(19) 대한민국특허청(KR)
(12) 공개특허공보(A)

(11) 공개번호 10-2012-0107750
(43) 공개일자 2012년10월04일

(51) 국제특허분류(Int. Cl.)
G06F 19/00 (2011.01)
(21) 출원번호 10-2011-0025461
(22) 출원일자 2011년03월22일
심사청구일자 2011년03월22일

(71) 출원인
계명대학교 산학협력단
대구광역시 달서구 달구벌대로 1095 (신당동)
(72) 발명자
손창식
대구광역시 중구 동덕로40길 27 (동인동3가)
김윤년
대구광역시 서구 달구벌대로 1707, 광장타운 107동 1203호 (내당동)
(뒷면에 계속)
(74) 대리인
김일환

전체 청구항 수 : 총 13 항

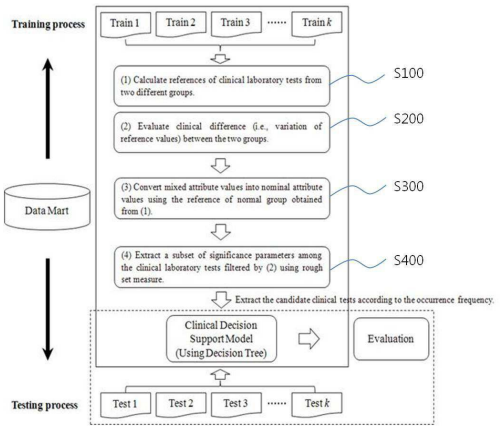
(54) 발명의 명칭 **엔트로피 러프 근사화 기술에 기반한 감별진단을 위한 주요 파라미터 추출방법 및 이를 이용한 통합 임상 의사결정 지원 시스템**

(57) 요 약

본 발명은 러프 근사화 기술기반 질환별 감별진단을 위한 주요 파라미터 추출방법 및 이를 이용한 통합 임상 의사결정 지원 시스템에 관한 것으로, (a) 다수의 검사항목별 임상 데이터가 저장된 데이터 베이스로부터 추출된 임상 데이터를 엔트로피 최대화 척도(entropy maximization measure)를 이용하여 서로 다른 2개의 데이터 그룹으로부터 임상적 참조치를 산출하는 단계; (b) 상기 서로 다른 2개의 그룹 사이의 임상적 차이를 평가하여, 후보 검사항목을 추출하는 단계; (c) 상기 어느 하나의 그룹에서 산출된 검사항목의 참조치를 기준으로 해당 검사항목의 속성값(attribute values)을 명목형 속성값(nominal attribute values)으로 변환하는 단계; 및 (d) 상기 (b) 단계에서 추출된 후보 검사항목들 사이에서 감별진단을 위한 중요 파라미터를 추출하는 단계를 포함한다.

이와 같은 본 발명을 제공하면 전처리 과정의 사용 없이 수집된 임상검사결과의 원 데이터(raw data)를 활용할 수 있다는 장점이 있고, 이 방법을 이용하여, 유사 질환에 대해서 부분적으로 설계된 임상 의사결정모형을 통합하고, 유사질환의 감별진단을 위한 통합 임상 의사결정 지원 시스템을 구축할 수 있게 된다.

대 표 도 - 도1



(72) 발명자

박희준

대구광역시 달서구 새동네로 34, 보람타운 102동 502호 (용산동)

서석태

울산광역시 남구 봉월로 133, 103동 1109호 (신정동, 남산포스코더샵아파트)

김민수

대구광역시 수성구 지범로40길 6, 신화맨션 101동 1310호 (범물동)

이 발명을 지원한 국가연구개발사업

과제고유번호	RTI04-01-01
부처명	지식경제부
연구사업명	지방기술혁신사업
연구과제명	u-Health 기반 진단/예측전문가시스템 상용화 기술개발
주관기관	계명대학교
연구기간	2010.07.01 ~ 2011.06.30

특허청구의 범위

청구항 1

(a) 다수의 검사항목별 임상 데이터가 저장된 데이터 베이스로부터 추출된 임상 데이터를 엔트로피 최대화 척도(entropy maximization measure)를 이용하여 서로 다른 2개의 데이터 그룹으로부터 임상적 참조치를 산출하는 단계;

(b) 상기 서로 다른 2개의 그룹 사이의 임상적 차이를 평가하여, 후보 검사항목을 추출하는 단계;

(c) 상기 어느 하나의 그룹에서 산출된 검사항목의 참조치를 기준으로 해당 검사항목의 속성값(attribute values)을 명목형 속성값(nominal attribute values)으로 변환하는 단계;

(d) 상기 (b) 단계에서 추출된 후보 검사항목들 사이에서 감별진단을 위한 중요 파라미터를 추출하는 단계를 포함하는 것을 특징으로 하는 엔트로피 러프 근사화 기술기반 질환별 감별진단을 위한 주요 파라미터 추출방법.

