데이터마이닝 기법을 활용한 고혈압 관리를 위한 의사결정지원시스템의 개발

Development of Decision Support System for the Management of hypertension using Datamining Technology

호승희[1], 채영문[1], 조승연[2], 최동훈[2], 송용욱[3], 박충식[1], 조경원[1], 송지원[1]

요 약

본 연구의 목적은 데이터마이닝 기법을 이용하여 임상적으로 중요한 위치를 차지하고 있는 고혈압 환자의 특성과 치료에 따른 예후를 예측할 수 있는 지식을 발굴하고 이의 임상적용의 타당성을 검증하여 의사결정지원시스템을 개발하고 이의 유용성을 평가하는데 있다. 이에 연세대학교 의과대학 부속 세브란스 병원의 환자를 대상으로 로지스틱 회귀분석을 이용하여 혈압조절상의 위험요인을 규명하고, 의사결정나무분석을 통해 치료약제별 혈압조절군과 비조절군의 특성을 도출하고 각 대상군을 결정짓는 규칙을 생성하였으며, 이를 활용한 의사결정지원시스템의 개발 및 평가를 시행하였다. 그 결과 기존 임상이론만을 활용한 시스템의 처방에 의한 혈압조절군보다 데이터마이닝 기법을 활용한 시스템의 처방에 의한 혈압조절군의 비율이 전체적으로 더 높게 나타남을 알 수 있었다. 본 연구의 결과는 우리나라 현실에 부합되는 고혈압 진료지침을 개발하고 적용, 평가하는데 기여할 수 있을 것으로 판단되며, 이와 같은 의사결정지원시스템을 운영을 통해 실제 임상 진료에 적용해 봄으로써 그 효과와 실증적 가치를 창출할 수 있을 것이다.

key words : Data Mining, Hypertension, Decision Support System, Decision Tree Analysis, Validation

1)연세대학교 보건대학원 보건정보관리학과 2)연세대학교 의과대학 심장내과학교실 3)경상대학교 경영대학 경영정보학과 4)영동대학교

1. 서 론

고혈압은 복잡한 현대사회를 사는 성인들에게 호발하는 만성퇴행성질환으로 높은 유병률과 함께 주요 사망원인질환 발생과 깊은 관련이 있어 이의 예방 및 치료, 관리 대책 수립이 요구된다. 또한 다양한 차이를 보이는 각 개인을 치료하는데 있어 균형적인 정보를 제공할 수 있는 실용 지침이 필요하다. 지금까지 많은 연구에서 고혈압의 위험요인에 근거한 치료방법의 중요성을 제시하였으며, 치료 예후와의 관련성 연구의 필요성에 대해 언급하였으나 (WHO, 1999), 이와 같은 실용화 지침에 활용할 수 있는 기본적인 연구자료가 부족한 실정이다. 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 임상의사나 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 의사결정지원시스템에 대한 연구가 진행중이다. 의료분야에서도 단순 정보 제공보다는 보다 해석적이며 감별 진단 및 치료에 도움을 줄 수 있는 clinical consultant의 역할을 수행하는 시스템이 요구된다. 이에 의사의 임상적인 의사결정을 지원하는 시스템으로 인공지능 기법 등을 이용한 의사결정지원시스템이 외국에서는 많이 개발되고 있다. 진료를 위한 의사결정지원시스템(Clinical decision support system)은 임상의사가 환자를 진료를 행하는 과정

중에 이루어지는 의사결정에 지식을 제공함으로써 이를 지원할 수 있는 시스템을 말한다(Robert, 1999). 현재까지 의사결정지원시스템을 개발하기 위하여 많이 활용된 인공지능기법은 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 획득(knowledge acquisition)에 의존해야 하므로 이로 인한 여러 한계점이 제기되고 있다. 필요한 지식을 임상에 적용시키기 위하여서는 문제해결을 위한 의사결정과정과 관련된 지식의 개념을 정립하여 표현해야 하는 어려움이 있다. 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. 의료분야에서도 다른 분야와 마찬가지로 의료정보자원의 효율적 활용을 위한 지식기반 경영과 이의 요소정보기술로써 데이터웨어하우스와 데이터마이닝의 중요성이 부각되고 있다.

본 연구에서는 의료분야의 의사결정지원을 위한 지식경영체제 도입을 촉진시키기 위한 방안으로써 데이터마이닝 기법을 적용하여 대규모 데이터 내에 존재하지만 숨겨져 있는 상호관련성과 패턴에 대한 탐색을 통해 유용한 지식을 이끌어 내고자 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 치료에 따른 예후를 예측할 수 있는 지식을 발굴하고자 하였다. 또한 이의 임상 적용의 타당성을 검증하여 의사결정지원시스템을 개발하고 향후 임상에서의 적용 방안을 제시하고자 하였다.

2. 연구 방법

2.1 연구 대상

연구 대상은 후향적 연구(retrospective study)에 의한 대상군과 전향적 연구(prospective study)에 의한 대상군으로 총 1,709명으로 구성하였다. 후향적 연구 대상군은 1999년 5월부터 7월 사이의 기간 중 연세대학교 의과대학 부속 세브란스 병원에 입원한 환자중에서 주진단(primary diagnosis)이나 부진단(secondary diagnosis) 중 고혈압이 있는 입원 환자중에서 일련의 검사 및 치료에 응한 환자 926명으로 선정하였다. 그리고 1999년 10월부터 2000년 2월까지 연세대학교 의과대학 부속 세브란스 병원 심장내과에 내원한 환자 중 고혈압이 의심되거나 또는 고혈압을 지니고 있어 설문조사에 응한 환자 중 약물 처방을 받고 follow-up하여 다시 임상적 평가를 시행한 환자 790명을 전향적 연구 대상군으로 선정하였다.

