

고혈압관리를 위한 의사지원결정시스템의 데이터마이닝 접근

김태수* · 채영문* · 조승연** · 윤진희*** · 김도마*

Data Mining Approach to Clinical Decision Support System for Hypertension Management

Tae S Kim* · Young M Chae* · Seung H Cho** · Jin H Yoon*** · Thoma Kim*

* Graduate School of Health Science and Management, Yonsei University

C.P.O. Box 8044, Seoul Korea, e-mail : tskim@yumc.yonsei.ac.kr

** Department of Cardiology, Yonsei University College of Medicine

*** Department of Mathematics, Yonsei University

요약

This study examined the predictive power of data mining algorithms by comparing the performance of logistic regression and decision tree algorithm, called CHAID (*Chi-squared Automatic Interaction Detection*). On the contrary to the previous studies, decision tree performed better than logistic regression. We have also developed a CDSS (Clinical Decision Support System) with three modules (doctor, nurse, and patient) based on data warehouse architecture. Data warehouse collects and integrates relevant information from various databases from hospital information system (HIS). This system can help improve decision making capability of doctors and improve accessibility of educational material for patients.

Key words : Hypertension, Data mining, Decision support system

1. 서론

고혈압은 복잡한 현대사회를 사는 성인들에게 호 발하는 만성퇴행성질환(chronic degenerative)으로 구미의 경우 전 인구의 15-20%가 고혈압환자인 것으로 알려져 있으며, 국내에서는 성인 인구의 20% 내외가 고혈압을 갖고 있는 것으로 추정되고, 특히, 고혈압성질환은 주요한 사망원인을 나타내고 있다. 이와 같이 높은 유병률을 차지하고 있는 고혈압은 본태성 고혈압(essential hypertension)과 이차성 고혈압(secondary hypertension)으로 대별하며, 그 중 85-90%가 본태성 고혈압이고 나머지 10-15%가 이

차성 고혈압이다. 고혈압은 유전인자와 환경인자에 따라 영향을 많이 받는 질환이며 고혈압의 전단계 를 지나 초기 고혈압, 확정적 고혈압 그리고 병발 증성 고혈압의 단계적 변화를 밟는 질환이다. 고혈 압은 그 자체로도 중요한 관리의 대상이 되기는 하 나 특히 이것은 심장질환이나 뇌혈관질환, 그리고 신장질환 같은 주요 사망원인 질환 발생과 깊은 관 련이 있다는데 더 큰 관심의 대상이 된다. 이와 같 이 고혈압의 치료 및 관리에 대한 유의성은 입증되 었으나, 지금까지 지속적으로 이루어진 고혈압의 인지도, 치료 및 관리현황과 관련된 행태조사를 살 펴보면 많은 나라에서 발표된 것 같이 고혈압 환자 의 70%이상이 불완전한 조절 상태, 또는 전혀 치

료를 받고 있지 않은 상태에 머물러 있고, 고혈압이 조절되고 있는 비율은 증가되지 않고 어떤 경우에 있어서는 오히려 낮아지고 있다는 것은 염려되는 점이다. 고혈압 환자의 상당수가 아직 치료를 받지 않은 상태이며, 치료중인 환자 중 많은 수에서 여전히 혈압이 높게 유지되고 있다. 지역적인 차이는 있으나, 평균적으로 치료 중인 환자의 약 반에서 혈압이 160/95mmHg 이상으로, 약 3/4에서 140/90mmHg 이상으로 유지되고 있다. 우리나라의 경우도 고혈압 유병률이 높음에도 불구하고 고혈압 환자의 75%가 자신의 혈압수준을 모르고 있다. 또한, 어떠한 종류의 치료라도 받고 있는 사람은 16%에 불과하고, 혈압이 적절히 관리되고 있는 사람은 5%밖에 되지 않는다고 한다. 지금까지의 연구는 대부분 치료의 상대적 효과를 측정할 무작위화 임상시험(randomized controlled trials)이 대부분이었는데, 그러나 의사나 환자에게 큰 관심 사항은 치료의 절대적인 효과로서 각 개인 환자에서 치료의 절대적인 효과는 여러 연구에서 밝혀진 환자에게 있는 위험인자들을 치료함으로써 얻어지는 상대적인 효과를 모두 더함으로써 예측할 수 있다. 즉 고혈압 환자의 치료는 여러 가지 복합적 요인이 작용하여 이루어지는 만큼 이러한 요소들을 고려하여, 다양한 차이를 나타내는 각 개인을 치료하는데 있어 경직된 규칙보다는 임상가를 이끌 수 있는 복합적 정보를 제공할 수 있는 지침 및 연구가 필요하다. 즉 고혈압의 진단 및 치료에 관여되는 요인들이 매우 많기 때문에 이들 정보를 종합하여 임상 의사나 보건관리자에게 제공할 수 있는 시스템과 이를 지원할 수 있는 새로운 정보체계가 필요하다.

최근 의학의 발달과 전문화와 함께 의학지식이 더욱 세분화되었을 뿐만 아니라 새롭고 다양한 진단과 치료방법이 개발됨에 따라 질병관리에 고도의 광범위한 전문지식이 필요하게 되었다. 이에 따라 고도의 전문지식을 공유하여 활용할 수 있는 방법으로 의사결정지원시스템(Decision Support System)에 대한 연구가 진행중이다. 의료분야에서도 단순 정보 제공보다는 보다 해석적이며 감별 진단 및 치료에 도움을 줄 수 있는 의료상담의 역할을 수행하는 시스템이 요구된다. 진료를 위한 의사결정지원시스템(Clinical Decision Support System)은 임상 의사가 환자를 진료할 때 행하는 과정중에 이루어지는 의사결정에 지식을 제공함으로써 이를 지원할 수 있는 시스템을 말한다. 현재까지 의사결정지원시스템을 개발하기 위하여 많이 활용된 인공

