

대한고혈압학회지 : 제 9권 제 2호

□ 원 저 □

고혈압 관리를 위한 의사지원결정시스템의 데이터 마이닝 접근

연세대학교 의과대학 심장혈관병원 심장내과¹, 예방의학교실²강석민¹·조승연¹·채영문²·김태수²·호승희²

=Abstract=

Data Mining Approach to Clinical Decision Support System for Hypertension Management

Seok-Min Kang, M.D.¹, Seung-Yun Cho, M.D.¹,Young Moon Chae, Ph.D.², Tae Soo Kim, Ph.D.², and Seung Heu Ho, M.D.²*Cardiology Division¹, Yonsei University College of Medicine**Graduate School of Health Science and Management², Yonsei University, Seoul, Korea*

Background: This study examined the utility of data mining algorithms for the management of hypertension.

Methods: We studied 2,446 hospitalized patients with hypertension and 3,835 clinic patients with hypertension. Among data mining algorithms, we used clustering analysis and compared decision tree analysis with logistic regression.

Results: 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.

Conclusions: This study suggests that data mining algorithms may be an useful method for hypertension management and CDSS 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

서 론

고혈압은 복잡한 현대사회를 사는 성인들에게 호발하는 만성 퇴행성질환으로 구미시의 경우 전 인구의 15-20%가 고혈압 환자인 것으로 알려져 있으며, 국내에서는 성인 인구의 20% 내외가 고혈압을 갖고 있는 것으로 추정되고 있다.¹⁻³⁾ 고혈압의 치료 및 관리에 대한 유의성은 여러 연구 결과에 의해 입증되었으나, 지금까지 지속적으로 이루어진 고혈압의 인지도, 치료 및 관리 현황과 관련된 행태조사를 살펴보면 우리나라를 포함하여 많은 나라에서 발표된 것 같이 고혈압 환자의 70% 이상이 불완전한 조절 상태, 또는 전혀 치료를 받고 있지 않은 상태에 머물러 있고, 고혈압이 조절되고 있는 비율은 증가되지 않고 어떤 경우에 있어서는 오히려 낮아지고 있다는 것은 염려되는 점이다. 지금까지의 고혈압 치료에 대한 연구는 무작위 임상시험이 대부분 이었는데, 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 임상에 적용할 수 있는 복합적 정보를 제공할 수 있는 지침이 필요하다.

진료를 위한 의사 결정지원시스템(clinical decision support system)은 임상 의사가 환자를 진료를 행하는 과정 중에 이루어지는 의사결정에 지식을 제공함으로써 이를 지원할 수 있는 시스템을 말한다.⁴⁾ 최근 몇 년에 걸쳐 대기업을 중심으로 대량의 데이터를 포함하고 있는 정보시스템을 구축, 관리할 수 있는 데이터 웨어하우스(data warehouse)가 새로운 아키텍처로 각광받고 있으며, 이와 함께 방대한 양의 데이터로부터 함축적이며 잠재적 유용성이 있는 정보를 발견할 수 있는 지식 탐사 방법인 데이터 마이닝(data mining)에

대한 중요도가 더욱 증대되고 있다. 데이터 마이닝은 패턴인식 기술이나 통계기법, 수학적 알고리즘을 이용하여 의미 있는 새로운 상관관계, 패턴, 추세 등을 발견하는 과정으로,⁵⁾ 의료분야에서도 다른 분야와 마찬가지로 의료정보자원의 효율적 활용을 위한 지식 기반 경영과 이의 요소정보기술로써 데이터 웨어하우스와 데이터 마이닝의 중요성이 부각되고 있다.

이에 본 연구에서는 의료분야의 의사결정지원을 위한 지식경영체제 도입을 촉진시키기 위한 방안으로써 데이터 마이닝 기법을 적용하여 질병 패턴 분석의 일환으로 임상적으로 중요한 위치를 차지하고 있는 고혈압 환자의 특성과 이에 따른 예후를 예측할 수 있는 지식을 발굴하고자 하였다.

대상 및 방법

1. 연구 대상

1999년 10월부터 2003년 1월까지 신촌 세브란스 병원에 입원한 환자 중에서 주 진단명이나 부 진단명 중 고혈압이 있는 입원 환자 2,446명과 심장내과에 내원한 환자 중 고혈압이 의심되거나 또는 고혈압을 지니고 있어 설문조사에 응한 외래 환자 3,835명을 연구 대상으로 하였다.