2.2 데이터마이닝 기법을 활용한 지식의 발굴

프로세스의 관점에서 데이터마이닝은 어떠한 패턴에 관한 가설검증이나 이미 알려진 견해 없이 데이터베이스에서 알려지지 않은 패턴을 발견하는 과정인 발견과정(discovery process)과, 데이터 내에서 패턴을 알아내는 발견과정 동안 새로운 데이터 아이템에 관한 값들을 예측하는 예측모델링(predictive modeling)으로 구분할 수 있다(Witten 등, 1999). 본 연구에서는 데이터마이닝 프로세스를 통해 환자의 특성, 진단, 치료방법, 치료결과간의 상관관계와 패턴을 발견하고 이를 통해 치료 결과를 예측할 수 있는 규칙을 도출하였다. 즉 치료 결과의 분류와 예측의 과정을 추론 규칙에 의해 표현하고, 이를 통해 대상군별 환자 및 치료의 특성을 분석하여 각 대상군, 즉 치료후 혈압조절군과 비조절군을 결정짓는 규칙을 생성하였다.

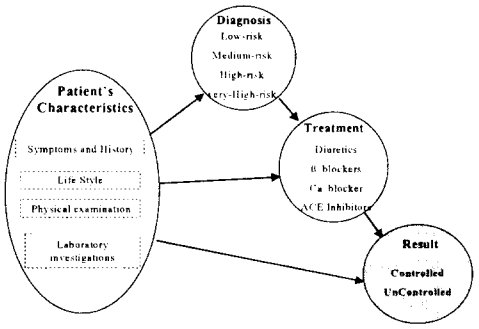


Figure 1. Framework of the analysis

2.2.1 분석 항목의 선정

2.2.1.1 환자의 특성에 따른 변수 선정

본 연구의 원래 데이터는 실제 병원 의무기록데이터 및 처방전달시스템의 결과 내역으로 이 중 고혈압 환자의 진단과 치료에 있어서 중요한 특성으로 파악되어지는 항목들을 선별하는 과정이 이루어졌다. 이 과정은 고혈압에 대한 임상진료지침 등의 임상이론, 고혈압 전문의의 견해, 실제 환자 데이터에 대한 기술통계량(descriptive statistics) 측정을 통한 데이터의 특성 파악 등을 통해 이루어졌다. 이에 따라 '나이, 성별, 비만도, 증상, 과거력 및 현재력, 흡연 및 음주량, 흉부촬영검사, 심전도검사, 생화학검사, 뇨화학검사' 등의 변수를 환자의 특성에 관련된 변수로 선정하였다.

2.2.1.2 진단

고혈압의 진단분류는 WHO(World Health Organization)의 심혈관 질환의 위험인자에 따른 환자의 계층화를 토대로 하여 'low-risk group, medium-risk group, high-risk group, very-high-risk group'의 네 군으로 범주화하였다.

2.2.1.3 치료방법

환자 치료에 사용된 혈압강하 약품을 일차적으로 WHO에서 제시한 'Diuretics, β-blocker, α-blocker, Calcium channel blocker'의 4가지 약제로 분류하고, 실제 임상에서의 항고혈압 약제의 선택시 단일 약제에 의한 치료보다 두 가지 이상의 약제를 병용한 복합 치료가 많은 점을 감안하여 약물 병용의 경우를 'Diuretics + β-blocker', 'Diuretics + Calcium channel blocker' 등의 10가지 복합처방으로 추가하였다.

2.2.1.4 치료결과

고혈압 환자의 치료결과는 혈압조절의 여부로 측정하였으며, 이의 분류 기준은 JNC VI와 WHO 위원회에서 발표한 정의와 분류를 수용하여 투약이후 수축기 혈압 140mmHg미만, 이완기 혈압 90 mmHg미만으로 떨어져 정상 혈압을 나타낸 경우는 치료후 혈압 조절군(controlled group)으로, 그렇지 않은 경우는 치료후 혈압 비조절군(uncontrolled group)으로 분류하였다.

2.2.2 로지스틱 회귀분석(Logistic regression)을 이용한 위험요인(risk factor) 규명

어떠한 조건하에 사건이 일어날 확률(P_x)과 그러한 조건을 가지지 않은 경우에 사건이 일어날 확률($1-P_x$)의 비, 즉 $P_x/(1-P_x)$를 오즈(odds)라고 하고 오즈를 자연대수한 값 $\ln(P_x/1-P_x)$을 로짓(logit : log unit)이라 했을 때 이러한 $\text{logit}(P_x)$를 종속변수로 하고, $X_1, \ldots, X_p$를 독립변수로 한 선형모형을 구성할 수 있는데, 이와 같은 분석방법을 로지스틱 회귀분석이라고 한다(David 등, 1989). 본 연구에서는 종속변수 혈압조절여부와 환자의 특성인 여러 독립변수와와 관계를 알아보고자 하는 회귀분석적인 방법으로서 로지스틱 회귀분석을 실시하였으며, 이로써 고혈압 치료상의 위험요인을 규명하고자 하였다. 즉 고혈압의 치료와 연관이 있다고 알려진 요인들이 얼마만큼 독립적으로 치료결과에 영향을 주는지를 비차비(odds ratio)를 측정하여 살펴보았다.

2.2.3 의사결정나무분석(Decision Tree Analysis)를 이용한 대상군의 분류 및 예측

의사결정나무분석은 의사결정규칙(decision rule)을 나무구조로 도표화하여 분류(classification)와 예측(prediction)을 수행하는 분석방법으로 예측의 과정이 나무구조에 의한 추론규칙(induction rule)에 의해 표현된다(Lee 등, 1999). 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. 의사결정나무분석은 하나의 나무구조를 이루고 있으며, 마디(node)라고 불리는 구성요소들로 이루어져 있고 목표변수(target variable)인 뿌리마디(root node)로부터 시작하여 예측변수(predictor variable)에 의해 각 가지(branch)가 끝마디(terminal node)에 이를 때까지 자식마디(child node)를 계속적으로 형성해 나감으로써 완성된다. 의사결정나무 형성을 위한 다양한 기준들이 제안되어 있는데, 이는 하나의 부모마디로부터 자식마디들이 형성될 때 예측변수의 선택과 병합이 이루어지는 기준인 분리기준(splitting criterion), 정지규칙(stopping rule), 가지치기(pruning)방법으로서 이들을 어떻게 결합하느냐에 따라 서로 다른 의사결정나무형성방법들이 만들어진다(Choi 등, 1998) .

