : 관동대학교 컴퓨터학과, bkleee@kd.ac.kr, 정희원

<sup>o</sup> Corresponding Author : 강원대학교 지역경제학과, jeongeh@kangwon.ac.kr, 정희원

논문번호 : KICS2013-10-465, 접수일자 : 2013년 10월 30일, 심사일자 : 2013년 11월 25일, 최종논문접수일자 : 2013년 11월 29일

## 1. 서 론

정보의 디지털화, 통신의 광대역화, 유무선 통신망을 통한 대용량의 정보를 빠르게 전송할 수 있게 되었으며, 멀티미디어 처리 및 저장 기술의 발전, RFID를 비롯한 각종 유비쿼터스 환경의 등장으로 유비쿼터스 헬스케어의 현실화가 빠르게 진행되고 있다. 또한 고령화 사회의 도래에 따른 환경, 생활, 건강 등 다양한 요소들의 유기적인 융합을 통해 USN 기반의 개인 맞춤형 운동 관리 및 예방의 사회적 요구에 따른 u-헬스(u-Health) 관련 신규서비스에 대한 수요가 증가하고 있다<sup>1)</sup>. 그리고 현재, 의료소비자는 언제, 어디서나 컴퓨터와 연결이 가능한 휴대형 진단기기를 이용하여 수시로 건강상태를 확인함으로써 조기 진단이 가능해졌다. 하지만, 이러한 헬스 케어 장비를 활용한 헬스 케어 시스템과 병원에서는 환자에 관한 수많은 양의 데이터를 매일 수집하고 있으므로, 이러한 빅 데이터 속에서 환자 질병에 대한 규칙을 추출하기 위한 기법이 필요하다.

본 논문에서는 Rough Set Theory(RST)를 이용하여 개인맞춤형 의료서비스와 정보를 제공할 수 있는 RST 기반 질병 모니터링 시스템인 RDMS(RST based Disease Monitoring System)를 제안한다. RDMS에서는 기존의 C4.5 알고리즘이 갖는 속성 간 관계를 무시하는 문제점을 보완하기 위해 RST의 core와 support를 이용하여 질병규칙을 생성하고, 질병에 대한 위험지수를 분석하고, 위험 지수에 따라 환자와 주치의에게 환자상태에 대한 메시지를 전달하고, 질병 규칙에 따라 환자의 정보를 접목시켜 개인맞춤형 의료서비스를 모니터링할 수 있게 한다. 그 결과, RDMS는 환자의 현재 상태에 대한 분석 결과를 환자에게 피드백 하거나, 환자의 상태에 따라 현재 환자가 조치할 수 있는 응급 조치방법을 알려주어 불의의 사고를 미연에 방지하고자 한다.

본 논문의 구성은 다음과 같다. 2장에서는 RST와 헬스 케어 모니터링에 대한 관련연구를 설명하고, 3장에서는 본 논문에서 제안하는 RDMS를 설계한다. 그리고 4장에서는 RDMS의 성능을 분석하고 5장에서 결론을 맺는다.

## II. 관련연구

### 2.1. Rough Set Theory

Rough Set Theory(RST)는 부정확하거나 불완전하고, 애매모호한 성질을 갖는 데이터를 하한 근사 공간(Lower Approximation)와 상한 근사 공간(Upper Approximation)으로 분류하는 알고리즘으로 1980년대 초에 Pawlak에 의해 소개 되었다<sup>2,3)</sup>.

RST는 동치관계에 의해 데이터를 동치류(equivalence class)로 구분한 후, 이들 동치류 원소의 집합을 기본 집합이라 한다. 그리고 이 기본 집합에 의해 정의되는 집합 공간을 근사(approximation) 공간이라 한다. RST는 근사 공간 상에 하나의 결정에 대한 데이터를 분류할 때, 동일한 기본 집합 내에 있으면서도 서로 다른 결정을 내는 경우가 발생하는데 이것을 결정상의 불일치(inconsistency)라고 한다. RST는 이러한 불일치를 처리하기 위해서 그림 1과 같이 상한근사(Upper approximation), 하한근사(Lower approximation)를 정의한다<sup>4,5)</sup>.

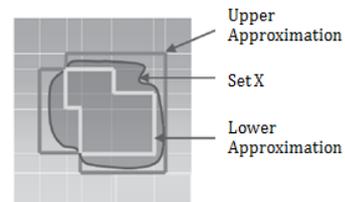


그림 1. Rough Set Theory  
Fig. 1. Rough Set Theory

그림 1에서 하한 근사는 X에 항상 포함되는 기본 집합이고, 상한근사는 X와 일치하는 부분이 하나라도 존재하는 모든 기본집합을 의미한다. 즉, U를 전체집합이라 하고 R을 U에 대한 동치관계라 하고, 근사 공간  $A=(U, R)$ 라고 할 때, 하한 근사와 상한근사는 다음과 같이 표현한다.

$$\text{하한근사집합} : \underline{R}X = \{x \in U | [x] \subseteq X\}$$

$$\text{상한근사집합} : \overline{R}X = \{x \in U | [x] \cap X \neq \emptyset\}$$

여기서  $[x]$ 는 원소  $x$ 를 포함하는 R의 동치류를 나타내므로, 집합 X의 러프 셋은  $AR(X) = (\underline{R}X, \overline{R}X)$ 로 정의된다. 또한 상한 근사에서 하한 근사를 제외시키면 불확실한 개체들만 또 다른 부분집합으로 표현될 수 있는데, 이를 경계 영역이라 부르며,  $BNR(X) = \overline{R}X - \underline{R}X$ 로 표현한다<sup>6,7)</sup>.

RST는 전체 집합 U에서 속성 집합 C의 하나의 원소 a가  $IND(C) = IND(C - \{a\})$ 를 만족할 경우, 속성 a는 C에서 불필요(dispensable)하고, 그렇지 않으면 a는 C에서 필요(indispensable)하다고 정의한다. 그

리고 속성 집합 C에서 임의의 속성을 추출할 때 불필요한 속성집합을 뺀 최소속성집합을 C'라고 할 때,  $C' \subset A$ 이고  $IND(C) = IND(C')$ 일 경우, C'를 C의 리덕트(reduct)라고 하고, C'에서 필요한 모든 속성들의 모임, 즉 reduct들의 교집합을 C의 코어,  $Core(C) = \cap Reduct(C)$ 라고 한다<sup>18)</sup>.