고혈압 환자의 진단과 치료에 있어서 중요한 특성으로 파악되어 지는 항목들을 선별하는 과정은 임상 진료지침 등의 임상이론, 순환기 내과 전문의의 견해, 실제 환자 데이터에 대한 기술 통계량 측정을 통한 데이터의 특성 파악 등을 통해 이루어졌다. 입원환자와 외래환자 두 집단은 동반질환 유무, 혈압의 중등도 및 합병증 유무 등의 특성이 다르기 때문에 같은 환자 군으로 병합하여 분석을 하기보다는 각각 개별적인 집단으로 보고 연구대상에 대하여 각 항목과 관련된 기본 통계량을 조사하였다.

2. 분석 방법

본 연구에서는 데이터 마이닝 프로세스 중 군집분

석(clustering analysis), 의사결정 나무분석(decision tree analysis) 및 로지스틱 회귀분석을 이용하였다. 데이터 마이닝은 대량의 데이터 사이에 묻혀 있는 패턴을 발견하고 규칙을 추론함으로써 의사결정을 지원하고 그 효과를 예측하기 위한 기법으로 지금까지 대답이 어려웠던 많은 질문에 대한 답을 제공하는 방법이다. 군집분석은 주어진 자료들 중에서 유효한 것들을 몇몇의 집단으로 그룹화하여, 각 집단의 성격을 파악함으로써 데이터 전체의 구조에 대한 이해를 돕고자 하는 방법이다. 대용량 데이터에서 각각의 관찰치를 요약 하는 것보다는 전체를 유사한 관찰치들의 군집으로 구분하여 각 군집들의 특성을 파악함으로써 전체 데이터에 대한 의미 있는 정보를 얻을 수 있다. 전통적으로 이분형인 종속변수와 여러 가지 독립변수와의 관계를 알아보고자 하는 방법으로서 많이 사용되어지는 것이 로지스틱 회귀분석이다. 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 의사결정 나무분석은 의사결정규칙을 나무구조로 도표화하여 분류와 예측을 수행하는 분석방법으로 예측의 과정이 나무구조에 의한 추론규칙에 의해 표현된다. 데이터 마이닝에서의 의사결정나무는 탐색과 모형화의 특성을 지

니며, 사전에 이상치(outlier)를 검색하거나 분석에 필요한 변수를 찾아내고 분석모형에 포함되어야 할 교호 효과를 찾아내는 데 사용될 수 있고, 그 자체가 분류 또는 예측 모형으로 사용될 수도 있다. 1980년대 이후 CHAID (Chi-squared Automatic Interaction Detection), 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.

3. 고혈압 관리 시스템 개발 방법

의사결정지원시스템은 여러 개의 원시 데이터 베이스들과 데이터 웨어하우스, 그리고 의사, 간호사, 환자의 3개의 모듈로 구성되어 개발하였다(Fig. 1). 데이터 웨어하우스는 데이터 마이닝에 필요한 통합적 데이터의 수집과 전처리 과정을 거치기 수행하기 때문에 데이터 웨어하우스 기반에서 데이터 마이닝을 수행하는 것이 바람직하다. 대부분의 데이터 마이닝의 시간이 이러한 데이터의 준비 과정에서 소모되기 때문이다.

결 과

1. 대상자 특성

각 집단의 증상, 가족병력 등에 따른 특성을 살펴보

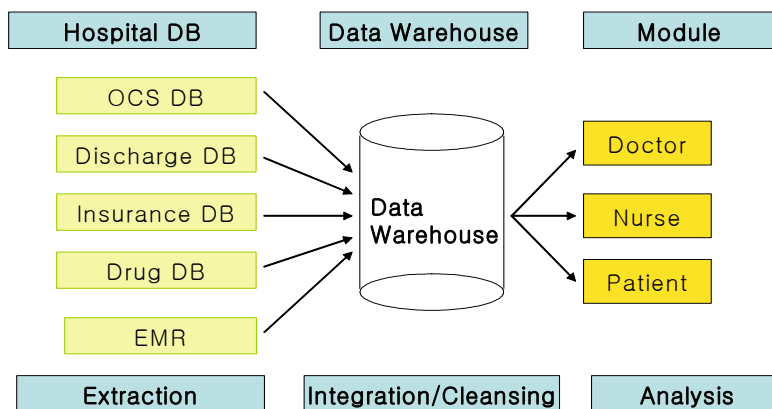


Fig. 1. Structure of hypertension management system.

면, 입원환자 성별 구성은 남성이 54.3%, 여성이 45.7%이고, 외래환자 성별 구성은 남성이 50.8%, 여성이 49.2%이었다. 입원환자 응답자의 연령은 50세 이하가 38.2%, 50세 이상이 61.8%이고 외래환자 응답자의 연령은 50세 이하가 23.5%, 50대 이상이 76.5%이었다(Table 1).