지능기법은 규칙기반추론이다. 규칙기반 추론기법에서는 전문가의 경험적인 휴리스틱 지식을 생성규칙(Production Rule)의 형태로 표현하는데, 이는 전문가의 지식이 체계적으로 잘 정리되어 있는 경우에는 별 문제가 없지만, 그렇지 못한 경우에는 그만큼 영역전문가(Domain Expert)로부터의 지식 획득에 의존해야 하므로 이로 인한 여러 한계점이 제기되고 있다. 필요한 지식을 임상에 적용시키기 위하여서는 문제해결을 위한 의사결정과정과 관련된 지식의 개념을 정립하여 표현해야 하는 어려움이 있다. 즉, 문헌과 환자의 사례, 과거의 경험 등 여러 형태의 지식의 원천으로부터 필요한 지식을 추출하여 이를 구조적으로 조직화하여 지식베이스화하는 것은 매우 힘든 과정이라 할 수 있다. 지식 획득방법의 핵심적 요소는 문제를 해결해 나가는 데 있어 도움이 될 수 있는 필요정보를 도출해 내는 일이다. 지식획득의 전과정에 있어서 시스템 개발자는 적절한 지식획득 방법을 이용하여 임상 의사가 그러한 처리과정을 효율적으로 발견하고 표현할 수 있도록 하여야 한다. 이와 같이 의사결정지원시스템을 개발하는 데 있어서 지식베이스의 구축은 가장 어려운 분야중의 하나라고 할 수 있다. 과거부터 현재까지 의료분야의 의사결정지원시스템이 질병의 진단과 치료 및 예후 예측에 있어 완벽한 의사결정지원기능을 수행하지 못하였던 것은 이와 같은 개발상의 어려움에 기인한다고도 볼 수 있다. 따라서 최근 들어서는 이러한 한계를 극복하기 위해 여러 대안적 기법을 활용한 시스템 개발 연구가 수행되고 있으며, 그 대표적인 연구가 임상자료로부터 직접 지식을 추출해 내는 방법이다. 정보시스템의 급격한 발전과 더불어 많은 양의 데이터가 축적되기 시작하면서, 필요한 정보를 찾아내어 가치 있는 지식으로 승화시키는 것에 대한 필요성이 야기되고 있다. 즉 대량의 데이터를 지식으로 효과적으로 저장, 관리, 활용할 수 있는 지식기반 경영시스템의 필요성이 대두되고 있다. 지식을 바탕으로 지속적인 업무지식을 창출하고, 효과적으로 지식기반 경영시스템을 운용할 수 있는 전사적인 지식경영(Knowledge Management)의 비전 및 이의 전략수립이 요구되고 있다. 또한 최근 몇 년에 걸쳐 대기업을 중심으로 대량의 데이터를 포함하고 있는 정보시스템을 구축, 관리할 수 있는 데이터웨어하우스(Data Warehouse)가 새로운 아키텍처로 각광받고 있으며, 이와 함께 방대한 양의 데이터로부터 함축적이며 잠재적 유용성이 있는 정보를 발

견할 수 있는 지식 탐사 방법인 데이터 마이닝(Data Mining)에 대한 중요도가 더욱 증대되고 있다. 데이터마이닝은 패턴인식 기술이나 통계기법, 수학적 알고리즘을 이용하여 의미 있는 새로운 상관관계, 패턴, 추세 등을 발견하는 과정으로, 의료분야에서도 다른 분야와 마찬가지로 의료정보자원의 효율적 활용을 위한 지식기반 경영과 의료 정보기술로써 데이터웨어하우스와 데이터마이닝의 중요성이 부각되고 있다. 본 연구에서는 의료분야의 의사결정지원을 위한 지식경영체계 도입을 촉진시키기 위한 방안으로써 데이터마이닝 기법을 적용하여 대규모 데이터 내에 존재하지만 숨겨져 있는 상호관련성과 패턴에 대한 탐색을 통해 유용한 지식을 이끌어 내하고자 하였다. 의료분야에서의 데이터 마이닝 응용 분야로는 의료이용도 분석, 삭감률 분석, 질병 패턴 분석, 건강증진 관련 분석, 경영 분석 등이 있을 수 있으며, 본 연구에서는 질병 패턴 분석의 일환으로 임상적으로 중요한 위치를 차지하고 있는 고혈압 환자의 특성과 이에 따른 예후를 예측할 수 있는 지식을 발굴하고자 하였다.

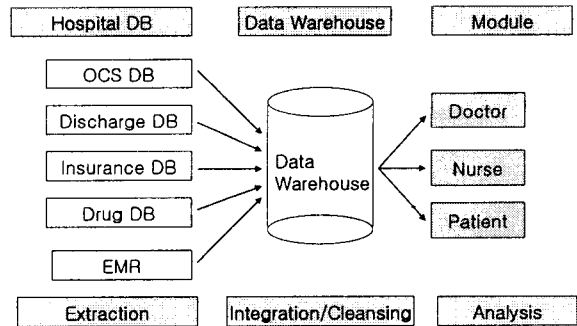
본 연구에서는 데이터마이닝 기법을 적용하여 치료결과에 어떠한 요인이 얼마나 중요한 지를 유추해 내고 치료 후 혈압조절군과 비조절군의 특성을 분류, 규명함으로써 새로운 치료 전략을 도출해 보고자 한다. 또한 여러 데이터마이닝 기법들의 결과에 대한 성과를 비교 분석하고 향후 진료분야의 지식경영으로서 고혈압 관리를 위한 의사결정지원 시스템 및 임상진료지침의 개발에 활용될 수 있는 최적 모형의 설계방안을 모색해 보고자 한다.

2. 연구 대상 및 방법

1999년 10월부터 2002년 5월 까지 일병원에 입원한 환자중에서 주진단(primary diagnosis)이나 부진단(secondary diagnosis) 중 고혈압이 있는 입원 환자 2,507명과 심장내과에 내원한 환자 중 고혈압이 의심되거나 또는 고혈압을 지니고 있어 설문조사에 응한 3,786명을 연구 대상으로 하였다.