RST을 이용한 연구는 현실세계에 존재하는 불확실한 데이터를 다루는 데 있어 매우 유용하다는 평가를 받고 있기 때문에 적용 분야가 대단히 넓다. 예를 들어 인공지능, 인지과학, 의료 데이터 분석, 패턴 인식 등과 같은 응용 분야가 있으며, 데이터 마이닝 분야에도 널리 활용되고 있다<sup>16)</sup>.

본 논문에서는 데이터의 각 속성에 대한 지지율을 계산하여 꼭 필요한 속성만을 이용하여 RST의 코어를 계산하고, 이 코어의 각 속성에 대한 지지율(support)을 계산하고, 지지율이 높은 속성을 루트노드로 설정하도록 설계함으로써 각 속성의 관련성을 분석하고, 그 결과를 RDMS에 반영하도록 설계하였다.

### 2.2. 헬스 케어 모니터링

유비쿼터스 헬스케어를 실현하기 위해서는 일상생활 중 안정적인 생체정보 모니터링이 필수적이다. 일상생활 중 인체로부터 건강상태를 모니터링할 수 있는 생체정보로는 체온, 심전도, 호흡, 폐음, 심음, 근전도, 피부저항, 목소리, 가속도, 뇌파, 혈당, 체중, 혈압 등 매우 다양한 것들이 있다<sup>19)</sup>.

한국전자통신연구원(ETRI)에서는 의복형 생체 신호 모니터링 시스템을 개발하였다. 의복형 생체신호 모니터링 시스템은 바이오셔츠라고 이름을 붙인 생체정보를 측정할 수 있는 천소재 센서가 내장된 의복과 바이오셔츠로부터 측정된 생체정보를 처리하여 무선 전송할 수 있는 생체신호 처리 모듈로 구성된다.

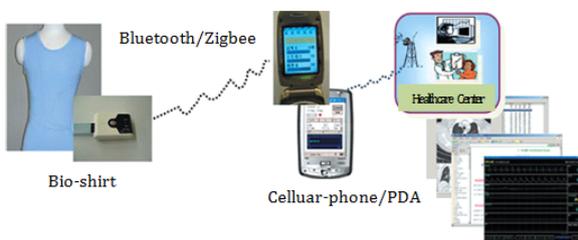


그림 2. 바이오셔츠를 이용한 유비쿼터스 헬스 케어 시스템  
Fig. 2. Ubiquitous Healthcare System using Bio-shirts

생체신호 처리 모듈로부터 전송된 생체정보는

PDA나 휴대전화 등 휴대단말을 통해 확인할 수 있으며, 휴대단말에 내장된 이동통신 또는 무선인터넷 전송 모듈로 서비스센터로 전송이 가능하고, 서비스센터에서는 응급상황 감시, 만성질환 관리, 운동 및 비만관리 등 각종 건강 관련 서비스를 제공할 수 있다<sup>9,10)</sup>.

바이오패치는 바이오 셔츠와 달리 몸에 직접 부착하여 생체신호를 측정하는 모듈로 일회용 심전도 전위 센서가 모듈에 착탈식으로 붙일 수 있도록 구성되어 있다.

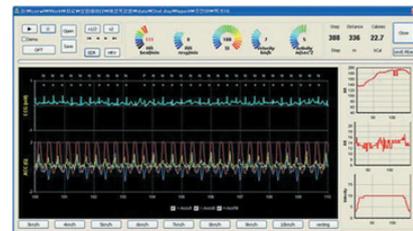


그림 3. 바이오패치와 생체신호 분석 프로그램  
Fig. 3. Bio-patch and bio-metric analysis program

바이오 패치는 심전도 외에 3축 가속도 신호를 측정 할 수 있으며 이를 통해 속도 운동량 활동량을 측정한다. 또한 바이오 패치에는 블루투스나 지그비 같은 근거리 무선 통신소자가 내장되어 측정된 생체 정보를 무선으로 전송할 수 있다. 바이오패치에서 측정된 심전도로부터 바이오 셔츠와 마찬가지로 심박수, 스트레스지수, 호흡수, HRV(Heart Rate Variability) 등 건강관련지표를 추출할 수 있다. 바이오패치의 크기와 무게 또한 바이오 셔츠의 생체신호처리모듈과 비슷하여 착용에 불편함이 없이 사용할 수 있다<sup>19)</sup>.

### III. RST기반 질병 관리 모니터링 시스템 설계

본 논문에서는 환자의 가족력과 환자 본인의 병력에 대한 정보를 수집한 후, 환자의 질병을 판단하거나 예측하는 RDMS를 설계하여, 환자의 생체정보를 모니터링하여 환자 맞춤형 의료서비스를 제공하고자 한다.

그림 4는 RDMS의 전반적인 흐름을 설명한 것이다.

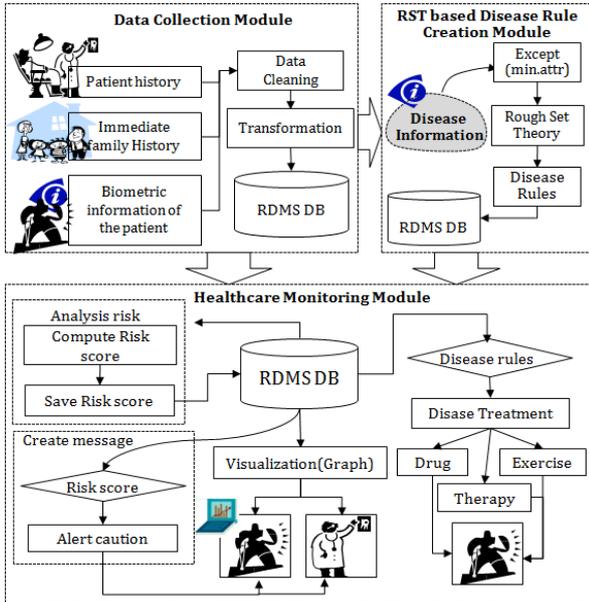


그림 4. RDMS의 전체 흐름도  
Fig. 4. The flowchart of RDMS

3.1. Data Collection Module 설계

RDMS의 DCM(Data Collection Module)은 바이오센서로부터 심전계, 뇌파계, 근전계, 혈당계, 혈압계에 대한 데이터와 환자의 질병에 대한 정보를 수집한다. 또한, 환자의 병력, 환자 가족력, 현재 환자의 건강상태를 수집하여 RDMS DB에 저장하는데, 아래의 조건들을 확인하고 수정한 후에 저장한다.