입원환자 주 호소 증상은 흉통이 30.1%로 가장 많았고 다음으로는 호흡곤란이 15.4%, 현기증이 8.0%, 두통이 7.2% 순이었다. 반면에 외래환자 주 호소 증상은 흉통이 29.7%로 가장 많았고 다음으로는 두통이 27.2%, 현기증이 23.9%, 심계항진이 18.7%이었다. 고혈압 환자의 임상평가에서 고려해야 하는 가족병력

Table 1. Clinical Characteristics of Subjects

Variables		In-Hospital (n=2446)		Clinic (n=3835)	
		n	%	n	%
Sex	Male	1328	54.3	1948	50.8
	Female	1118	45.7	1887	49.2
Age	age <= 50	935	38.2	901	23.5
	age > 50	1511	61.8	2935	76.5
Chief Complaints	Headache	141	7.2	361	27.2
	Syncope	25	1.3	25	1.9
	Chest Pain	593	30.1	394	29.7
	Palpitation	36	1.8	248	18.7
	Dizziness	157	8.0	317	23.9
	Edema	91	4.6	156	11.7
	Numbness	5	0.3	35	2.6
	Dyspnea	303	15.4	240	18.1
	Visual disturbance	28	1.4	181	13.6
Family History	Hypertension	469	26.1	1320	42.7
	Diabetes	235	13.1	289	9.5
	Hyperlipidemia	2	0.1	20	0.7
	Stroke	174	9.7	218	7.2
	Renal disease	10	0.6	11	0.4
	Sudden death	7	0.4	5	0.2
Associated Diseases	Cerebrovascular disease	240	9.8	172	4.7
	Cardiovascular disease	579	23.7	1106	30.1
	Renal disease	374	15.3	215	5.9
	Vascular disease	129	5.3	109	3.0
	Hypertensive retinopathy	48	2.0	30	0.8
	Diabetes	886	36.0	777	21.2
	Hyperlipidemia	31	1.3	123	3.4
	Gout	94	3.8	37	1.0
	Bronchial asthma	7	0.3	29	0.8
Sexual dysfunction	0	0	4	0.1	
Smoking	Yes	450	49.4	803	27.8
Alcohol	Yes	744	25.5	947	32.7

을 조사한 결과 입원환자 가족력에 고혈압이 26.1%, 당뇨병이 13.1% 순으로 나타났고 외래 환자 가족력에 고혈압이 42.7%, 당뇨병이 9.5%로 나타났다. 입원한 고혈압 환자의 동반질환에 있어서 당뇨병이 36.0%로 가장 많았다. 반면 외래 환자의 동반질환은 심장질환이 30.1%로 가장 많았다. 흡연 여부에 있어서는 입원 환자 49.4%, 외래 환자 27.8%가 흡연력이 있었다. 또한 음주 여부에 있어서는 입원환자 25.5%, 외래 환자 32.7%가 음주력이 있었다.

흉부 엑스선 촬영 검사상에서 심비대가 입원 환자의 73.3%, 외래 환자의 24.8%가 관찰되었다. 심전도 검사상 허혈성 심질환 소견을 보였던 환자는 입원 환자의 39.9%, 외래 환자의 24.7% 이었다. 심초음파 검사상 심비대가 동반된 경우는 입원 환자의 67.1%, 외래 환자의 15.3% 이었다(Table 2).

공복시 평균 혈당 수치는 입원 환자가 131.2 ± 73.2

mg/dl, 외래 환자가 118.8 ± 39.9 mg/dl 이었다. 그 외의 대부분의 임상검사 수치가 입원 환자와 외래 환자 간에 통계적으로 유의한 차이가 없었다(Table 3).

2. 군집분석에 의한 혈압변이 요인분석

입원환자와 외래환자의 연구대상에 대하여 각 항목과 관련된 기본 통계량을 조사한 바, 입원환자군과 외래환자군의 두 연구대상집단에 대하여 혈압변이의 변수를 목표변수로 하여 군집수를 세 군집으로 나누어서 군집분석을 하였다. 각 군집의 관측 값을 군집의 중심값 중에서 가장 가까운 군집 안에 있는 관측 값간의 평균 거리와 군집간의 평균 거리를 비교 함으로써 평가하였다. 그 결과 군집 1의 혈압변이는 18 mmHg, 군집 2의 혈압변이는 55 mmHg 만큼 혈압이 하강한 군집이고, 군집 3의 혈압변이는 8 mmHg 만큼 혈압이 상승한 군집이었다. 이 3개 군집의 비유사성의 척도로

Table 2. Chest X-ray, ECG and Echocardiographic Findings

Variables	In-Hospital (n=2446)		Clinic (n=3835)	
	n	%	n	%
LVH on Chest PA	641	73.3	437	24.8
Ischemia on EKG	349	39.9	527	24.7
LVH on Echocardiography	587	67.1	252	15.3