그림1과 같이 의사결정지원시스템은 여러개의 원시데이터베이스들과 데이터웨어하우스, 그리고 의사, 간호사, 환자의 3개의 모듈로 구성되어 있다. 데이터웨어하우스는 데이터마이닝에 필요한 통합적 데이터의 수집과 전처리 과정을 거치기 수행하기 때문에 데이터웨어하우스 기반에서 데이터마이닝을

수행하는 것이 바람직하다. 대부분의 데이터마이닝 그림 1. 고혈압관리 시스템의 구조



의 시간이 이러한 데이터의 준비과정에서 소모되기 때문이다. 최근에 데이터와 정보는 폭발적인 증가를 하고 있으며 이러한 데이터와 정보의 증가는 데이터 저장소의 역할과 그 영역의 증대를 가져왔다. 그러나 이러한 데이터 저장소의 역할과 영역 확대는 필요로 하는 데이터를 찾기 어렵게 한다는 문제점을 야기시켰을 뿐만 아니라 데이터 처리를 위한 시간과 비용을 크게 증대시켰다. 이와 같은 데이터 크기의 비약적인 증대는 사용하고자하는 데이터의 추출과 데이터 저장공간 확보 등의 문제점을 유발하였고 이러한 문제를 해결하기 위해 데이터베이스 관리시스템, 데이터웨어하우스 그리고 데이터마이닝과 같은 데이터관리와 데이터추출에 관한 연구가 진행되었다. 이 중 데이터마이닝은 대량의 데이터 사이에 묻혀 있는 패턴을 발견하고 규칙을 추론함으로써 의사결정을 지원하고 그 효과를 예측하기 위한 기법으로 지금까지 대답이 어려웠던 많은 질문에 대한 답을 제공한다. 본 연구에서는 데이터마이닝 프로세스중 군집분석(Clustering analysis), 의사결정나무분석(Decision tree analysis), 그리고 연관성규칙(Association rules)을 이용하고, 이와의 비교를 위하여 통계적 방법으로서 로지스틱회귀분석(Logistic regression analysis)을 한다. 군집분석은 주어진 자료들 중에서 유효한 것들을 몇몇의 집단으로 그룹화하여, 각 집단의 성격을 파악함으로써 데이터 전체의 구조에 대한 이해를 돕고자 하는 방법이다. 대용량데이터에서 객개의 관찰치를 요약하는 것 보다는 전체를 유사한 관찰치들의 군집으로 구분하여 각 군집들의 특성을 파악함으로써 전체 데이터에 대한 의미있는 정보를 얻을 수 있다. 전통적으로 이분형인 종속변수와 여러 가지 독립변수와의 관계를 알아보하고자 하는 방법으로서 많이 사용되어지는 것이 로지스틱 회귀

분석이다. 혈압변이를 목표변수로 하여 혈압조절군과 비조절군으로 분류하여 각 군집의 대상군별 환자 및 치료의 특성을 분석하고, 종속변수 혈압조절여부와 환자의 특성인 여러 독립변수와의 관계를 알아보고자 로지스틱 회귀분석을 실시하였으며, 이와 함께 비교하게 될 의사결정나무분석은 의사결정규칙(decision rule)을 나무구조로 도표화하여 분류(classification)와 예측(prediction)을 수행하는 분석 방법으로 예측의 과정이 나무구조에 의한 추론규칙(induction rule)에 의해 표현된다. 데이터마이닝에서의 의사결정나무는 탐색(exploration)과 모형화(modeling)의 특성을 지니며, 사전에 이상치(outlier)를 검색하거나 분석에 필요한 변수를 찾아내고 분석모형에 포함되어야 할 교호효과를 찾아내는 데 사용될 수 있고, 그 자체가 분류 또는 예측 모형으로 사용될 수도 있다. 1980년대 이후 CHAID, CART, C4.5 등 다양한 알고리즘이 제안되어 왔는데, 본 연구에서는 변수가 범주형 데이터이고 예측변수와 목표변수간의 관계를 찾아야 할 때 가장 유용한 방법인 CHAID 방법을 적용하였다. 마지막으로 연관성 규칙은 데이터에 안에 존재하는 항목간의 유용한 관련 규칙을 발견하고자 할 때 이용될 수 있는 기법으로서 의사결정나무 모형에서 추출된 지식과 임상이론 및 고혈압 전문의의 지식을 비교하고 검증하는 방법으로 사용하였다.

3. 결과

고혈압 환자의 진단과 치료에 있어서 중요한 특성으로 파악되어지는 항목들을 선별하는 과정은 임상진료지침(clinical practice guideline) 등의 임상이론, 고혈압 전문의의 견해, 실제 환자 데이터에 대한 기술통계량(descriptive statistics) 측정을 통한 데이터의 특성 파악 등을 통해 이루어졌다. 입원 환자와 외래환자의 연구대상에 대하여 각 항목과 관련된 기본 통계량을 조사한 결과 다음의 표1,2,3과 같은 특성을 나타내고 있다. 두 집단은 이질적인 특성을 보이므로 같은 환자군으로 병합하여 분석을 하기 보다는 각각 개별적인 집단으로 보고 각 집단의 특성을 살펴보기로 한다.