그리고 1980년대 이후 CHAID, CART 등 다양한 알고리즘이 제안되어 왔는데, 본 연구에서는 이러한 알고리즘을 구현할 수 있는 의사결정나무분석 솔루션 중의 하나인 SPSS사의 Answer Tree 1.0을 사용하였고, 변수의 성격이 범주형 데이터이고 예측변수와 목표변수간의 관계를 찾아야 할 때 가장

유용한 방법인 CHAID 방법을 적용하였다. 환자의 특성 중 어떠한 특성이 고혈압을 결정짓는데 가능성이 높은가를 예측하려 할 경우, CHAID 알고리즘은 최상의 예측변수로서 결정된 변수를 이용하여 혈압 정도를 결정짓는데 가장 큰 차이를 갖는 두 개 이상의 구분된 집단으로 나누고 그 결과를 트리구조로 나타낸다. CHAID 알고리즘의 카이제곱 통계량을 분리기준으로 사용하여 p값이 지정한 유의수준($\alpha=0.1$) 보다 크면, 그 예측변수(환자의 특성)는 목표변수(혈압조절 여부)의 분류에 영향을 주지 않는 것으로 간주하여 자식 마디를 형성할 대상에서 제외하였다. 결국 분리기준을 카이제곱통계량으로 한다는 것은 p값이 가장 작은 예측변수와 그 때의 최적분리에 의해 자식마디를 형성시킨다는 것을 의미하는데, 목표변수 자체의 빈도가 나무의 맨 위에 위치하게 되고 각 예측변수 중 가장 윗쪽에 위치하는 변수가 목표변수에 가장 영향력이 높은 변수이다. 그리고 의사결정나무를 형성한 후에는 형성된 나무가 얼마나 타당성을 가지고 만들어졌는지를 평가하는 것이 중요한데, 이는 오분류확률(misclassification probability)을 나타내는 위험도표(risk chart)를 가지고 수행하였다.

2.2.4 데이터마이닝 모형의 분석 및 평가

의사결정나무 모형에서 추출된 지식과 임상이론 및 고혈압 전문의의 지식을 비교하여 그 일치 여부 및 정도에 따라 consistent rule 과 inconsistent rule로 분류하였다. 그리고 inconsistent rule에 대해 일치하지 않는 이유를 분석하여 그 인과관계를 규명하고 기존 임상이론에 추가 또는 보완할 수 있다고 판단되는 규칙인 acceptable rule과 bias 또는 confounding variable 등으로 인하여 신뢰성이 의심되는 규칙인 unacceptable rule로 분류한다. 그리고 실제 환자의 데이터 값과 데이터마닝 모형의 결과의 일치 여부를 측정함으로써 모형의 예측력을 평가했다. 이에 따라 true positive와 true negative를 합한 값인 예측도(prediciive rate)와 true positive인 민감도(sensitivity), true negative인 특이도(specificity)를 산출하였다.

2.3 고혈압 관리를 위한 의사결정지원시스템의 개발

임상이론 및 데이터마이닝 기법에 의해 도출된 규칙을 활용하여 규칙베이스를 구성하고 웹 기반 추론엔진을 사용하여 고혈압 관리를 위한 의사결정지원시스템을 개발하였다.

HTML의 하이퍼링크를 이용하면 의사결정나무를 구현할 수 있다. 또한 의사결정문제를 구현한 HTML파일들이 반드시 나무(Tree) 구조일 필요는 없으며, 그래프(Graph)의 구조를 갖는 것도 가능하다. 이 특성을 이용하면 하이퍼링크 된 HTML 파일들의 나열을 통하여 역방향 추론을 구현하였다.

한편, 역방향 추론 기관용 지식베이스의 규칙들을 연쇄(Chaining) 구조의 관점에서 살펴보면 그들이 의사결정 그래프 구조를 갖고 있다. AND, OR, NOT 등에 의해 연결된 다수개의 조건들을 조건절에 가진 규칙들은 의사결정그래프로 변환이 가능하다. 따라서 하이퍼링크된 HTML 파일들의 나열만으로도 역방향 추론을 구현할 수 있는데, 이를 위해서는 사실(Fact)의 입력, 하위목표(Subgoal)에 의한 연쇄, 추론결과의 출력 등이 지원이 필요했다. 하위목표에 의한 연쇄는 상기된 바와 같이 하이퍼링크에 의해 구현할 수 있으며, 추론결과의 출력은 웹 부라우저가 디스플레이할 HTML 파일을 통해 이루어졌다. 따라서, 남은 문제는 사실의 입력을 지원하는 방안이었는데 역방향 추론에서 사실 값은 변수에 저장된다.

역방향 추론에서 변수의 형태는 진위형(Fact Type), OAV(Object-Attribute-Value Type)형, 수치형(Numeric Type)의 세 가지가 있는데 여기서 진위형과 OAV형 변수는 하이퍼텍스트의 나열에 의해 쉽게 해결되었으며, 수치형은 값을 입력 받은 후 계산하여야 되기 때문에 하이퍼텍스트의 단순 나열로는 해결되지 않았다. 해결을 위한 방안으로는 수치형 변수 처리용 CGI를 구현하는 것이었으나, 이것보다 더 간단하고 효과적인 방법은 JavaScript(Danesh, 1996), VBScript(Jerke et al., 1997) 등과 같은 클라이언트 측 스크립트 언어(Client-side Script Language)를 이용하는 것이었다. 따라서 고혈압 관리를 의한 의사결정시스템의 추론엔진은 HTML언어와 JavaScript를 통해 구현되었다.