Table 3. Laboratory Findings

Variables	In-Hospital (n=2446)		Clinic (n=3835)		P
	Mean	SD	Mean	SD	
FBS (mg/dl)	131.2	73.2	118.8	39.9	0.03
BUN (mg/dl)	26.5	20.4	18.3	10.9	NS
Creatinine (mg/dl)	2.9	3.4	3.6	61.7	NS
Uric acid (mg/dl)	6.4	2.4	5.6	1.9	NS
K ⁺ (mmol/dl)	4.4	5.7	4.2	0.5	NS
T-cholesterol (mg/dl)	176.7	43.5	193.0	38.7	NS
LDL-Cholesterol (mg/dl)	106.3	34.0	117.5	31.4	NS
HDL-Cholesterol (mg/dl)	46.8	17.0	47.8	13.5	NS

서 거리를 이용하여 관찰치 간의 서로 위치하고 있는 거리가 가까운 군집 1과 3을 혈압 비조절군으로 하고 군집 2를 혈압조절군으로 분류하여 특성을 살펴보면 다음과 같다. 입원환자의 혈압변이에 따른 군집분류 결과, 혈압 비조절군의 혈압변이는 5.6±12.0 mmHg, 혈압조절군의 혈압변이는 41.3±14.2 mmHg이었다. 혈압 비조절군 중 두통을 호소한 환자는 45명 (6.2%) 였으나, 혈압 조절군은 49명(10.6%)로 약 4 %정도 더 높았다(Table 4). 외래환자의 혈압변이에 따른 군집 분류 결과, 혈압 비조절군의 혈압변이는 2.9±10.4 mmHg, 혈압 조절군의 혈압 변이는 41.3±14.2 mmHg 이었다(Table 5). Table 4 와 Table 5 에서는 두 집단 간의 변수 중에서 통계적으로 유의한 차이 (p<0.05)를 보이고 있는 변수들만을 나열 하였으며, 이들의 변수

는 다음에 진행 할 의사결정나무분석과 로지스틱 회귀분석에서 주로 사용될 독립변수들을 의미한다.

3. 의사결정나무와 로지스틱 회귀분석에 의한 혈압 변이 요인 분석

위 군집분류의 결과를 기초로 하여, 의사결정나무 분석과 로지스틱 회귀분석의 비교를 한 바 두 연구 집단에 대하여 의사결정나무분석이 로지스틱 회귀분석보다 모형에 대한 오차(root adjusted standard error)도 적을 뿐 아니라 오분류율(misclassification rate)도 작음을 알 수 있었다(Table 6).

의사결정나무분석의 한 예를 살펴보면, 외래환자 3,835명 중 분석 가능한 환자는 1458명 이었다. 이 중 혈압 조절군은 381명으로 전체 1458명의 26.1%,

Table 4. Clustering Analysis of Hospitalized Patients with Hypertension (*p<0.05)

Variables	Non-Responder (5.6±12.0 mmHg)			Responder (41.3±14.2 mmHg)			t-value/ x ² -value*
	Mean	N	%	Mean	N	%	
Initial SBP (mmHg)	146.2±13.4			172.6±20.1			30.2
Headache		45/728	6.2		49/464	10.6	6.9*
Family Diabetes		97/672	14.4		36/387	9.3	5.4*
Hyperlipidemia		15/953	1.6		4/525	0.8	1.2*
Drinking		161/526	30.6		94/262	35.9	0.6*
Smoking		275/606	45.4		138/285	48.4	2.0*

Table 5. Clustering Analysis of Clinic Patients with Hypertension (*p<0.05)

Variables	Non-Responder (2.9±10.4 mmHg)			Responder (33.8±12.9 mmHg)			t-value/ x ² -value*
	Mean	N	%	Mean	N	%	
Initial SBP (mmHg)	144.7±10.1			169.2±16.7			34.9
FBS (mg/dl)	118.3±48.8			106.9±29.1			-3.5
Diabetes		94/512	18.7		222/956	23.2	3.7*
Symptomatic arterial disease		3/5	60		1/18	5.6	4.7*

Table 6. Comparison between Logistic Regression Analysis and Decision Tree Analysis

Patient Type	Model	Root ASE	Misclassification Rate
In-Hospital	Logistic Reg.	0.3957	0.1942
	Decision Tree	0.3757	0.1867
Clinic	Logistic Reg.	0.4398	0.2613
	Decision Tree	0.3916	0.2222