조건 1. DCM은 수집된 데이터가 정확한지를 확인한다. 즉, DCM은 데이터의 입력오류를 확인하고, 입력된 환자의 생체정보가 허용되는 범위이내 데이터인지도 확인한다.

조건 2. DCM은 환자가 특별한 실험에 참여하였을 경우에 발생한 데이터는 단지 실험에 대한 데이터로 처리하고, 이 데이터를 환자의 질병 예측하는 데이터로 활용하지 않는다.

조건 3. DCM은 누락된 데이터(missing data)를 처리해야 하는 경우에는 환자의 최신 정보 또는 데이터 집합(dataset)의 평균값을 사용한다.

조건 4. 모든 생체정보에 대한 실질적으로 중요한 매개변수는 적용범위이내의 있는 측정값으로 조정한다.

RDMS는 환자정보와 센서 데이터 형식은 그림 5

와 같은 형식으로 RDMS DB에 저장한다. 여기서 환자의 고유 ID는 환자 정보와 센서 데이터의 P\_Identifier 필드에 기록함으로써 각각의 환자 정보를 구분한다. 그리고 current 필드, Disease\_History 필드, Family\_History 필드에는 환자의 현재 상태, 환자의 질병 이력, 환자의 가족병력을 각각 기록한다. 센서 데이터의 Disease\_No 필드는 측정된 센서 데이터를 통해 환자에게 가장 위험한 질병이면서 환자가 가장 유의해야할 질병에 대한 식별 번호를 저장하고, Sensor\_Data 필드에는 각 센서의 ID와 측정 데이터를 기록한다. 그리고 Risk\_quotient 필드는 환자의 위험 분석 후 측정된 위험 지수를 저장한다. 그림 5에서 설명하고 있듯이 환자 정보와 센서 데이터는 P\_Identifier로 1대다 관계를 가진다.

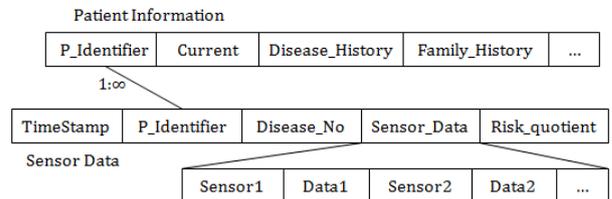


그림 5. 환자 정보와 센서 데이터의 형식  
Fig. 5. Format of patient information and sensor data

표 1은 심장질환, 당뇨병, 고혈압, 저혈압 등의 질병을 측정할 때, 가장 유의해야할 수집 데이터의 종류를 설명한 것이다.

표 1. 각 질병에 대한 센서 데이터의 종류  
Table 1. The kind of sensor data on each disease

Disease No.	Bio Sensor Data /Disease Name	ECG	EEG	Blood sugar	BP
000	Myocardial infarction	O			
010	Epilepsy		O		
011	Diabetes			O	
100	Heart(angiographic)	O		O	O
101	Hypoglycemia			O	
110	Hypertension				O
111	Hypotension				O

\*ECG: electrocardiogram, EEG:electroencephalogram, BP: Blood Pressure

3.2. RST based Disease Rule Generation Module 설계

본 논문에서 제안하는 RDRGM(RST based Disease Rule Generation Module)은 RST를 이용하여 각 속성에 대한 지지율을 계산하여 불필요한 속

성을 제거한 후 코어(Core)를 구한다. 그리고 코어의 속성들에 대한 지지율을 계산하여 지지율이 높은 속성을 기준으로 의사결정 규칙인 질병 규칙을 생성하도록 설계한다. 즉, RDRGM은 기존의 C4.5 알고리즘이 개체들 속성간의 관계성을 고려하지 않는 문제점을 보완하였을 뿐만 아니라, 필요한 속성만을 이용하여 좀 더 효율적으로 질병규칙을 생성할 수 있다.

RDRGM은 다음과 같은 절차에 따라 질병규칙들을 생성한다.

[1 단계] RDRGM은 표 2의 결정(decision) 속성인 Cough의 결과 값에 대한 속성 {Temperature, Headache, Nausea} 지지율을 계산한다. 특히 Nausea 속성은 다른 {Temperature, Headache}에 비해 결정 속성이 "yes"인 경우의 Support(yes) = 4/8, Support(no)=4/8로 불분명하므로 조건 속성에서 제외시킨다.

[2 단계] RDRGM은 표 2에서 속성 {Nausea}를 제외한 정보 테이블을 이용하여 식별이 불가능한 집합인 IND를 찾는다.

$$IND(\{temperature, Headache, Cough\}) = \{\{1,4,7,14\},\{2,8\},\{3,9,15\},\{5\},\{10,11,13\},\{6\},\{12\}\}$$

표 2. 정보 테이블  
Table 2. Information Table

Case	Conditional Attributes			Decision Attribute
	Temperature	Headache	Nausea	Cough
1	high	yes	yes	yes
2	very high	yes	yes	yes
3	high	no	no	no
4	high	yes	yes	yes
5	normal	yes	no	yes
6	normal	no	yes	no
7	high	yes	no	yes
8	very high	yes	no	yes
9	high	no	yes	no
10	normal	yes	no	no
11	normal	yes	yes	no
12	high	no	yes	yes
13	normal	yes	yes	no
14	high	yes	no	yes
15	high	no	no	no

[3 단계] RDRGM은 하한 근사 집합과 상한 근사 집합을 찾아 boundary 영역 집합을 구한다. 이때, 조건 속성값들은 모두 일치하지만, 결정 속성값이 다른 {{3,9,15},{12}}와 {{5},{10,11,13}}를 찾