2.4 데이터마이닝 기법을 활용한 의사결정지원시스템의 타당성 검증

기계학습(machine learning) 기법으로부터 도출된 지식을 전문가의 검증없이 그대로 사용하는 것은 때때로 매우 위험하며, 이러한 지식들을 최종 사용자가 어떻게 받아들이냐는 것은 데이터마이닝 기법의 실제 활용 측면에서 매우 중요한 문제라고 할 수 있다. 그러나 현재까지 인공지능과 통계학적 방법을 포함한 데이터마이닝 기법과 전문가 판단과의 체계적인 상호작용(structured interaction)에 대한 연구는 많이 이루어지지 않고 있다(Kim 등, 1999). 따라서 본 연구는 데이터마이닝 기법으로부터 도출된 지식과 임상이론 및 전문가의 지식으로부터 추출된 지식을 실증적으로 비교연구함으로써 궁극

적으로 전문가의 의사결정에 도움이 되는 데이터마이닝 기법의 활용 방안을 제시하고자 하였다. 이에 따라 임상이론과 데이터마이닝 기법의 단독 및 혼합에 의한 처방규칙에 대한 치료결과를 분석함으로써 데이터마이닝 기법을 활용한 의사결정지원시스템의 타당성을 검증하였다. 이를 위해 일차적으로 기존 임상진료지침에서 제시하는 치료방법에 따른 규칙기반시스템을 개발하고 이에 따른 치료 결과를 분석하였다. 임상진료지침에서의 각 적응증은 대해 해당 약제를 처방할 경우 혈압이 조절될 것이라는 이론하에 설정된 것이라고 볼 수 있으며, 본 연구에서는 이러한 임상이론에 의한 처방규칙을 규칙화하여 시스템을 구현하고 이 때의 혈압조절군과 비조절군의 비를 비교하여 평가했다. 또한 데이터마이닝 기법을 통해 도출된 치료약제별 혈압조절군과 비조절군의 특성을 도출하고, 이 중 혈압조절군의 특성을 규칙화하여 제시하고 앞에서와 마찬가지로 이 때의 혈압조절군과 비조절군의 비를 비교하여 평가했다. 그 다음 기존 임상이론과 데이터마이닝 기법을 통해 도출된 규칙 중 수용가능한 규칙과의 결합을 통한 혼합모형을 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.

3. 연구 결과

3.1 혈압 조절상의 위험 요인

혈압조절상의 위험 요인 규명을 위한 환자의 특성 및 치료방법에 관련되는 변수 중 회귀계수의 추정 및 유의성 검정 결과에서 10% 유의수준 검정에서 유의한 변수는 (figure 2)에 제시된 6개의 변수였다. 이때 비차비가 가장 높은 변수는 혈압의 정도를 나타내는 BP stage로 비차비가 2.22로서, BP stage가 한 단위 증가할수록 혈압 조절이 안 될 확률이 2.22배 높아짐을 알 수 있었다. 그 다음으로 비차비가 높은 변수는 수축기 고혈압(systolic hypertension)으로 비차비가 2.03으로서, 수축기 고혈압일 경우 아닐 경우보다 혈압 조절이 안 될 확률일 2.02배 높아짐을 알 수 있었다. 이와 같이 로지스틱 회귀분석을 통하여 혈압 조절 여부, 즉 치료결과와 위험요인과의 관계를 추정한 결과 혈압 조절상의 위험요인은 BP stage, 수축기 고혈압 여부, 좌심실비대(LVH), Creatinine, RBC, 비만도(BMI)의 순으로 나타났다.

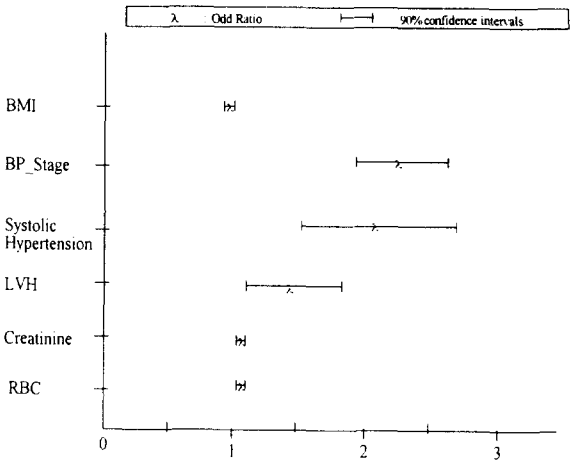


Figure 2. Odds ratios for the control of hypertension

3.2 치료약제별 혈압조절군과 비조절군의 결정 규칙

로지스틱 회귀분석을 통하여 혈압조절상의 위험요인을 규명할 수는 있었으나 각 요인들이 상호 결합된 상황에서의 치료결과를 예측하고 이를 규칙화할 수는 없는 제한점이 있었다. 이에 따라 의사결정나무분석을 통해 치료약제별 처방 환자를 대상으로 치료결과의 분류와 예측의 과정을 나무구조에 의한 추론규칙에 의해 표현하였으며, 치료결과에 영향을 미치는 요인을 도출하고 이를 통해 대상군별 환자의 특성을 규명하고 이로써 각 대상군, 즉 혈압조절군과 비조절군을 결정짓는 규칙을 생성할 수 있었다. 또한 이와 같은 예측 분류에 대한 신뢰도를 측정함으로써 예측가능성을 점수화하였다.

Calcium channel blockers의 경우 (Table 1)에서 나타난 바와 같이 이를 처방한 환자군 중 혈압 조절 여부를 결정짓는 첫 번째 특성은 진단분류로 low-risk 그룹과 medium-risk 그룹에서의 비조절군의 비율은 38.98%, high-risk 그룹과 very-high-risk 그룹에서의 비조절군의 비율은 59.60%로 혈압 비조절군의 비율이 상위 위험 군에서 1.5배 높음을 알 수 있다. 두 번째 특성은 대상군의 특성에 따라 달라지는데 하위 위험군에서는 연령이 혈압 조절여부를 결정짓는 가장 큰 특성으로 61세 이하의 군에서는 대체적으로 비조절군의 비율이 높았으나 61세 이상의 군에서는 조절군의 비율이 82.35%로 현저히 높게 나타났다. 이는 Calcium channel blockers가 고연령층에 대한 효능·효과가 있었음을 나타낸다. 그리고 상위위험군에서 혈압조절여부를 결정짓는 가장 유의한 변수는 '부종'으로 '부종'의 증상이 있는 군이 없는 군에 비조절군의 비율이 1.2배 높게 나타났다. 상위위험군에서 이 외

의 유의한 결과는 부종이 없고 심장질환의 합병증이 있으며 고혈압와 가족력이 있는 군에서 'CA'의 수치가 혈압 조절에 매우 결정력이 있는 것으로 나타났는데, CA가 8.9 이상인군에서 비조절군의 비율이 78.26%로 8.9 미만인군의 20%에 비해 3.9배 높게 나타났다. 그리고 이 때 CA가 8.9 이상인 군에서 흉통이 있는 군이 없는 군에 비해 혈압조절군의 비율이 4.3배 높게 나타나 Calcium channel blockers가 흉통에 대한 효능·효과가 있는 것으로 나타났다. 또한 β-blockers의 경우 혈압 조절여부를 결정짓는데 가장 영향력 있는 변수는 당뇨에 대한 가족력이었으며, 그 다음으로 흉통, 단백뇨, 심장질환유무, 음주량, 고혈압 가족력의 순으로 영향력있는 변수가 나타났다.