표 1. 증상, 가족병력등에 따른 환자기본분류

표 2. 이학적 검사 결과

측정변수	변수값	입원환자(환자수)	내원환자(환자수)
BMI	과체중 이하	30.69(%)	26.71(%)
흉부촬영	심비확대	13.51(54)	9(21)
	폐혈종	1.06(46)	5(89)
	심비대&폐혈종	4.46(64)	2(33)
EKG	Ischemia	16.43(30)	31(27)
	Arrhythmia	7.15(63)	8(34)
	Ischemia&Arrhythmia	0.37(7)	4(89)
ECHO	비정상	15.18(39)	40(27)
안저촬영	비정상	1.84(46)	2(80)
주호소증상	현기증	3.48	25.19
	시력장애	0.33	19.24
	부종	4.35	14.94
	호흡곤란	11.97	23.92
	흉통	20.02	23.92
	발저림	0.33	26.71
	빈맥	2.94	26.96
	마비	4.24	0.76
가족병력	실어증	1.96	0
	실신	0.87	0
	고혈압	15.78	45.82
	심혈관질환	02.83	9.62
	신질환	0	0.76
	당뇨	1.09	9.49
동반질환	고지혈증	0	0.89
	뇌졸중	0.44	12.53
	뇌혈관질환	21.44	4.94
	심장질환	49.62	38.73
	신장질환	21.65	1.65
	혈관질환	6.31	1.01
	고혈압성망막증	0	0.38
	당뇨	28.94	5.06
흡연여부	고지혈증	6.09	0.89
	통풍	0.65	0.51
	과거	4.68	26.46
음주여부	현재	21.87	11.27
	음주	24.81	31.52

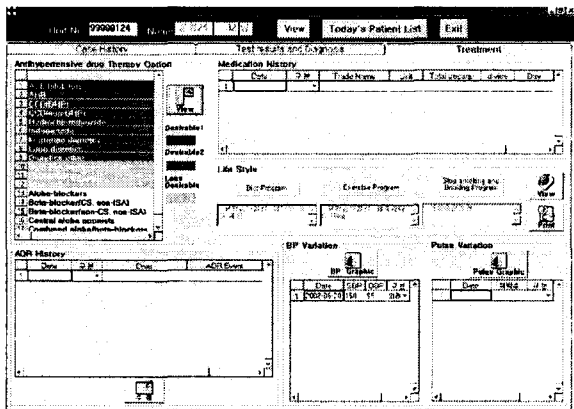
표 3. 임상검사 결과

위에 언급된 고혈압 환자의 진단과 치료에 있어서 중요한 항목들을 기초로 하여 임상지식 및 고혈압 전문의의 지식 등 관련 정보를 체계화된 지식으로 표현함으로써 의사결정지원시스템의 지식베이스의 생성 규칙을 이끌어냈다. 의사모들은 WHO와 JNC(Joint National Committee on Prevention, Detction, Evaluation, and Treatment of High Blood Pressure)에 근거한 임상처방을 기초로 한 진단 정보를 제공한다. 그림 2는 의사모들에서 각 증상에 대하여 우선순위(Priority)에 기초하여 약제 처방을 제안해주는 화면이다. 이곳에서는 해당하는 약제에 대한 성상, 용법, 그리고 부작용등을 알 수

측정변수	입원환자 (Mean, S.D)		외래환자 (Mean, S.D)	
BUN	24.69	19.94	17.40	8.79
Calcium	8.97	0.75	9.32	0.51
Cholesterol	178.03	51.53	199.02	39.98
Creatinine	2.35	2.90	1.24	0.70
GlucoseAC	137.48	76.83	122.15	63.30
HbA1c	7.60	2.00	7.75	1.98
HDLCholesterol	43.57	14.02	46.32	11.57
K	4.16	0.67	4.22	0.47
LDLCholesterol	106.47	34.11	111.68	31.47
Na	138.64	3.88	140.04	2.81
Protein	240.71	399.40	152.79	321.96
TProtein	6.54	0.90	7.26	0.72
Triglyceride	145.93	100.85	163.00	95.15
b_Hct	35.50	2.40	38.71	5.82
Hemoglobin	11.80	2.40	13.28	1.84
RBC	3.93	0.78	4.27	0.63
WBC	8.35	3.97	7.07	2.26

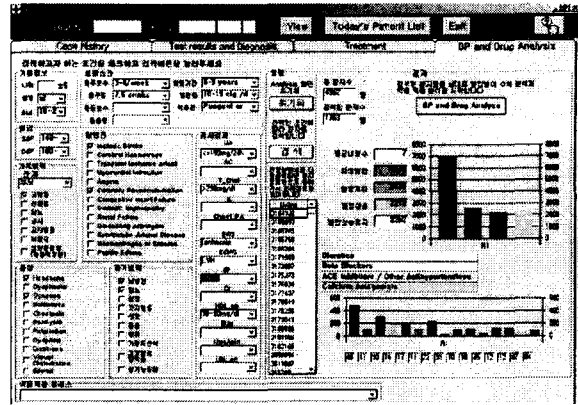
도 있다.

그림 2. 의사 모듈에서의 약제처방의 화면



또한 의사모듈에서는 데이터베이스를 검색하여 여러 증상에 해당하는 환자들을 검색하여 약제처방의 결과 발생되는 혈압의 변이에 대한 통계량 및 개개 환자의 자세한 특성을 검색 할 수 있도록 구성 하였다.

그림 3. 의사모듈에서의 환자검색화면



입원환자와 외래환자의 연구대상에 대하여 각 항목과 관련된 기본 통계량을 조사한 결과 입원환자군과 외래환자군의 두 연구대상집단에 대하여 혈압변이의 변수를 목표변수로 하여 군집분석을 실시하여 혈압조절군과 비조절군으로 분류하여 특성을 살펴 보면 각각 표4,5와 같다. 표4와 5에서는 두 집단간의 변수중에서 통계적으로 유의한 차이를 보이고 있는 변수들만을 나열 하였으며, 이들의 변수는 다음에 진행할 의사결정나무분석과 로지스틱회귀분석 주로 사용될 독립변수들을 의미한다.