아낸다. 이 두 값은 하한 근사값에서 포함되지 않고 상한 근사값에 포함된다. 따라서 Boundary Region(BR)은 다음과 같다.

$$\text{Cough가 Yes인 경우: 상한근사집합-하한근사집합} = \{1,2,4,5,7,8,12,14\} - \{1,2,4,7,8,14\} = \{5,12\}$$

$$\text{Cough가 No인 경우: 상한근사집합-하한근사집합} = \{3,6,9,10,11,13,15\} - \{6\} = \{3,9,10,11,13,15\}$$

[4 단계] RDRGM은 2 단계와 3 단계의 결과값을 이용하여 각 속성에 대한 reduct를 구한다. 이때, 조건 속성값이 같고 결정 속성값이 다른 {{3,9,15}, {12}}와 {{5},{10,11,13}}은 제외시키고, 3단계의 결과 값인 IND(C)에서 속성 값이 같은 집합에서 한 개만 추출하여 reduct를 구한다. 표 3은 RDRGM의 reduct한 결과 값을 설명한 것이다.

표 3. 속성 {Temperature, Headache}에 대한 reduct 결과  
Table 3. The reduct result of attributes {Temperature, Headache}

Case	Conditional Attributes		Decision Attribute
	Temperature	Headache	Cough
1	high	yes	yes
2	very high	yes	yes
6	normal	no	no

[5 단계] RDRGM은 reduct의 교집합을 찾아 core를 구한다. 만약 속성이 3개 이상일 경우에는 2개의 속성에 대한 reduct 결과가 3개가 생성되므로, 3개의 reduct 결과 값에 대한 교집합이 core가 된다. 본 논문에서 제시한 예제는 3개의 속성 중에서 decision 속성에 대한 지지율이 불분명한 속성을 제외한 2개의 속성에 대해서만 reduct 하였다. 그 결과 reduct가 1개이므로, 이 reduct가 곧 core가 된다. 이때 코어의 속성이 1개인 경우, 코어의 속성이 복수 개인 경우, 코어의 속성이 없는 경우가 발생할 수 있다.

RDRGM은 4단계 reduct의 결과 값인 표 3의 {1,2,6}이 코어가 되고, 코어의 속성의 개수는 2개임을 알 수 있다.

[6 단계] RDRGM은 코어의 속성 개수에 따라 root 노드를 선택한다. 즉, 코어의 속성이 1개인 경우, 코어를 루트로 선택하고, 코어의 속성이 복수 개인 경우, 코어에 해당되는 속성들의 지지율(Support)을 계산한다. 이때, 지지율이 가장 높은 속성을 루트 노드로 선택한다. 그리고 코어의 속성이 없는 경우, 전체 속성에 대해 지지율이 가장 높은

속성을 루트 노드로 선택한다.

RDRGM은 코어의 속성이 여러 개이므로 표 4처럼 코어 속성에 대한 지지율을 계산하고, 속성에 대한 지지율이 가장 높은 속성을 Decision Tree의 루트 노드로 선정한다. 즉, 속성 {Headache}가 루트 노드가 된다.

표 4. 코어와 코어 속성의 지지율  
Table 4. Core and support of Core attributes

Case	Conditional Attributes		Decision Attribute
	Temperature	Headache	Cough
1	high	yes	yes
2	very high	yes	yes
6	normal	no	no
support	high-1/3, very high-1/3 normal-1/3	yes-2/3, no-1/3	

[7 단계] RDRGM은 루트 노드가 선택되면, 선택된 속성의 값으로 분지하고, 각 속성 값에 대해 2 단계부터 다시 반복하여 자식노드를 선택한다. 그리고 더 이상 분지할 수 없고, 결정속성이 결정되면 분지를 종료한다. 즉, 표 2에서 속성 {Headache, Nowsea}를 제외한 나머지 속성에 대한 reduct를 생성한다. 본 논문에서 제시한 예제는 지지율이 낮은 속성 {Nowsea}를 제외한 두 개의 속성만을 이용하므로 자동적으로 Temperature가 Headache의 자식 노드가 된다.

본 논문에서 제안한 RDRGM에서 생성한 질병 결정 규칙(RDRGM rules)과 C4.5 알고리즘<sup>[11]</sup>으로 생성된 질병 결정 규칙(C4.5 rules)을 정리해보면, 표 5와 같다.

표 5. C4.5 알고리즘과 RDRGM의 질병 결정 규칙들  
Table 5. The disease decision rules of C4.5 algorithm and RDRGM

C4.5 Rules	Rule 1	if (Temperature=high) and (Headache=yes) then (Cough=yes)(100%)
	Rule 2	if (Temperature=high) and (Headache=no) then (Cough=no) (75%)
	Rule 3	if (Temperature=very high) then (Cough=yes)(100%)
	Rule 4	if (Temperature=normal) then (Cough=no) (80%)

RDRGM Rules	Rule 1	if (Headache=yes) and (Temperature=very high) then (Cough=yes)(100%)
	Rule 2	if (Headache=yes) and (Temperature=high) then (Cough=yes)(100%)
	Rule 3	if (Headache=no) and (Temperature=normal) then (Cough=no)(100%)

표 5의 규칙들을 살펴보면, RDRGM에 의해 생성된 규칙 3개 모두 100%인 반면에 C4.5 알고리즘에 의해 생성된 규칙 중에 규칙 2는 75%, 규칙 4는 80%임을 알 수 있다. 따라서 RDRGM에 의해 생성된 규칙이 더 정확하다고 할 수 있겠다.

### 3.3. Healthcare Monitoring Module 설계

RDMS의 HMM(Healthcare Monitoring Module)은 DCM이 수집한 데이터를 실시간으로 정상 기준치와 비교하여 환자의 상태를 분석하고, 환자의 상태가 위험하다고 판단되면 환자에게 알리는 메시지를 생성한다. 그리고 메시지를 환자, 주치의 등에게 전달한다. 또한, RDMS의 데이터베이스에 기록된 데이터를 시각화하여 환자, 주치의, 의료기관 등에 제공하며, 환자 ID(P-Identifier)로 병원 내 기록된 환자의 위험 요소도 검색하여 수집된 환자의 데이터와 함께 출력하도록 설계한다.