이 결과 β-blocker는 당뇨에 대한 가족력이 있는 경우, 당뇨에 대한 가족력이 없는 군에서는 단백뇨가 있거나 음주량이 많을 경우, 또는 흉통 및 심장질환이 있는 경우에 효능·효과가 있었음을 알 수 있었다(Table 2).

(Table 3)은 Calcium channel blockers 처방군에서의 혈압 조절군과 비조절군의 특성을 규칙화한 것인데, 규칙 1의 경우 '진단분류가 low-risk 또는 medium-risk이고 연령이 19세에서 59세 사이며 여자이고 고혈압 가족력이 없는 경우'에는 혈압 조절군으로 분류되며, 이에 대한 예측신뢰도는 100%란 것을 나타낸다. 마찬가지로 β-blockers 처방군에서의 혈압조절군과 비조절군의 특성과 규칙을 제시하였다(Table 4).

Table 1. Tree structure for the treatment by Calcium channel blockers

<table>
<tr><td rowspan="19">DX [1:2]
uncontrol 38.98
control 61.02</td><td rowspan="3">AGE [19,59]
uncontrol 55.56
control 44.44</td><td rowspan="2">GENDER [female]
uncontrol 41.67
control 58.33</td><td>Fhx_HTN [0]
uncontrol 00.00
control 100.00</td><td></td></tr>
<tr><td>Fhx_HTN [1]
uncontrol 71.43
control 28.57</td><td></td></tr>
<tr><td>GENDER [male]
uncontrol 83.33
control 16.67</td><td colspan="2"></td></tr>
<tr><td>AGE [59,61]
uncontrol 100.00
control 00.00</td><td colspan="3"></td></tr>
<tr><td>AGE [61,87]
uncontrol 17.65
control 82.35</td><td colspan="3"></td></tr>
<tr><td rowspan="14">DX [3:4]
uncontrol 59.60
control 40.40</td><td rowspan="13">EDEMA [0]
uncontrol 59.60
control 40.40</td><td rowspan="5">COM_HD [0]
uncontrol 65.06
control 34.94</td><td rowspan="2">MENOP [0]
uncontrol 77.27
control 22.73</td><td>RBC [1.91,3.64]
uncontrol 42.86
control 57.14</td><td></td></tr>
<tr><td>RBC [3.64,5.36]
uncontrol 83.78
control 16.22</td><td></td></tr>
<tr><td rowspan="3">MENOP [1]
uncontrol 51.28
control 48.72</td><td>AGE [19,59]
uncontrol 10.00
control 90.00</td><td></td></tr>
<tr><td>AGE [59,69]
uncontrol 83.33
control 16.67</td><td></td></tr>
<tr><td>AGE [69,87]
uncontrol 36.36
control 63.64</td><td></td></tr>
<tr><td rowspan="8">COM_HD [1]
uncontrol 50.00
control 50.00</td><td rowspan="4">Fhx_HTN [0]
uncontrol 41.94
control 58.06</td><td rowspan="2">A_DRINK [0,151.2]
uncontrol 31.25
control 68.75</td><td>GENDER [female]
uncontrol 45.83
control 54.17</td><td></td></tr>
<tr><td rowspan="2">GENDER [male]
uncontrol 16.67
control 83.33</td><td>LVH [0]
uncontrol 5.56
control 94.44</td></tr>
<tr><td>LVH [1]
uncontrol 50.00
control 50.00</td></tr>
<tr><td>A_DRINK [151.2,1209.6]
uncontrol 78.57
control 21.43</td><td colspan="2"></td></tr>
<tr><td rowspan="4">Fhx_HTN [1]
uncontrol 67.86
control 32.14</td><td>CA [5.7,8.9]
uncontrol 20.00
control 80.00</td><td colspan="2"></td></tr>
<tr><td rowspan="3">CA [8.9,10.7]
uncontrol 78.26
control 21.74</td><td rowspan="2">CHEST_PAIN [0]
uncontrol 88.24
control 11.76</td><td>A_SMOKE [0,18.4]
uncontrol 100.00
control 00.00</td></tr>
<tr><td>A_SMOKE [18.4,22.5]
uncontrol 60.00
control 40.00</td></tr>
<tr><td>CHEST_PAIN [1]
uncontrol 50.00
control 50.00</td><td></td></tr>
<tr><td>EDEMA [1]
uncontrol 76.00
control 24.00</td><td colspan="4"></td></tr>
</table>

Table 2. Tree structure for the treatment by Beta blockers

<table>
<tr><td rowspan="5">Fhx_DM [0]
uncontrol 58.62
control 41.38</td><td rowspan="3">CHEST_PAIN [0]
uncontrol 66.67
control 33.33</td><td rowspan="2">PROTEIN(U/A) [0]
uncontrol 72.22
control 27.78</td><td>A_DRINK [0]
uncontrol 84.62
control 15.38</td><td>Fhx_HTN [0]
uncontrol 100.00
control 00.00</td></tr>
<tr><td>A_DRINK [0]
uncontrol 40.00
control 60.00</td><td>Fhx_HTN [1]
uncontrol 60.00
control 40.00</td></tr>
<tr><td>PROTEIN (U/A) [1]
uncontrol 33.33
control 66.67</td><td colspan="2"></td></tr>
<tr><td rowspan="2">CHEST_PAIN[1]
uncontrol 37.50
control 62.50</td><td>COM_HD [0]
uncontrol 66.67
control 33.33</td><td colspan="2"></td></tr>
<tr><td>COM_HD [1]
uncontrol 20.00
control 80.00</td><td colspan="2"></td></tr>
<tr><td>Fhx_DM [1]
uncontrol 0.00
control 100.00</td><td colspan="4"></td></tr>
</table>