표 4. 입원환자의 혈압변이에 따른 군집분류결과

변수	혈압비조절군		혈압조절군	
	Mean	S.D	Mean	S.D
혈압변이	5.33	16.23	47.38	20.15
Initial SBP	146.55	14.05	173.08	25.04
Initial DBP	89.68	10.03	99.00	17.16
Headache	0.08	0.27	0.13	0.34
Family Diabets	0.16	0.36	0.10	0.30
Protein UA	0.39	0.86	0.51	1.01
Glucose AC	61.30	99.09	80.29	106.39
T.Cholesterol	120.12	95.44	137.99	91.37
EKG	0.08	0.27	0.15	0.35
ECHO	0.09	0.29	0.15	0.35
HDL	27.03	27.60	30.37	23.29
UricAcid	3.48	3.25	3.96	3.21

표 5. 외래환자의 혈압변이에 따른 군집분류결과

위 군집분류의 결과를 기초로 하여, 의사결정나무분석과 로지스틱회귀분석의 비교를 통하여 표 6에서는 두 연구집단에 대하여 의사나무결정분석이 로지스틱회귀분석보다 모형에 대한 오차도 적을 뿐 아니라 오분류율도 작음

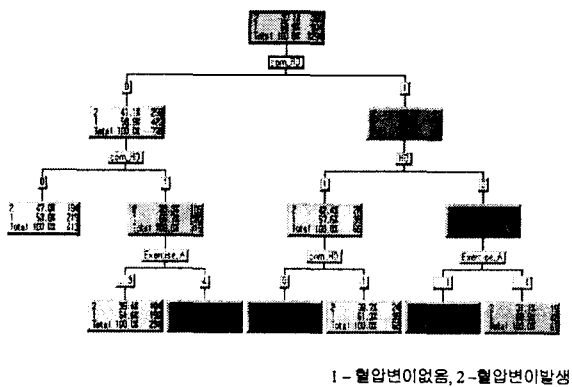
변수	혈압비조절군		혈압조절군	
	Mean	S.D	Mean	S.D
혈압변이	20.00	15.03	67.79	19.74
Gender	1.44	0.50	1.58	0.49
Age	61.37	10.93	62.56	10.64
Initial SBP	152.05	14.70	189.30	21.73
Initial DBP	93.17	11.24	106.72	23.94
Blood Disease	0.06	0.23	0.09	0.28
Heart Disease	0.29	0.45	0.45	0.50
Diabets	0.19	0.39	0.11	0.31
Family Hyper	0.37	0.48	0.49	0.50
Family Stroke	0.08	0.27	0.12	0.32
흡연량	0.56	1.06	0.43	0.96
운동량	0.66	1.20	0.82	1.24
com HD	0.49	0.50	0.35	0.48
EKG	0.30	0.46	0.38	0.48
ECHO	0.11	0.32	0.16	0.36
EF	4.61	16.73	8.45	22.55

을 알 수 있다.

표 6. 로지스틱회귀분석과 의사결정나무분석에 의한 오분류율

환자유형	모델	Root ASE	오분류율
입원환자	Logistic Reg.	0.5084	0.5722
	Decision Tree	0.4188	0.3978
내원환자	Logistic Reg.	0.4793	0.3139
	Decision Tree	0.4330	0.2967

그림 4. 외래환자에 대한 의사결정나무분석에 의한 나무구조



한 편 그림 4는 의사결정나무분석 결과로서 의사나무구조를 보여준다. 외래환자의 경우에 혈압변이에 영향을 주는 변수로는 신질환 및 심장질환의 동반질환의 유무와 주호소 증상으

로서 과거심장병의 유무 및 운동량등이 주요 영향 변수로 나타났다.

그림5는 두 모형에 관한 비교로서 리프트를 비교할 수 있다. 두 모형의 결과를 비교하여 보면 사후확률의 상위20%미만에서는 다소 로지스틱회귀분석이 좋은 결과를 보이지만 20% 이상에서는 의사결정나무분석이 더 나은 결과를 보여주고 있다. 입원환자의 경우에는 모든 구간에서 의사결정나무분석이 보다 더 좋은 결과를 보여 주고 있다.

그림 5. 외래환자에 대한 로지스틱회귀모형과 의사결정나무분석에 의한 리프트(lift) 곡선

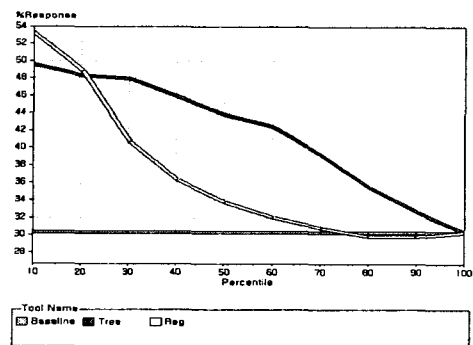
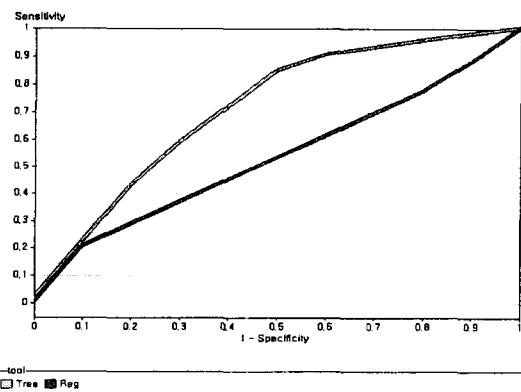


그림6은 민감도(Sensitivity)와 특이도(Specificity)를 이용하여 두 모형의 비교를 할 수 있는 ROC곡선으로서 좌측상단방향으로 곡선이 굽을 수록 좋은 모형을 나타내고, 직선에 가까우면 모형의 적용이 의미가 없음을 나타내고, 곡선이 우측하단으로 굽