그림 6은 HMM의 역할에 대한 전체적인 흐름을 설명한 것이다.

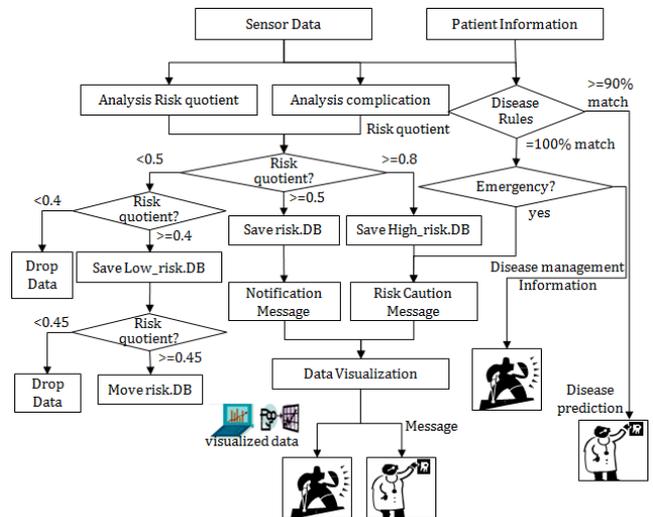


그림 6. HMM의 흐름도  
Fig. 6. The flowchart of HMM

#### 3.3.1. 위험 지수 분석

RDMS의 헬스 케어 모니터링 모듈은 질병을 판단하는 임계치가 저장되어 있는 질병임계치데이터를 이용하여 환자의 질병을 판단하고, 질병의 위험 지

수를 산출하는 비율이 저장되어 있는 질병위험지수 테이블을 이용하여 질병에 대한 위험지수를 분석한다. 따라서 본 논문에서는 표 6(a)와 표 6(b)처럼 질병임계치테이블과 질병위험지수테이블을 사전에 정의한다.

예를 들어 당뇨병인 경우, 정상인의 정상시 혈당치는 70mg/dL~100mg/dL 수준이며, 식사 등으로 당분을 섭취하여도 120mg/dL 이상으로는 올라가지 않는 것으로 알려져 있다. 즉, 이 범위를 초과하는 경우 내당능장애 및 당뇨 진단을 받게 된다. 공복혈당치가 126mg/dL이상, 식후 2시간 혈당치는 200mg/dL이상, 임의측정시에 혈당치가 2번 이상 200mg/dL이상일 때 고혈압 진단을 한다<sup>12,13,14,15</sup>.

본 논문에서는 위험지수분석단계에서 표 6(a)의 질병임계치테이블에 저장되어 있는 공복혈당 평가기준값 126mg/dL이상, 혈당계의 정상 수치인 100-140mg/dL, 고혈당 기준값 200mg/dL이상, 저혈당 기준값 60mg/dL이하를 기준으로 환자의 데이터와 비교하여 환자의 당뇨병을 판단한다. 그리고 당뇨병에 대한 합병증 분석은 표 6(b)에서 설명하고 있는 혈당, 혈압, 심혈관질환, 신장질환, 우울증에 대한 위험지수를 이용하여 당뇨병 합병증에 대한 위험지수를 분석한다.

표 6. 당뇨병자의 예(Disease No=011)

Table 6. The example of diabetes

(a) The critical value of diabetes in the disease critical table

Disease	Blood sugar	Fasting blood sugar	High blood sugar(>=)	Hypoglycemia (<)
Diabetes	100-140	126	200	60

(b) The risk quotient of diabetes in the disease risk quotient table.

Sensor Data/Disease	Blood sugar	BP	Cardiovascular	Kidneys	ST depression
Diabetes	0.5	0.2	0.2	0.05	0.05

본 논문에서는 혈당의 위험지수는 0.5, 당뇨병자의 합병증으로 가장 많이 발견되는 고혈압과 심혈관질환의 위험지수는 각각 0.2, 그리고 신장질환과 우울증은 각각 0.05로 사전에 설정하였다. 즉, 위험지수분석단계에서는 표 6(a)로 당뇨병을 진단한 후, 표 6(b)로 당뇨병 환자에 대한 위험지수를 계산하여, 그림 6의 위험지수에 따라 위험지수가 0.4미만인 경우에는 데이터를 삭제하고, 위험지수가 0.5 이상인 경우 환자의 상태에 변화가 생긴 것으로 판단하여 알림 메시지를 생성하여 환자의 데이터를 응

급센터 및 주치의에게 1회 전송한다. 그리고 위험지수가 0.8이상인 경우 매우 응급한 상황이라고 판단하여 환자의 상태 데이터를 포함한 응급 메시지를 실시간으로 전송하도록 설계한다.

### 3.3.2. 메시지 생성

HMM은 위험지수 분석 결과에 대한 알림 메시지 혹은 응급 메시지를 그림 7의 메시지 구조로 생성하여 주치의 또는 응급센터에 전달한다. 이때 알림 메시지는 m\_Identifier 필드에 "00"을 설정하고, 응급 메시지는 m\_Identifier 필드에 "01"을 설정한다.

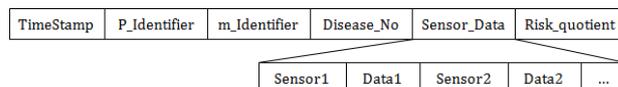


그림 7. 메시지 형식  
Fig. 7. The message

#### (1) 알림 메시지

헬스 케어 모니터링 모듈은 바이오센서를 통해 수집된 데이터를 분석한 결과, 위험지수가 0.5 이상인 경우 응급 상황이 발생할 가능성이 있다고 판단하여 환자 상태가 위험하다는 것을 알리는 알림 메시지를 그림 7의 형식으로 생성한다. 그리고 환자, 주치의에 알림 메시지를 1회만 전송된다.

#### (2) 응급 메시지

헬스 케어 모니터링 모듈은 위험 지수 분석에서 측정된 위험지수가 0.8 이상인 응급한 환자의 경우 응급 대처를 위하여 그림 7의 형식으로 응급 메시지를 생성한다. 그리고 단 1회만 전송하는 알림 메시지와는 달리 환자의 상태를 지속적으로 주치의 또는 응급센터에 전달한다.