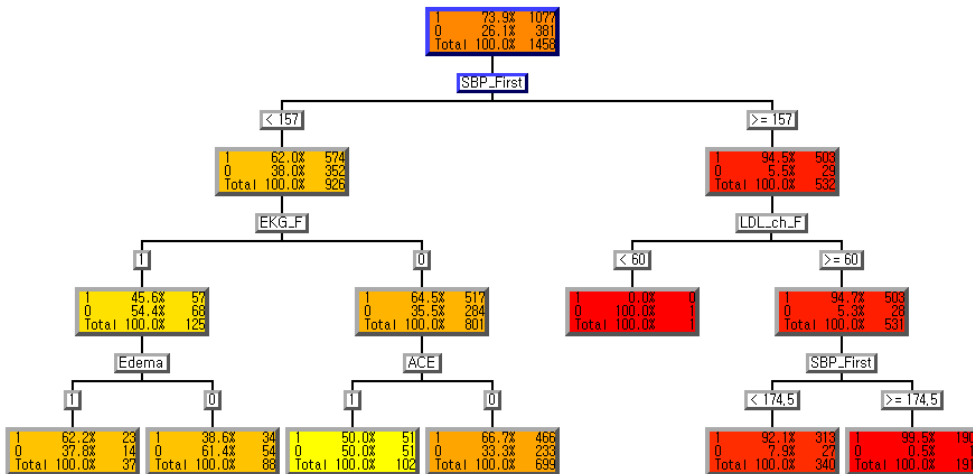


Fig. 2. Clinical decision support system for hypertensive patients at out-patient clinic (0-controlled hypertensive patients, 1-uncontrolled hypertensive patients).

비 조절군은 1077명으로 전체 1458명의 73.9%로 나타났다. 초기 수축기압력이 첫번째 분리변수로 초기 수축기압력에 따라 혈압변이의 범주를 나눌 수 있다. 즉, 혈압변이에 가장 영향을 미치는 변수는 외래 환자의 경우에 초기 수축기 혈압이 157 mmHg이하인 환자의 혈압이 조절될 확률은 38.0% (352명)으로 나타났다. 또한 심전도 검사로 허혈성 증세가 없는 환자이면서 안지오텐신 전환효소 억제제를 복용한 환자군의 혈압 조절 확률은 50.0%이었다(Fig. 2).

4. 고혈압 관리 시스템

저자들은 위에 언급된 고혈압 환자의 진단과 치료에 있어서 중요한 항목들을 기초로 하여 임상지식 및 순환기 내과 전문의의 지식 등 관련 정보를 체계화된

지식으로 표현함으로써 의사결정지원시스템의 지식 베이스의 생성 규칙을 이끌어냈다. 의사모듈은 WHO와 JNC (Joint National Committee on prevention, detection, evaluation, and treatment of high blood pressure) VII에 근거한 임상처방을 기초로 한 진단 정보를 제공한다.⁶⁻⁷⁾ 의사모듈에서는 각 증상에 대하여 우선순위에 기초하여 약제처방을 제안해 줄 수 있다. 이곳에서는 해당하는 약제에 대한 성상, 용법, 그리고 부작용 등을 알 수도 있다(Fig. 3). 또한 의사모듈에서는 데이터베이스를 검색하여 여러 증상에 해당하는 환자들을 검색하여 약제처방의 결과 발생되는 혈압의 변이에 대한 통계량 및 개개 환자의 자세한 특성을 검색할 수 있도록 구성하였다(Fig. 4).

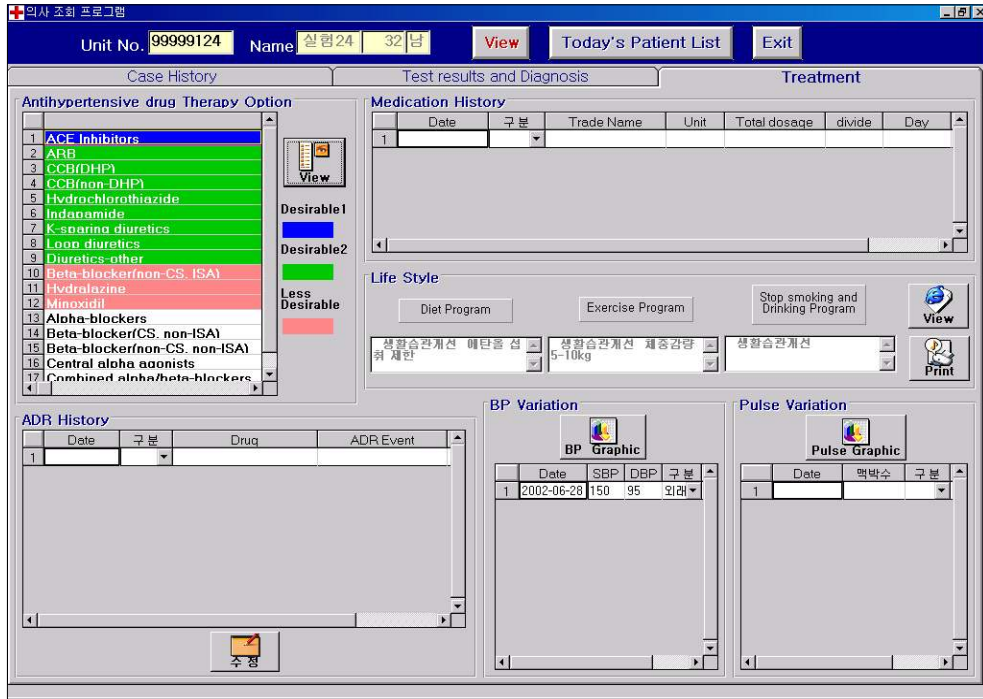


Fig. 3. Drug prescription module.