그림 6. 외래환자에 대한 로지스틱회귀모형과 의사결정나무분석에 의한 ROC곡선



어 있다면 그 모형은 적용하므로써 아무런 것도 적

용하지 않았을 때 보다도 못함을 나타낸다. 따라서 외래환자의 경우에는 의사결정 나무분석이 로지스틱회귀모형보다 좋음을 알 수 있다. 입원환자군의 경우도 같은 결과를 얻을 수 있었다. 이와 같이 두 모형을 비교분석한 결과는 의사나무결정 분석이 로지스틱회귀분석 모형 보다 더 정확한 결과를 주고 있을 알 수 있다. 임상이론과 데이터마이닝 기법의 단독 및 혼합에 의한 처방규칙에 대한 치료결과를 분석함으로써 데이터마이닝 기법을 활용한 의사결정지원시스템의 타당성을 검증하였다. 이를 위해 일차적으로 기존 임상진료지침에서 제시하는 치료 방법에 따른 규칙기반시스템을 개발하고 이에 따른 치료 결과를 분석하였다. 각 적응증에 대해 해당 약제를 처방할 경우 혈압이 조절될 것이라는 이론 하에 설정된 것이라고 볼 수 있으며, 본 연구에서는 이러한 임상이론에 의한 처방규칙을 규칙화하여 시스템을 구현하고 이 때의 혈압조절군과 비조절군의 비율 비교하여 평가했다. 또한 데이터마이닝 기법을 통해 도출된 치료약제별 혈압조절군과 비조절군의 특성을 도출하고, 이 중 혈압조절군의 특성을 규칙화하여 제시하고 앞서와 마찬가지로 이 때의 혈압조절군과 비조절군의 비율 비교하여 평가한 후, 기존 임상이론과 데이터마이닝 기법을 통해 도출된 규칙 중 수용가능한 규칙과의 결합을 통한 혼합 모형을 개발하고 이에 대한 타당성 검증을 실시하였다. 위의 세가지 방법을 시행하여 각 방법에 의한 결과, 즉 혈압조절군의 비율 및 원인을 상호 비교 분석하여 고혈압 관리 의사결정지원을 위한 최적 모형을 설계·구현하고자 하였다.

표 7. Practice guideline for the drug treatment

Class of Drug	Indication		Contraindication
	Compelling	Possible	
Diuretics	<ul style="list-style-type: none"> • No complication • Elderly patients • Heart failure • Systolic hypertension 	<ul style="list-style-type: none"> • Protein uria 	<ul style="list-style-type: none"> • Gout • Hyperlipidemia
Beta blockers	<ul style="list-style-type: none"> • No complication • LVH • Angina 	<ul style="list-style-type: none"> • Elderly patients 	<ul style="list-style-type: none"> • Heart Failure • Hyperlipidemia • Peripheral vascular disease • Diabetes
Calcium channel blockers	<ul style="list-style-type: none"> • Elderly patients • Systolic hypertension • Hyperlipidemia 	<ul style="list-style-type: none"> • Peripheral vascular disease • DM • Angina 	
ACE inhibitors	<ul style="list-style-type: none"> • Heart failure • Protein uria • Diabetes 	<ul style="list-style-type: none"> • LVH • Hyperlipidemia 	<ul style="list-style-type: none"> • Hyperkalaemia

WHO에서 제시하는 고혈압 약제 처방에 대한 임상

진료지침을 요약하면 표 7과 같다. 각 약제에 대한 일차적응증(compelling indication)과 이차적응증(possible indication) 및 금기증(contra indication)이 제시되어 있다. 임상적으로 각 증상에 따라 단일처방 또는 복합처방을 한다. 표 9, 10은 입원환자군과 외래환자군에 대하여 각각 처방을 한후 결과에 따른 연관성 분석을 한 결과들의 일부이다. 임상적으로 시행되고있는 처방의 결과가 모두 혈압 조절의 효과를 가져오지는 않고 있음을 알 수 있다. 표9는 입원환자의 경우 많은 처방규칙들 중 다음과 같은 5번째에 해당하는 규칙에 대하여 규칙에 따른 처방을 하였을때, 혈압 조절이 되었을 경우를 나타내는 측정치들로서 지지도(Support), 신뢰도(Confidence), 그리고 향상도(Lift)를 보여주고 있다.

표 8. 고혈압시스템의 처방규칙의 예

Rule_5 : 증상 - LVH & Diabets & Angina

- First Desirable Drug : CCB(non-DHP) ,
- Second Desirable Drug : ACE Inhibitors, ARB, Beta-blocker(CS, non-ISA), Beta-blocker(non- CS, non-ISA), CCB(DHP), Hydrochlorothiazide, Indapamide, K-sparing diuretics, Loop diuretic, Diuretics-other,
- Less Desirable Drug : Alpha-blockers, Beta-blocker(non-CS, ISA), Hydralazine, Minoxidil

표 9. 입원환자군의 처방규칙에 대한 연관성분석

Lift	Support(%)	Confidence(%)	Rule
1.31	0.7	50	05thRule & ruleOk ==> delSBPOK
1.4	2.24	53.33	07thRule & ruleOk ==> delSBPOK
2.62	0.84	100	08thRule & ruleOk ==> delSBPOK
1.25	1.54	47.83	10thRule & ruleOk ==> delSBPOK
1.25	1.4	47.62	11thRule & ruleOk ==> delSBPOK
1.37	1.54	52.38	12thRule & ruleOk ==> delSBPOK
1.57	0.84	60	14thRule & ruleOk ==> delSBPOK

표 10. 입원환자군의 처방규칙에 대한 연관성분석

Lift	Support(%)	Confidence(%)	Rule
1.49	1.88	79.31	02thRule & ruleOk ==> delSBPOK
1.46	4.82	77.63	03thRule & ruleOk ==> delSBPOK
1.13	1.96	60	04thRule & ruleOk ==> delSBPOK
1.54	1.88	82.14	05thRule & ruleOk ==> delSBPOK
1.57	1.23	83.33	07thRule & ruleOk ==> delSBPOK
1.42	4.5	75.34	08thRule & ruleOk ==> delSBPOK
1.36	1.06	72.22	09thRule & ruleOk ==> delSBPOK
1.55	3.11	82.61	11thRule & ruleOk ==> delSBPOK
1.57	0.41	83.33	12thRule & ruleOk ==> delSBPOK
1.76	1.23	93.75	15thRule & ruleOk ==> delSBPOK