### 3.3.3. 데이터 관리

HMM은 수집된 데이터의 위험 지수 분석 결과에 따라 데이터를 지속적으로 저장할 것인지 또는 데이터를 삭제할 것인지를 판단하여 RDMS DB 공간을 절약하도록 설계한다.

HMM의 데이터 관리 규칙은 다음과 같다.

[규칙 1] HMM은 위험 지수가 0.5이상일 경우부터 알림 메시지를 생성하므로, 위험 지수가 0.5 이상인 데이터는 위험군(Risk) 테이블 저장한다.

[규칙 1-1] 이때 HMM은 위험 지수가 0.8 이상인 고위험군의 환자들에 대한 정보는 고위험군

(High\_risk) 테이블에 저장한다.

[규칙 2] HMM은 위험 지수가 0.5 미만이고 0.4 이상인 데이터는 RDMS DB의 저위험군(Low\_risk) 테이블에 저장한다.

[규칙 2-1] 임시 테이블에 저장되어 있던 환자의 위험 지수가 최근에 0.05이상 변화한다면 임시 테이블의 데이터를 위험군 테이블로 이동시킨다.

[규칙 2-2] 임시 테이블에 저장되어 있던 환자의 위험 지수의 변화가 없으면, 환자의 상태가 안전한 것으로 판단하여 저위험군 테이블에서 데이터를 삭제한다.

### 3.3.4. 질병 예측 설계

HMM은 수집한 환자의 정보와 RDRGM에서 생성한 질병 규칙을 접목시켜 환자의 질병을 판단하거나 예측하도록 다음과 같이 설계한다.

[1 단계] 질병 예측은 입력받은 환자의 정보와 RDRCM에서 생성한 결정 규칙을 비교한다.

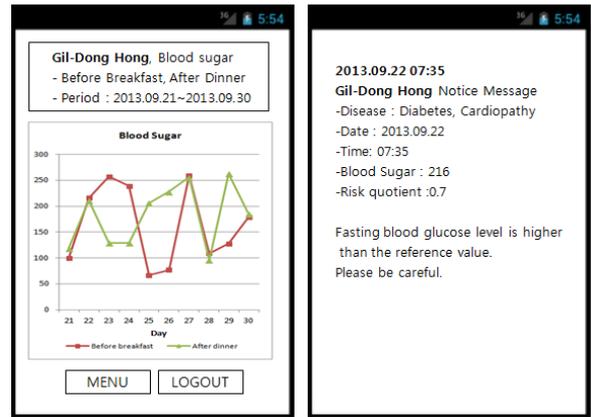
[2 단계] 질병 예측은 결정 규칙의 조건에 일치하면 결정규칙의 결과값을 환자에게 전달한다.

[3 단계] 이때, 질병 예측 모듈은 환자의 상태가 위급하다고 판단되면, 환자의 주치의 또는 응급센터에 환자의 정보를 전달한다.

[4 단계] 질병 예측은 환자에게 응급상황을 대처할 수 방법을 전달하거나, 질병을 예방하거나 완화시킬 수 있는 Therapy 등을 전달한다.

### 3.3.5 시각화 설계

HMM은 수집한 환자의 정보에 대한 분석결과를 환자 또는 주치의에게 시각화하는 정보를 그림 8과 같이 전달하도록 설계한다. 즉, 센서 노드가 부착된 환자의 생체 정보는 RDMS DB에 저장되고, RDMS 서버에는 모니터링을 위한 API를 구성하여 그래프를 작성하고, RDMS 서버에 접속하는 주치의 PC 또는 모바일 장치는 에 HTML5 언어로 작성된 웹 페이지로 생체정보 그래프를 그림 8(a)처럼 제공하도록 설계한다. 또한, 헬스 케어 모니터링 모듈이 분석한 환자의 정보에 따라 그림 8(b),(c)와 같이 알림메시지, 응급 메시지를 전달하도록 설계한다.



(a) Blood sugar

(b) Notice Message



(c) Emergency message

그림 8 시각화 설계

Fig. 8. Visualization design

## IV. 분석

### 4.1. 위험지수분석

본 논문에서 제안한 HMM의 위험지수분석에 대한 실험데이터는 Heart Disease<sup>[16]</sup>를 이용하였다. Heart Disease는 결정속성을 포함하여 14개의 속성으로 구성되어 있으며, 속성에 대한 설명은 표 7에서 설명하고 있다.

이 실험 데이터는 총 270개의 데이터로 구성되어 있으며, 본 논문에서는 이 실험 데이터의 공복혈당 수치(a6)를 이용하여 당뇨병을 판단하고, 당뇨병일 경우 혈압(a4), 우울증(a10), 심장질병(a14)의 속성을 이용하여 위험지수를 산출하도록 하였다.

표 7. 환자의 위험지수 실험 데이터 속성들과 값

Table 7. The attributes and values of patient's risk quotient experiment data

Field Name	Description	Data Ranges
a1	age	
a2	sex	1=male, 0=female

a3	chest pain type	1=typical angina, 2=atypical angina 3=non-angina pain, 4=asymptomatic
a4	blood pressure	
a5	serum cholestorol in mg/dL	
a6	fasting blood sugar	>=120
a7	ECG results	0=normal, 1=abnormality 2=left ventricular hypertrophy
a8	maximum heart rate	
a9	exercise induced angina	0=no, 1=yes
a10	ST depression	
a11	the slope of ST segment	1=upsloping, 2=flat 3=downsloping
a12	number of major vessels	
a13	thal	3=normal, 6=fixed defect, 7=reversible defect
a14	Class	1=absence 2=presence

그 결과, HMM의 위험지수분석 결과는 다음과 같다.

첫째, 270명의 환자 중에서 공복혈당수치가 120 이상인 40명을 당뇨병환자로 판단하였다.

둘째, 40명의 당뇨병 환자 중에서 고혈압 환자는 15명, 심장질환환자는 17명, 우울증환자는 20명임을 분석하였다.

셋째, 당뇨병 합병증에 대한 위험지수 분석결과는 표 8과 같이 산출되었다.