Table 3. Decision rules for the treatment by Calcium channel blockers

RULE	DX	Edema	Com_HD	Menop	RBC	A_Drink	Age	Gender	LVH	F_HTN	CA	Reliability of prediction
1	Med or Low						19<=Age<=59	female		no		100.0%
2	VHigh or High	no	yes			0<=A_Drink<=151.2		male	no			94.4%
3	VHigh or High	no	no	yes			19<=Age<=59					90.0%
4	Med or Low						61<Age<=87					82.4%
5	VHigh or High	no	yes							yes	5.7<=CA<=8.9	80.0%
6	VHigh or High	no	no	yes			69<Age<=87					63.6%
7	VHigh or High	no	no	no	1.9<=RBC<=3.6							57.1%
8	VHigh or High	no	yes			0<=A_Drink<=151.2		female		no		54.2%

Table 4. Decision rules for the treatment by Beta blockers

RULE	FHX_DM	Chest_Pain	Com_HD	Protein(U/A)	A_Drink	Reliability of prediction
1	yes					100.0%
2	no	yes	yes			80.0%
3	no	no		yes		66.7%
4	no	no		no	A_Drink>13.5	60.0%

3.3 데이터마이닝 모형의 결과 분석 및 예측력 평가

단일 약제 처방에 대하여 앞에서 제시한 임상이론에 의한 처방규칙과 데이터마이닝기법에서 추출된 처방규칙을 비교검증하였다. 단일 약제인 네 가지 약제 중에서 일치하는 경우는 calcium channel blockers에 대한 적응증인 '고연령(elderly patient)'뿐이었다. 그리고 inconsistent rule 중 고혈압 전문의의 검토하에 설정된 acceptable rule은 calcium channel blockers는 '심장질환, 흉통, 저칼슘증'에 대한 적응증, β_blockers의 '당뇨 가족력, 흉통, 단백뇨, 심장질환'에 대한 적응증, ACE는 'uric acid 수치가 높고 BUN 수치가 높거나 또는 심장질환이 있는 경우'에 대한 적응증 이었다. 그러나 Diuretics는 기존 임상 이론에 추가할 만한 별다른 특성이 도출되지 않았다. 복합 약제 처방의 경우 이에 대한 임상이론이 없어 비교검증을 수행하지는 못하였다.

또한 이와 같은 혈압조절군에 대한 특성을 근거로 약제에 대한 적응증을 정의하여 규칙화하였으며, 이를 실제 환자 사례에 적응시킴으로써 치료결과에 대한 예측력을 평가하였다. 이 중 단일약제에 대한 결과만을 〈Table 5〉에 제시하였다.

Table 5. Estimated predictive rates

drug	ensitivity	specificity	predictive rate
Diuretics	95.8 %	37.5 %	81.3 %
β-blockers	82.4 %	76.5 %	80.0 %
ACE	100.0 %	72.2 %	88.64 %
Calcium Channel blockers	75.0 %	80.1 %	78.3 %

3.4 고혈압 관리를 위한 의사결정지원시스템

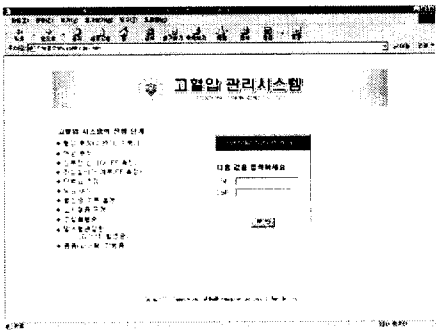
증상, 위험요인, 검사결과, 동반질환 등 환자의 특성에 관련된 요소들을 층별화 하여 환자의 특성에 따라 약물 요법을 제시하는 기능을 수행하는 의사결정지원시스템이며, 문답형으로 진행되고 총 12단계로 이루어져 있다. 웹상에서의 고혈압 추론엔진은 Javascript를 사용한 HTML파일로 이루어져 있으며 다음과 같은 방식으로 구현된다. 예를 들어 다음과 같은 규칙이 있다고 하자.

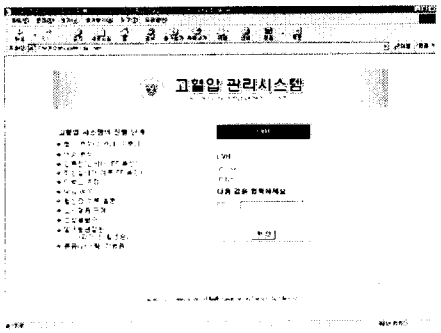
IF SBP 〉= 140 and DBP 〈 90
THEN
Hypertension
ELSE
Non-Hypertension

위 규칙은 상기한 바와 같이 JavaScript를 사용하여 HTML 파일에 의해 구현될 수 있다. 다음 그림에서 보듯이 모든 변수들의 값은 FORM 태그에서 입력받으며, 입력된 변수를 포함한 수식의 평가 및 이에 따른 하이퍼링크는 SCRIPT 태그 내에서 이루어지고 있다. JavaScript 자체가 하나의 프로그래밍 언어이므로 조건절의 수식이 임의의 모든 수식이 될 수 있다는 것이 이 방식의 또 다른 장점이다.

```
〈html〉
〈head〉
〈title〉 다음 값을 입력하세요. 〈/title〉
〈SCRIPT LANGUAGE=JavaScript〉
〈!--
function WBIBranch(form)
{
 if (form.SBP.value 〉= 140 && form.DBP.value 〈 90)
 {
 location = "condition-2-0.html";
 }
 else
 {
 location = "condition-2-1.html";
 }
}
//--〉
〈/Script〉
```