4. 결론

고혈압 상태에서의 병리 및 생리학적 변화나 그 치료방법의 엄청난 발전에도 불구하고 고혈압은 아직도 그 예방과 치료가 어려운 상태이며, 그것은 무엇보다 아직 그 원인과 이에 따른 치료방법이 확실히 밝혀지지 않은 데 있다. 따라서 고혈압 관리에서 위험요인을 찾고 이를 해결하기 위한 적절한 치료방법을 찾는 것은 무엇보다 중요하다고 할 수 있다. 이에 본 연구에서는 데이터마이닝 기법을 적

용하여 치료결과에 어떠한 요인이 얼마나 중요한지를 유추해 내고 치료후 혈압조절군과 비조절군의 특성을 분류, 규명함으로써 새로운 치료 전략을 도출해 보고자 하였으며, 이를 이용한 의사결정지원 시스템을 개발하고 이의 유용성을 평가하고자 하였다. 이에 따라 고혈압 환자를 대상으로 로지스틱 회귀분석과 의사결정나무분석 기법을 이용하여 혈압조절상의 위험요인을 규명하고, 치료약제별로 혈압조절군과 비조절군을 결정짓는 규칙을 도출하여 실제 임상이론 및 환자 사례와의 비교검증을 통한 예측력을 평가하였다. 그리고 임상이론 및 데이터마이닝 기법을 통해 도출된 지식을 기반으로 하는 지식베이스를 구축하고 추론엔진을 구성하여 의사결정지원시스템을 개발하고 임상이론과 데이터마이닝 기법의 단독 및 혼합에 의한 처방규칙에 대한 치료결과를 분석함으로써 데이터마이닝 기법을 활용한 의사결정지원시스템 개발에 대한 타당성 검증을 실시하였다. 임상진료지침을 따른 경우의 치료결과에서 혈압 조절군의 비율이 예상보다 높지 않은 것은 고혈압의 치료약제에 대한 효능효과 및 적응증에 대한 지속적인 연구와 평가가 이루어져야 함을 시사한다. 또한 이러한 과정이 지금까지와 같이 주관적으로 치우치거나 현재까지 알려진 혹은 임상적 이론에 의거하는 것 보다는 방대한 데이터로 이루어진 데이터웨어하우스로부터의 지식 획득을 통하여 이루어져야 함을 제시한다. 그리고 본 연구에서 고혈압의 특성을 대표할 수 있는 대상자의 선정과 이들의 최적을 특성을 선별하고 이에 대한 충분한 임상적 평가가 이루어졌는지에 대한 충분한 검토가 이루어져야 함을 나타낸다.

본 연구에서는 여러 기존 임상이론 및 데이터마이닝 기법을 통해 추출된 규칙을 활용한 각 시스템 및 혼합모형에 대한 결과를 비교 분석하고 타당성 검증을 실시함으로써 향후 진료분야의 지식경영으로서 고혈압 관리를 위한 의사결정지원시스템 및 임상진료지침의 개발에 활용될 수 있는 데이터마이

닝의 활용 방안을 제시할 수 있었다. 고혈압 관리를 위해서 많은 나라에서 진료지침을 개발하여 사용하고 있으며, 이 진료지침의 수행 실태에 관한 연구가 활발히 이루어지고 있다. 고혈압의 경우 그동안 발병요인, 단일 약제의 효능·효과에 관한 연구 등은 지속적으로 이루어져 왔으나, 환자의 특성, 진단, 치료방법 및 치료결과간의 상관관계 및 패턴을 동시에 파악하여 이를 치료전략으로 활용하지는 못하였다. 의료분야에서의 지식베이스를 활용한 의사결정지원시스템의 확산은 학계와 임상사와의 사이를 좁혀 줄 수 있다. 또한 지식베이스의 수가 양적으로 증가하면서 일관성과 정확성을 보다 확실하게 하기 위한 과정이 더욱 필요하게 되었다. 지금까지 많은 지식베이스가 소수 전문가들에 의해 구축되었으며, 지식베이스 내용들에 대한 체계적인 타당성 검증을 위한 기준도 아직 마련되지 않고 있다. 또한 데이터의 수집 및 관리를 위한 신뢰성 있는 방법의 개발이 의료분야의 지식베이스 구축 및 이의 임상적 실용화를 위해 필요하다. 궁극적으로 의료분야에서도 다른 분야와 마찬가지로 의료정보자원의 효율적 활용을 위한 데이터마이닝과 같은 새로운 전략적 기법이 요구된다고 할 수 있다. 향후에는 본 연구에서 개발한 시스템을 병원환경에 적용시킬 수 있는 추론 엔진으로 재구성하여 병원의 운영계 시스템과 연계시켜 운영하여 실제 진료분야에서의 지식경영시스템의 가치를 재평가하고, 운영 데이터의 축적·활용을 통한 지식의 검증 및 보완이 이루어져야 할 것이며, 운영 결과를 분석하여 시스템의 유용성을 평가함으로써 임상에서의 지식기반의 의사결정지원시스템 응용의 활성화를 도모해야 할 것이다. 또한 본 연구의 결과는 우리나라 현실에 부합되는 고혈압 진료지침을 개발하고 적용, 평가하는데 기여할 수 있을 것으로 판단되며, 이를 의사결정지원시스템의 개발을 통해 실제 임상진료에 적용해 봄으로써 그 효과와 실증적 가치를 창출할 수 있을 것이다. 급변하는 의료 환경에 따