표 8. 위험지수분석 결과  
Table 8. The result of risk quotient analysis

No	Diseases				Risk quotient	Patient	Message
	diabetes (0.5)	High BP (0.2)	Heart disease (0.2)	ST depression (0.05)			
1	○	○	○	○	0.95	5	Emergency
2	○	○	○	×	0.9	5	
3	○	○	×	○	0.75	10	Notice
4	○	×	○	○	0.75	13	
5	○	○	×	×	0.7	15	
6	○	×	○	×	0.7	17	
7	○	×	×	○	0.55	20	
8	○	×	×	×	0.5	40	

그림 9는 표 8의 위험지수와 환자인원에 대한 결과를 그래프로 설명한 것이다. 위험지수의 산출결과에 따라 환자, 주치의에 응급 메시지는 10건 지속적으로 전송되었고, 알림 메시지는 115건이 1회 전송되었음을 알 수 있다.

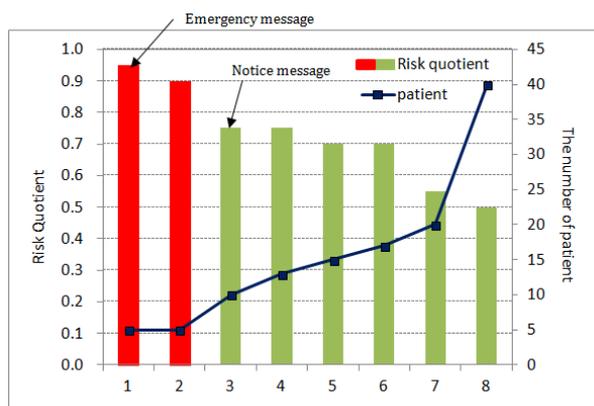


그림 9. 위험지수 분석 결과  
Fig. 9. The result of risk quotient analysis

#### 4.2. 질병규칙분석

본 논문에서 제안한 RDRGM의 질병규칙 실험 데이터는 Wisconsin Diagnostic Breast Cancer (WDBC)의 Breast Cancer<sup>[17]</sup>를 이용하였다. 표 9는 본 논문의 실험에 사용된 데이터 Breast Cancer의 속성들, 결정속성 그리고 각 속성들의 값에 대해 설명이다.

표 9. 규칙 실험 데이터의 속성들과 값  
Table 9. The attributes and values of rule experiment data

Field Name	Description	Data Ranges
a1	clump thickness	1-10
a2	Uniformity of cell size	1-10
a3	Uniformity of cell shape	1-10
a4	Marginal Adhesion	1-10
a5	Single Epithelial cell size	1-10
a6	Bare Nuclei	1-10
a7	Bland Chromatin	1-10
a8	Normal Nucleoli	1-10
a9	Mitoses	1-10
d	Class	b(benign), m(malignant)

이 실험 데이터는 benign이 458개, malignant가 241개로 총 699개로 구성되어 있으나, 이중 16개의 데이터가 missing attribute value를 갖고 있다. 본 논문에서는 missing attribute value를 갖는 데이터 16개를 제외시킨 683개의 데이터 중에서 랜덤하게 100개의 데이터를 추출하였고, RDRGM과 C4.5 알고리즘을 이용하여 질병규칙을 생성하였다. 그리고 이렇게 생성된 질병규칙들은 missing attribute value를 갖는 데이터 16개를 제외시킨 683개의 데이터에 적용시켜 정확도를 평가하였다.

표 10은 랜덤하게 추출한 100개의 데이터를 이용하여 C4.5알고리즘과 본 논문에서 제안한 RDRGM에 의해 생성된 5개의 규칙들이다. 이 5개의 규칙을 100개의 데이터에 적용시켜 정확도를 평가한 결과로써 정확도가 5개의 규칙 중에서 C4.5 rule 1은 98%, RDRGM rule 5는 93%, 그리고 나머지 규칙들은 100%로 나타났다. 표 11은 표 10의 규칙들을 683개의 데이터에 적용시켜 정확도를 평가한 결과로써 정확도가 RDRGM이 91.80%로 89.36%인 C4.5보다 정확한 것으로 나타났다. 따라서 본 논문에서 제안한 RDRGM이 기존의 C4.5보다 우수하다고 할 수 있다.

표 10. 질병 규칙들  
Table 10. The disease rules

C4.5 rules	Rule 1	if (a6<1.5) then (d=b)(98%)
	Rule 2	if (a6>1.5 and a3<1.5) then (d=b)(100%)
	Rule 3	if (a6>1.5 and a3>=1.5 and a1<5.5 and a6<9) then (d=b)(100%)
	Rule 4	if (a6>1.5 and a3>=1.5 and a1<5.5 and a6>=9) then (d=m)(100%)
	Rule 5	if (a6>1.5 and a3>=1.5 and a1>=5.5) then (d=m)(100%)
RDRGM rules	Rule 1	if (a2<=1 and a1<=5) then (d=b)(100%)
	Rule 2	if (a2<=1 and a1>5) then (d=m)(100%)
	Rule 3	if (a2>1 and a3<3 and a6<5) then (d=b)(100%)
	Rule 4	if (a2>1 and a3<3 and a6>=5) then (d=m)(100%)
	Rule 5	if (a2>1 and a3>=3) then (d=m)(93%)

표 11. RDRGM과 C4.5 알고리즘의 정확도 비교  
Table 11. The accuracy comparison of RDRGM and C4.5 Algorithm

Method	C4.5	RDRGM
Accuracy(%)	89.36%	91.80%

## V. 결 론

본 논문에서는 헬스케어시스템에서 Rough Set Theory를 이용하여 개인맞춤형 의료서비스와 정보를 제공할 수 있는 Rough Set Theory 기반 질병 모니터링 시스템인 RDMS를 제안하였다. RDMS는

DCM, RDRGM, 그리고 HMM로 구성된다.

DCM은 바이오센서로부터 심전계, 뇌파계, 근전계, 혈당계, 혈압계에 대한 데이터와 환자의 질병에 대한 정보 및 환자 가족력을 수집하여 RDMS DB에 저장하도록 설계하였다. 이때, DCM은 데이터의 입력오류를 확인하고, 입력된 환자의 생체정보가 허용되는 범위이내 데이터인지도 확인하여 RDMS DB에 저장함으로써 HMM의 위험지수 분석 결과 정확도를 향상시킬 수 있다.