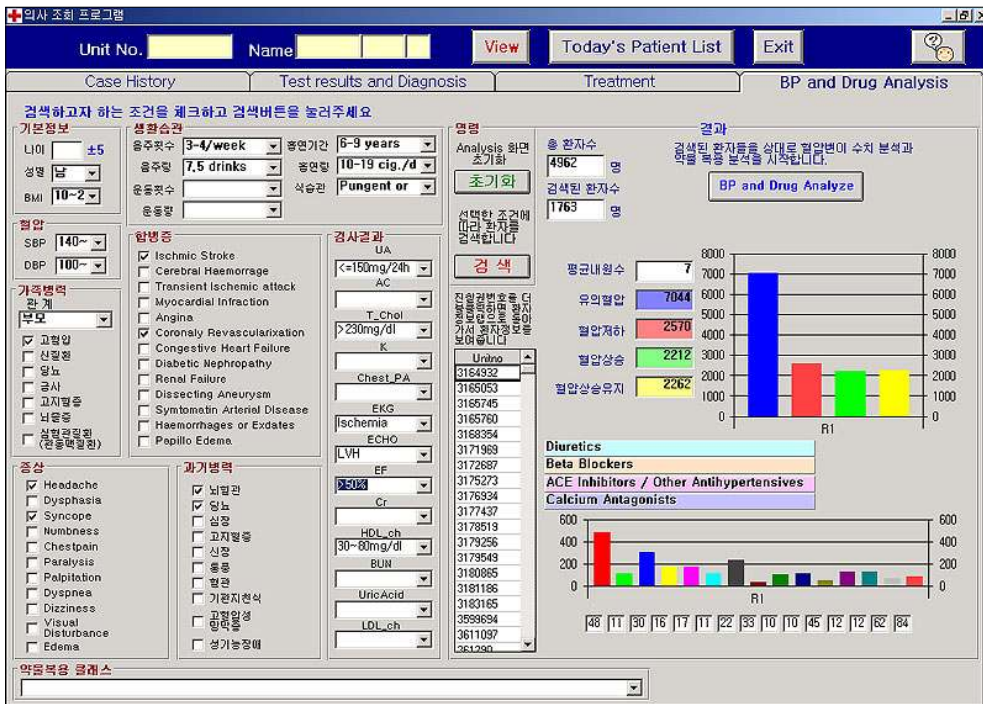


Fig. 4. Patient search module.

고찰

고혈압은 유전인자와 환경인자에 따라 영향을 많이 받는 질환이며, 그 자체로도 중요한 관리의 대상이 되는 하나 특히 심장질환이나 뇌혈관질환, 그리고 신장질환 같은 주요 사망원인 질환 발생과 깊은 관련이 있으므로 더욱 중요하다.⁸⁾ 고혈압의 병리 및 생리학적 변화나 그 치료방법의 엄청난 발전에도 불구하고 고혈압은 아직도 완전한 예방과 치료가 어려운 상태이다. 따라서 고혈압 관리에서 위험요인을 찾고 이를 해결하기 위한 적절한 치료방법을 찾는 것은 무엇보다 중요하다고 할 수 있다.

이에 본 연구에서는 데이터 마이닝 기법을 적용하여 치료결과에 어떠한 요인이 얼마나 중요한지를 유추해 내고 치료 후 혈압 조절군과 비조절군의 특성을 분류, 규명함으로써 새로운 치료 전략을 도출해 보고자 하였으며, 이를 이용한 의사결정지원시스템을 개발하고 이의 유용성을 평가하고자 하였다. 또한 고혈압 환자를 대상으로 로지스틱 회귀분석과 의사결정나무분석 기법을 이용하여 혈압조절상의 위험요인을 규명하고, 치료약제별로 혈압 조절군과 비조절군을 결정짓는 규칙을 도출하여 실제 임상이론 및 환자 사례와의 비교검증을 통한 예측력을 평가하였다.