라 국내 의료계에서는 지역적으로 산재되어 있는 의료 자원을 정비하고 이를 효율적으로 활용하여 보다 효과적인 의료 서비스를 제공하기 위한 대책을 마련하여야 하는 과제를 안고 있다. 이와 같은 과제를 해결하기 위한 방법으로써 의사결정지원시스템 등 지식경영시스템의 의료체계로의 도입이 요구되며 그 방법론 및 실효성에 대한 연구가 요구되고 있다. 환자에게 양질의 의료서비스를 제공할 수 있게 되는 등 의료진료의 효율성을 제고시키며, 지역간 의료서비스 수준의 격차를 줄일 수 있는 유용한 수단이라고 할 수 있다. 고혈압은 선진국, 후진국의 구별없이 유병율이 약 10-20%로 가장 흔한 질환일 뿐 아니라 뇌혈관계, 심장 그리고 신장에까지 합병증을 초래하는 만성퇴행성 질환으로 전세계적으로 중요한 보건학적 관심사이다. 이에 따라 데이터웨어하우스 기반의 데이터마이닝 프로세스를 통해 환자의 특성, 진단, 치료방법, 치료결과간의 상관관계와 패턴을 발견하고 이를 통해 치료 결과를 예측할 수 있는 규칙을 도출하여 이를 고혈압 관리를 위한 의사결정지원시스템의 개발에 활용하였다. 의료분야의 정보가 양적으로 증가하면서 일관성과 정확성을 보다 확실하게 하기 위한 과정이 더욱 필요하다. 또한 데이터의 수집 및 관리를 위한 신뢰성 있는 방법의 개발이 의료분야의 지식베이스 구축 및 이의 임상적 실용화를 위해 필요하다. 향후연구관제로는 진료지침을 개발하고 실제 진료활동에 적용해 봄으로써 그 효과를 평가하는 중재적 연구가 필요하고, 고혈압의 특성을 대표할 수 있는 대상자의 선정과 이들의 최적의 특성을 선별하여 데이터마트를 구축하여야 할 것이며, 또한 이에 맞는 최적의 지식습득 모형을 개발해야 할 것이다. 뿐만 아니라 신뢰성 있는 시스템을 개발하기 위해서는, 특성의 선별과 분석결과 활용 등 시스템의 개발 단계에 있어서 충분한 임상적 평가를 통해 임상적 실용화의 가치를 높여나가야 할 것이다. 의사결정지원시스템이 임상전문의의 의사결

정 지원뿐 아니라 지역사회 일선 의료인들이 교육 및 일차진료에 활용된다면, 교육효과의 상승과 함께 지역적 시간적 한계를 받지 않는 유용성을 얻을 수 있으며, 전문의료인에 대한 수요증가 추세를 둔화시킬 수 있어 의료비 절감과 의료혜택 균등의 기회를 만들 수 있는 계기가 될 것이다. 본 연구의 결과에서 데이터마이닝 모형의 예측력과 이에 대한 타당성 검증 결과가 아직 임상에 적용되기에는 미흡한 부분이 있지만, 임상이론과 전문의, 그리고 이에 의해 축적된 정보인 환자 데이터를 이용한 데이터마이닝 기법을 효율적으로 활용한 시스템을 구축하여 그 타당성을 점차적으로 높여감으로써 실용화 단계에 이를 수 있도록 하여야 할 것이다.

참 고 문 헌

- [1] Bonner G, "Diagnosis of Hypertension", *Acta Medica Austriaca*, 22(3), pp.32-34, 1995
- [2] Buchanan BG and Smith RG, "Fundamentals of expert system" *Ann R Comp Sci*, 3, pp.23-58, 1998
- [3] Burt EJ, "Prevalence of hypertension in the US adult population", *Hypertension*, 25(3), pp.305-13, 1995
- [4] Chae YM, Ho SH, Cho KW, Lee DH, Ji SH. Data mining approach to policy analysis in a health insurance domain. *International Journal of Medical Informatics*. 2001; 62:103-111
- [5] Fayyad U, Piatetsky-Shapiro G and Smyth P, "Advances in Knowledge Discovery and Data Mining", Menlo Park, Calif, AAAI Press, 1996
- [6] Giuse DA and Miller NB, "Strategies for medical knowledge acquisition", *Medical informatics*(Bemmel JH and Musen MA), Houten, Springer, pp.277-292, 1997
- [7] Guideline subcommittee of the World Health Organization-International society of hypertension, "1999 World Health Organization-International society of hypertension guidelines for the management of hypertension", *Journal of hypertension*, 17, pp.151-183, 1999
- [8] Hart, A., "Knowledge Acquisition for expert systems(2nd Ed.)", New York:Macgraw-Hill, 1992
- [9] Joint National Committee on Detection, Evaluation, and Treatment of High Blood Pressure, "The sixth report of the Joint National Committee on Prevention, Detection, and Treatment of High Blood Pressure (JNC VI)",
- [10] Langley P and Simon HA, "Application of machine learning and rule induction", *Communication of the acm*, 38(11), pp.55-64, 1995
- [11] Long WJ, Griffith JL, Selker HP, D'Agostino RB. A comparison of logistic regression to decision-tree induction in a medical domain. *Computers and Biomedical Research*, 1993; 26:74-97
- [12] Marques P and Tuomilehto J, "Hypertension awareness, treatment and control in the community: is the 'rule of halves' still valid?", *J Hum Hyperten*, 11, pp.213-220, 1997
- [13] Robert J, "Clinical Decision support system. Electronic health records :chaining the vision(Gretchen FM, Mary AH, Kathleen AW)", Philadelphia, WB Saunders, pp.305-315, 1999
- [14] Swales JD, "Management guidelines for hypertension", *Journal of Human Hypertension*, 9(2), pp.9-13. 1995
- [15] Zopounidis C, Doumpos M and Matsatsinis NF, "On the use of knowledge based DSS in financial management", *A survey, Decision support system*, 20(2), pp.259-277, 1997