RDRGM은 각 속성과 결정 속성간의 지지율을 계산하여, 지지율이 낮은 속성은 제거한 후에, 질병을 판단할 수 있는 질병 규칙을 생성하도록 설계하였다. 그 결과, 기존의 C4.5 알고리즘이 갖는 속성간 관계를 무시하는 문제점을 보완하였을 뿐만 아니라 불필요한 데이터를 제외시킴으로써 좀 더 정확한 질병 규칙을 생성하여 기존의 C4.5 알고리즘으로 생성한 질병규칙보다 정확성을 향상시켰다.

HMM은 DCM이 수집한 데이터를 실시간으로 정상 기준치와 비교하여 환자의 위험 지수를 분석하도록 설계하였다. 특히 위험지수분석은 특정 질병에 대한 합병증까지 분석한 후, 위험지수 분석 결과에 따라 알림메시지와 경고메시지를 환자와 주치의에게 1회 또는 지속적으로 전달함으로써 환자의 건강상태를 피드백 하였다. 또한, HMM은 환자의 상태에 따라 현재 환자가 조치할 수 있는 응급조치방법을 알려줌으로써 불의의 사고를 미연에 방지할 수 있으며, RDRGM이 생성한 질병규칙을 활용하여 환자의 병력, 환자 가족력 등의 정보와 적용하여 환자의 질병을 예방하거나 개인맞춤형 의료서비스를 제공할 수 있다.

## References

- [1] H. Seo, D.-H. Ryu, and T. Choi, "Design and implementation of u-Healthcare system for u-Wellness," *J. Korea Academia-Industrial Cooperation Soc. (KAIS)*, vol. 13, no. 11, pp. 5506-5511, Nov. 2012.
- [2] Z. Pawlak, "Rough sets," *Int. J. Comput. Inform. Sci. (IJCIS)*, vol. 11, no. 5 pp. 341-356, Sep. 1982.
- [3] Z. Pawlak, "Rough sets and decision algorithms," *Lecture Notes Comput. Sci.*, vol. 2005, pp. 30-45, 2001.
- [4] B. Walczak and D.L. Massart, "Tutorial

- Rough sets Theory,” Chemometrics and Intelligent Laboratory System, vol. 47, no. 1, pp. 1-16, Jan. 1999.
- [5] Q. Shen and R. Jensen, “Rough sets, their extensions and applications,” *Int. J. Automation Comput. (IJAC)*, vol. 4, no. 3, pp. 217-228, July 2007.
- [6] S.-J. Oh and C.-W. Park, “Development of automatic rule extraction method in data mining : an approach based on hierarchical clustering algorithm and rough set theory,” *J. Korea Soc. Comput. Inform. (KSCI)*, vol. 14, no. 6, pp. 135-142, June 2009.
- [7] S.-W. Han and J.-Y. Kim, “A new decision tree algorithms based on rough set theory and entity relationship,” *J. Korean Inst. Ind. Eng. (KIIE)*, vol. 33, no. 2, pp. 183-190, June 2007.
- [8] R. Yellasiri, C. R. Rao, and V. Reddy, “Decision tree induction using rough set theory - comparative study,” *J. Theoretical Appl. Inform. Technol.*, vol. 3, no. 4, pp. 110-114, Dec. 2007.
- [9] S. H. Kim, “Biological signal monitoring technique for ubiquitous healthcare,” *IT SoC Mag.*, vol. 25, pp. 40-47, 2008, [Online]. Available: [http://www.ssforum.org/business/file/vol25\\_Design\\_Methodology.pdf](http://www.ssforum.org/business/file/vol25_Design_Methodology.pdf).
- [10] K.-M. Kim, “Implementation of patient monitoring system based on mobile healthcare,” *J. Korea Soc. Comput. Inform. (KSCI)*, vol. 17, no. 2, pp. 1-10, Dec. 2012.
- [11] J. R. Quinlan, *C4.5: Programs for Machine Learning*, Morgan Kaufmann Publishers, 1993.
- [12] H.-J. Jin and H. Y. Kim, “Relationship between impaired fasting glucose and periodontal health among adults,” *J. Korea Academia-Industrial Cooperation Soc. (KAIS)*, vol. 12, no. 11, pp. 5034-5042, Nov. 2011.
- [13] M.-K. Kim, “Diabetes complications-strike!,” *A Monthly Diabetes Mag.*, vol. 2010, no. 11, pp. 42-43, Nov. 2010.
- [14] Diabetic Association of Korea, “Horrible of complications of diabetes-heart disease,” *A Monthly Diabetes Magazine*, vol. 2011, no. 1, pp. 22-25, Jan. 2011.
- [15] American Diabetes Association, “Diagnosis and classification of diabetes mellitus,” *Diabetes Care*, vol. 33, no. 1, pp. 562-569, Jan. 2010.
- [16] University of California, Irvine, *Heart Disease*, retrieved Sep., 1, 2013, from <http://archive.ics.uci.edu/ml/machine-learning-databases/statlog/heart/>.
- [17] University of California, Irvine, *Wisconsin Diagnostic Breast Cancer*, retrieved Sep., 1, 2013, from <http://archive.ics.uci.edu/ml/machine-learning-databases/breast-cancer-wisconsin/>.

**이 병 관 (Byung-Kwan Lee)**



1979년 2월 부산대학교 기계설계학과 졸업  
 1986년 2월 중앙대학교 전자계산공학과 석사  
 1990년 2월 중앙대학교 전자계산공학과 박사  
 1988년 3월~현재 관동대학교

공과대학컴퓨터학과 교수

<관심분야> 네트워크 보안, 인터넷 보안, 빅데이터, IoT

**정은희 (Eun-Hee Jeong)**



1991년 2월 강릉대학교 통계학과 졸업  
 1998년 2월 관동대학교 전자계산공학과 석사  
 2003년 2월 관동대학교 전자계산공학과 박사  
 2003년 9월~현재 강원대학교

인문사회과학대학 지역경제학과 부교수

<관심분야> 인터넷 보안, 전자상거래 보안, 데이터 마이닝, 빅데이터 보안