Figure 3. JavaScript를 이용한 수치형 변수의 구현





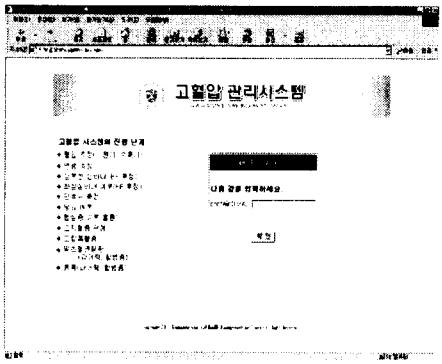


Figure 3. 변수 형태별 질문 화면

3.5 데이터마이닝 기법으로부터 발굴된 지식에 대한 타당성 검증

첫번째, 임상이론에 의한 규칙기반시스템의 타당성 검증을 실시하였다. 실제 환자 데이터를 대상으로 적응증을 살펴보면, '수축기성 고혈압, 심부전 등'의 특성을 가진 경우로 나눌 수 있으며, 본 연구에서 각 적응증에 대하여 임상진료지침에서 제시하는 약제를 처방한 경우의 치료결과를 살펴 본 결과 혈압 비조절군의 사례가 예상외로 높게 나타남을 알 수 있었다. 두 번째 데이터마이닝 기법으로부터 도출된 규칙을 활용하여 적응증에 대한 새로운 처방 규칙을 찾아내어 규칙기반시스템을 개발하여 적용한 결과 '고연령, 협심증, 고지혈증, 통풍'을 가진 환자에 대한 처방 시 기존 임상이론에 의한 처방시보다 혈압조절군의 비율이 증가하였고, '단백뇨'에 대한 처방시는 오히려 감소하였다. 그 다음으로 두 가지 방법을 상호보완한 혼합 모형에 대한 타당성 검증을 실시한 결과, 기존 임상이론만을 활용한 시스템의 처방에 의한 혈압조절군보다 데이터마이닝 기법을 활용한 시스템의 처방에 의한 혈압조절군의 비율이 전체적으로 더 높게 나타남을 알 수 있었다 (Table 6.7).

4. 고찰 및 결론

현대의 질환은 점차로 고혈압이나 당뇨와 같은 만성퇴행성 질환으로 변해가는 추세이며, 데이터마이닝과 같은 새로운 기법이 이러한 질환의 예방과 관리에 적극 활용될 필요가 있다(Ho 등, 1999).

Table 6. Comparison of practice guideline with the modified practice guideline (complication)

Patient Characteristics	Clinical Practice Guideline			Data Mining			Hybrid Model		
	Desirable Agents	Treatment Result		Desirable Agents	Treatment Result		Desirable Agents	Treatment Result	
		Control	Uncontrol		Control	Uncontrol		Control	Uncontrol
Systolic Hypertension	1,4,9	39.76 %	60.34%				1,4,9	39.76 %	60.34%
Elderly Patients	1,2,4,5,9,11	51.04 %	48.06%	4	62.08 %	38.92%	4	62.08 %	37.02%
Heart Failure	1,3,6	42.5%	53.50%				1,3,6	80.00 %	20.00%
LVH	2,3	52.94 %	47.04%				2,3	52.94 %	67.06%
Proteinuria	1,3,6	72.22 %	27.78%	2	66.67 %	33.33%	1,3,6	72.22 %	27.78%
Diabetes	3,4,8	48.00 %	52.00%				3,4,8	48.00 %	53.00%
Angina	2,4,7	40.23 %	59.77%	2,4,8	40.65 %	59.35%	2,4,7,8	40.37 %	59.63%
Hyperlipidemia	3,4,8	43.75 %	56.25%	8	50.00 %	50.00%	8	50.00 %	50.00%
Peripheral Vascular Disease	4	25.00 %	75.00%				4	25.00 %	75.00%
Hyperkalaemia	1,2,4,5,7,9,11	60.00 %	40.00%				1,2,4,5,7,9,11	60.00 %	40.00%
Gout	2,3,4,7,8,13	55.56 %	44.44%	7	100.00 %	0%	2,3,4,7,8,13	55.56 %	44.54%

Table 7. Comparison of practice guideline with the modified practice guideline (no complication)

Clinical Practice Guideline				Data Mining				Hybrid Model			
Patient Characteristics	Desirable Agents	Treatment Result		Patient Characteristics	Desirable Agents	Treatment Result		Patient Characteristics	Desirable Agents	Treatment Result	
		Control	Uncontrol			Control	Uncontrol			Control	Uncontrol
No Complication	1.2.5	63.16 %	37.84%	Na <= 140	6	78.13 %	22.87%	Na <= 140	6	78.13 %	22.87%
				Paralysis	7	100.00 %	0%	Paralysis	7	100.00 %	0%
				Headache	8	68.18 %	31.82%	Headache	8	68.18 %	31.82%
				Fhx_Stroke	12	81.82 %	18.18%	Fhx_Stroke	12	81.82 %	18.18%
				Fhx_DM	2	100.00 %	0%	Fhx_DM	2	100.00 %	0%
				Fhx_HTN	2	60.00 %	40%	Fhx_HTN	2	60.00 %	40%

Index	Drug Class
1	Diuretics
2	Beta-blockers
3	ACE inhibitors /A-II receptor blockers
4	Calcium antagonists
5	Diuretics + Beta-blockers
6	Diuretics + ACE inhibitors
7	Beta-blockers + Calcium antagonists
8	ACE inhibitors + Calcium antagonists
9	Diuretics + Calcium antagonists
10	Diuretics + Beta-blockers + ACE inhibitors
11	Diuretics + Beta-blockers + Calcium antagonists
12	Diuretics + ACE inhibitors + Calcium antagonists
13	Beta-blockers + ACE inhibitors + Calcium antagonists
14	Diuretics + Beta-blockers + ACE inhibitors + Calcium antagonists