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 의사결정지원시스템에 대한 연구가 도입되었다. 현재까지 의사결정 지원시스템을 개발하기 위하여 많이 활용된 인공지능기법은 규칙 기반 추론기법이다. 규칙 기반 추론기법에서는 전문가의 경험적인 휴리스틱 지식을 생성규칙의 형태로 표현하는데, 이는 전문가의 지식이 체계적으로 잘 정리되어 있는 경우에는 별 문제가 없지만 그렇지 못한 경우에는 그만큼 영역 전문가로부터의 지식 획득에 의존해야 하므로 이로 인한 여러

한계점이 제기되고 있다. 필요한 지식을 임상에 적용시키기 위하여서는 문제 해결을 위한 의사결정 과정과 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 지식 베이스화하는 것은 매우 힘든 과정이라 할 수 있다.⁹⁾ 이와 같이 의사결정지원시스템을 개발하는 데 있어서 지식 베이스의 구축은 가장 어려운 분야중의 하나라고 할 수 있다. 과거부터 현재까지 의료분야의 의사 결정지원시스템이 질병의 진단과 치료 및 예후 예측에 있어 완벽한 의사결정지원 기능을 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,¹⁰⁻¹¹⁾ 대표적인 연구가 임상자료로부터 직접 지식을 추출해 내는 방법이다. 정보시스템의 급격한 발전과 더불어 많은 양의 데이터가 축적되기 시작하면서, 필요한 정보를 찾아 내어 가치 있는 지식으로 승화시키는 것에 대한 필요성이 야기되고 있다. 최근 몇 년에 걸쳐 대기업을 중심으로 대량의 데이터를 포함하고 있는 정보시스템을 구축, 관리할 수 있는 데이터 웨어하우스가 새로운 아키텍처로 각광받고 있으며, 이와 함께 방대한 양의 데이터로부터 함축적이며 잠재적 유용성이 있는 정보를 발견할 수 있는 지식 탐사 방법인 데이터 마이닝에 대한 중요도가 더욱 증대되고 있다.

이 연구에서는 입원과 외래 환자 모두에서 의사결정나무가 로지스틱 회귀분석보다 오분류율이 낮아서 혈압변이를 예측하는데 더 정확함을 알 수 있었다. 이는 Long 등¹²⁾의 연구와 다른 결과를 나타내는 것으로 종전의 역학 연구에서 가장 널리 사용하던 로지스틱 회귀분석의 대안으로 의사결정나무분석이 유용하게 사용될 수 있음을 알 수 있었다.

또한 군집분석 결과 입원환자의 혈압조절 평균치는 41.3 mmHg으로 외래환자의 혈압조절 평균치 33.8 mmHg 보다 7.5 mmHg가 더 높았다. 이는 외래환자

의 평균 초기혈압(169.2 mmHg)보다 높은 입원환자(172.6 mmHg)가 병원에서 집중적으로 약물치료를 받아서 혈압이 효과적으로 낮아지게 된 데 기인하였다고 생각해 볼 수도 있으나, 두 군 간의 혈압 측정 시기 및 측정 방법의 차이가 있으므로 이는 결과를 해석하는데 문제점이 있으리라 생각된다. 그리고 본 연구에서는 임상이론 및 데이터 마이닝 기법을 통해 도출된 지식을 기반으로 하는 지식베이스를 구축하고 추론엔진을 구성하여 의사결정지원시스템을 개발하고 임상이론과 데이터 마이닝 기법의 단독 및 혼합에 의한 처방규칙에 대한 치료결과를 분석함으로써 데이터 마이닝 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 것보다는 방대한 데이터로 이루어진 데이터 웨어하우스로부터의 지식 획득을 통하여 이루어져야 함을 제시한다. 그리고 본 연구에서 고혈압의 특성을 대표할 수 있는 대상자의 선정과 이들의 최적적 특성을 선별하고 이에 대한 충분한 임상적 평가가 이루어졌는지에 대한 충분한 검토가 이루어져야 함을 나타낸다.

향후에는 본 연구에서 개발한 시스템을 병원환경에 적용시킬 수 있는 추론 엔진으로 재구성하여 병원의 처방전달 시스템과 연계시켜 운영하여 실제 진료분야에서의 지식경영시스템의 가치를 재평가하고, 운영데이터의 축적 및 활용을 통한 지식의 검증 및 보완이 이루어져야 할 것이며, 운영 결과를 분석하여 시스템의 유용성을 평가함으로써 임상에서의 지식 기반의 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 임상 진료에 적용해 봄으로

써 그 효과와 실증적 가치를 창출할 수 있을 것이다. 급변하는 의료 환경에 따라 국내 의료계에서는 지역적으로 산재되어 있는 의료 자원을 정비하고 이를 효율적으로 활용하여 보다 효과적인 의료 서비스를 제공하기 위한 대응책을 마련하여야 하는 과제를 안고 있다. 이와 같은 과제를 해결하기 위한 방법으로써 의사결정지원시스템 등 지식경영시스템의 의료체계에의 도입이 요구되며 그 방법론 및 실효성에 대한 연구가 요구되고 있다. 환자에게 양질의 의료서비스를 제공할 수 있게 되는 등 의료진료의 효율성을 제고시키며, 지역 간 의료서비스 수준의 격차를 줄일 수 있는 유용한 수단이라고 할 수 있다.

본 연구의 결과에서 데이터 마이닝 모형의 예측력과 이에 대한 타당성 검증 결과가 아직 임상에 적용되기에는 미흡한 부분이 있지만, 임상이론과 전문의, 그리고 이에 의해 축적된 정보인 환자 데이터를 이용한 데이터 마이닝 기법을 효율적으로 활용한 시스템을 구축하여 그 타당성을 점차적으로 높여감으로써 실용화 단계에 이를 수 있도록 하여야 할 것이다.

요 약

연구배경 : 의료분야의 의사결정지원을 위한 지식경영체제 도입을 촉진시키기 위한 방안으로써 데이터 마이닝 기법을 적용하여 질병 패턴 분석의 일환으로 임상적으로 중요한 위치를 차지하고 있는 고혈압 환자의 특성과 이에 따른 예후를 예측할 수 있는 지식을 발굴하고자 하였다.

방법 : 1999년 10월부터 2003년 1월까지 신촌 세브란스 병원에 입원한 고혈압 환자 2,446명과 외래를 내원한 고혈압 환자 3,835명을 연구 대상으로 하였다. 데이터 마이닝 프로세스 중 군집분석, 의사결정 나무 분석 및 로지스틱 회귀분석을 이용하였다.

결과 : 입원환자의 혈압변이에 따른 군집분류 결과, 혈압 비조절군의 혈압변이는 5.6 ± 12.0 mmHg, 혈압 조절군의 혈압변이는 41.3 ± 14.2 mmHg이었다. 외

래환자의 혈압변이에 따른 군집분류 결과, 혈압 비조절군의 혈압변이는 2.9 ± 10.4 mmHg, 혈압 조절군의 혈압 변이는 41.3 ± 14.2 mmHg이었다. 위 군집분류의 결과를 기초로 하여, 의사결정나무분석과 로지스틱 회귀분석의 비교를 한 바 두 연구 집단에 대하여 의사결정나무분석이 로지스틱 회귀분석보다 모형에 대한 오차도 적을 뿐 아니라 오분류율도 작음을 알 수 있었다.

결론 : 의사결정지원시스템이 임상전문의의 의사결정 지원뿐 아니라 지역사회 일선 의료인들이 교육 및 일차진료에 활용된다면, 교육 효과의 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.

참고문헌

- 오병희, 김창엽, 이진세, 강영호, 이영조, 강위창. 우리나라 농어촌 지역 성인의 고혈압 유병률, 대한내과학회지 1999;56(3):299-316.
- 허갑범. 내분비 및 대사 이상과 고혈압, 대한의학협회지 1992;35(2):186-192.
- Marques P, Tuomilehto J. Hypertension awareness, treatment and control in the community: is the 'rule of halves' still valid ?, J Hum Hyperten 1997;11:213-220.
- Robert J. Clinical Decision support system. Electronic health records: chainging the vision (Gretchen FM, Mary AH, Kathleen AW), Philadelphia, WB Saunders, 1999:305-315.
- Fayyad U, Piatetsky-Shapiro G, Smith P. Advances in Knowledge Discovery and Data Mining, Menlo Park, Calif, AAAI Press, 1996.
- Guideline subcommitte of the World Health Organization-International society of hypertension. 1999 World Health Organization-International society of hypertension guidelines for the management of hypertension, J Hypertens 1999; 17:151-183.
- 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 VII), JAMA 2003;289: 2560-2572.
- 맹광호. 본태성 고혈압의 원인, 한국역학회지 1992; 24(2):111-116.
- Giuse DA, Miller NB. Strategies for medical knowledge acquisition, Medical informatics (Bemmel JH and Musen MA), Houten, Springer, 1997:277-292.
- Chae YM, Ho SH, Cho KW, Lee DH, Ji SH. Data mining approach to policy analysis in a health insurance domain. Int J Med Inf 2001; 62:103-111.
- 호승희, 채영문, 조경원, 지선하, 이동하. 의료분야의 지식경영을 위한 데이터마이닝 응용, 대한의료정보학회지 1999;5(3):37-41.
- Long WJ, Griffith JL, Selker HP, D'Agostino RB. A comparison of logistic regression to decision-tree induction in a medical domain. Comput Biomed Res 1993;26:74-97.