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(JNC, 1997). 따라서 고혈압 관리에서 위험요인을 찾고 이를 해결하기 위한 적절한 치료방법을 찾는 것은 무엇보다 중요하다고 할 수 있다. 이에 따라 본 연구에서는 데이터마이닝 기법을 적용하여 치료결과에 어떠한 요인이 얼마나 중요한 지를 유추해 내고 치료후 혈압조절군과 비조절군의 특성을 분류, 규명함으로써 새로운 치료 전략을 도출해 보고자 하였다. 데이터마이닝은 패턴인식 기술이나 통계기법, 수학적 알고리즘을 이용하여 의미있는 새로운 상관관계, 패턴, 추세 등을 발견하는 과정으로(Cho 등, 1998), 본 연구에서는 고혈압 관리를 위한 의사결정지원에 활용될 수 있는 유용한 지식을 이끌어내기 위한 데이터마이닝 기법으로써 통계적 기법인 logistic regression과 의사결정나무 분석의 CHAID분석을 수행하였다. 또한 여러 데이터마이닝에 의해 발굴된 지식을 기존 임상이론과 비교검증하고 이를 활용한 의사결정지원시스템을 개발하고 이의 유용성을 평가함으로써 향후 진료분야의 지식경영으로서 고혈압 관리를 위한 의사결정지원시스템 및 임상진료지침의 개발에 활용될 수 있는 최적 모형의 설계방안을 모색해 보고자 하였다. 본 연구에서 WHO의 국제적인 고혈압 관리 지침에 기초하여 종합병원 전문의의 고혈압 진료 실태를 평가한 결과 고혈압의 진단 및 치료 과정에 있어서 임상진

료지침을 잘 따르지 않는 것으로 나타났다. 이는 WHO에 근거한 원칙적인 고혈압 진료지침의 내용와 의미가 의사들에게 충분히 전달되지 않았기 때문이며, 다른 한편으로는 진료지침이 비현실적이어서 그대로 따르기가 어렵기 때문임을 시사하는 결과라고 판단된다. 이와 같이 의사의 진료과정이 기존의 진료지침에 부합되지 않는 결과를 나타냄으로써, 그 이유를 조사하여 진료행태를 바람직한 방향으로 객관화시킬 필요성이 있음을 알 수 있다. 그리고 임상진료지침을 따른 경우의 치료결과에서 혈압 비조절군의 비율이 조절군 보다 높은 경우가 많음을 알 수 있었다. 그러나 임상이론과 데이터마이닝 기법에서 도출된 규칙을 혼합하여 시스템을 개발하고 이에 대한 타당성을 검증한 결과, 처방시 기존 임상이론만을 활용한 시스템보다 혈압조절군의 비율이 더 높게 나타남을 알 수 있었다. 이는 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.

의료분야에서의 지식베이스를 활용한 의사결정지원시스템의 확산은 학계와 임상의사와의 사이를 좁혀 줄 수 있다. 일반적으로 의학 교과서를 개정하는 기간이 대개 몇 년이 되기 때문에 새로운 임상지식을 쉽게 반영할 수 없다. 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(Giuse 등, 1997). 외국에서도 미국, 영국 등을 중심으로 1990년대 들어 evidence-based medicine이 새로운 파라다임으로 받아들여지면서 종래의 직관이나 비체계적 임상경험 또는 이론적(rational) 근거에 의존하던 의사결정 방식에 일대 변화가 예측되고 있다. 특히 임상적 상황에서 의사와 환자의 결정을 도와주기 위해 묵시적 합의에서 탈피하여 보다 체계적인 합의도출 과정을 거치며, 광범위한 자료검색, 메타분석, 베이시안 분석, 비용효과분석 등의 계량적이고 명시적인 방법론을 활용한 임상지침 개발의 필요성이 대두되고 있다. 궁극적으로 의료분야에서도 다른 분야와 마찬가지로 의료정보자원의 효율적 활용을 위한 데이터마이닝과 같은 새로운 전략적 기법이 요구된다고 할 수 있다.

본 연구의 결과는 우리나라 현실에 부합되는 고혈압 진료지침을 개발하고, 적용, 평가하는데 기여할 수 있을 것으로 판단되며, 향후에는 진료지침을 개발하고 실제 진료활동에 적용해 봄으로써 그 효과를 평가하는 중재적 연구가 필요하다고 사료된다. 그리고 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 분석결과의 활용 등 시스템의 개발 단계에 있어서 충분한 임상적 평가를 통해 임상적 실용화의 가치를 높여나가야 할 것이다. 본 연구는 과거의 축적된 환자 데이터를 이용하여 데이터마이닝이란 새로운 기법을 사용함으로써 기존의 조사나 연구에서 발견할 수 없었던 의미있는 새로운 상관관계, 패턴, 추세를 밝혀내고 이와 같은 유용한 정보를 집적하고 체계화하여 고혈압의 치료 및 관리를 위한 의사결정지원시스템 및 임상진료지침의 개발에 활용할 수 있는 근거를 제시했다는 데 그 의의가 있다고 할 수 있다. 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 실용화단계에 이를 수 있도록 하여야 할 것이다. 또한 이러한 결과는 우리나라 현실에 부합되는 고혈압 진료지침을 개발하고 적용, 평가하는데 기여할 수 있을 것으로 판단되며, 이와 같은 의사결정지원시스템을 운영을 통해 실제 임상 진료에 적용해 봄으로써 그 효과와 실증적 가치를 창출할 수 있을 것이다.

참고 문헌

Cho JH. Reengineering of Enterprise Information Analysis Environment by Data Warehousing. SIGDM'98, 5-33, 1998

Choi JH, Han ST, Kang HC, Kim ES. Data mining Decision Tree Analysis using Answer Tree. Korea: SPSS Academy, 1998

David W. Stanley Lemeshow. Applied logistic regression: A Willy-interscience publication, 1989

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. Journal of hypertension 17 :151-183, 1999

Giuse DA and Miller NB. Stategies for medical knowledge acqusition. Medical informatics(Bemmel JH and Musen MA), Houten, Springer, 277-292, 1997

Ho SH, Chae YM, Cho KW, Jee SH, Lee DH. Data mining application for knowledge management in medical field. Journal of Korean society of medical infometics 5(3), 169-179, 1999

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). Arch Intern Med 157:2413-2446, 1997

Kim HS, Lee CH. Explanation-based data mining in data warehouse. Journal of intelligent information systems 5(2), 15-27, 1999

Lee GH, Na GH. Development and evaluation of Decision Tree Model for bankruptcy prediction. SIGDM'99, 179-187, 1999

Robert J. Clinical Decision support system. Electronic health records :chaining the vision(Gretchen FM, Mary AH, Kathleen AW), Philadelphia, WB Saunders, 305-315, 1999