청구항 2

제1항에 있어서,

상기 서로 다른 2개의 데이터 그룹은,

하나의 질환을 갖는 그룹과 다른 질환을 갖는 그룹,

또는 하나의 질환을 갖는 그룹과 나머지 질환을 갖는 그룹인 것을 특징으로 하는 엔트로피 러프 근사화 기술기반 질환별 감별진단을 위한 주요 파라미터 추출방법.

청구항 3

제1항에 있어서,

상기 엔트로피 최대화 척도는,

$$Maximize\ to\ H(T) = H_{R1}(T) + H_{R2}(T),$$

$$where\quad H_{R1}(T) = -\sum_{g=a_{min}}^{T} P_{R1}(g) \cdot \log P_{R1}(g),$$

$$H_{R2}(T) = -\sum_{g=T+1}^{a_{max}} P_{R2}(g) \cdot \log P_{R2}(g),$$

$$P(g) = \sum_{i=a_{min}}^{g} p(i)$$

(여기서, P(g)는 도메인 범위에서의 누적 확률값을 의미하고, HR1(T)과 HR2(T)는 임계값으로, 해당 검사항목의 참조치가 T일때 두 영역 R1과 R2의 엔트로피를 의미하며, H(T)는 이때의 엔트로피 합을 나타낸다.)와 같은 식으로 계산되는 것을 특징으로 하는 엔트로피 러프 근사화 기술기반 질환별 감별진단을 위한 주요 파라미터 추출방법.

청구항 4

제1항에 있어서,

상기 (b) 단계는,

상기 참조치가 한 개인 경우,

하나의 검사항목에 대한 상기 서로 다른 2개의 그룹의 참조치가 서로 다른 경우만을 후보 검사항목을 추출하고,

상기 참조치가 2개인 경우,

어느 하나의 참조치 범위가 다른 하나의 참조치 범위에 포함되지 않는 경우만을 후보 검사항목으로 추출되는 것을 특징으로 하는 엔트로피 러프 근사화 기술기반 질환별 감별진단을 위한 주요 파라미터 추출방법.

청구항 5

제1항에 있어서,

상기 (c) 단계는,

상기 참조치가 한 개인 경우,

상기 참조치를 기준으로 2개의 영역의 검사항목의 값들을 명목화하고,

상기 참조치가 두 개인 경우,

상기 두 개의 참조치를 기준으로, 3개 영역의 검사항목의 값들을 명목화 하는 것을 특징으로 하는 엔트로피 러프 근사화 기술기반 질환별 감별진단을 위한 주요 파라미터 추출방법.

청구항 6

제1항에 있어서,

상기 (d) 단계는,

상기 추출된 후보 검사항목과 상기 검사항목별 상기 명목형 값으로 변환되는 의사결정표를 생성하는 단계;

상기 의사결정표를 근거로 식별가능행렬을 생성하는 단계;

상기 식별가능행렬로부터 식별가능함수를 계산하여 감별진단을 위한 주요 파라미터를 추출하는 단계를 포함하는 것을 특징으로 하는 러프 근사화 기술기반 질환별 감별진단을 위한 주요 파라미터 추출방법.

청구항 7

제6항에 있어서,

상기 식별가능행렬은,

$$(c_{ij}) = a \in A : a(x_i) \neq a(x_j),\ \exists\, i,j,\ \text{for}\ d_i \neq d_j$$

(여기서, 전체 A는 검사항목을 나타내는 입력변수의 전체집합이고, a는 임의의 입력변수이고, x_i는 i 번째 case이고, d_i는 i 번째 출력 속성값으로, 질환을 나타낸다. 그리고, c_{ij} 는 서로 다른 두 case들 사이에서 속성값들의 차이를 보인 입력변수들을 의미하고, N은 전체 case의 수를 의미한다.)와 같은 식으로 생성하는 것을 특징으로 하는 러프 근사화 기술기반 질환별 감별진단을 위한 주요 파라미터 추출방법.

청구항 8

제7항에 있어서,

상기 식별가능함수는,

$$f(A)=\prod_{(x,y)\in U^2}\left(\sum\delta(x,y):(x,y)\in U^2 \text{ and } \delta(x,y)\neq\emptyset\right)$$

(여기서, $\sum\delta(x,y)$ 는 (x,y)원소 내에 포함된 속성값들 사이에 논리합(OR) 연산을 의미하고, $\prod_{(x,y)\in U^2}(\cdot)$ 는 해당 case 내에서 서로 다른 원소들 사이에 논리곱(AND) 연산을 나타낸다.)와 같은 식으로 나타내는 것을 특징으로 하는 러프 근사화 기술기반 질환별 감별진단을 위한 주요 파라미터 추출방법.

청구항 9

제7항에 있어서,

상기 의사결정표에서 상기 어느 하나의 명목형 값이 없거(null)나

알려지지 않은값(unknown value)인 경우에도 대응 가능한 모든 값을 가질 수 있는 것을 특징으로 하는 러프 근사화 기술기반 질환별 감별진단을 위한 주요 파라미터 추출방법.

청구항 10

다수의 검사항목별 임상 데이터를 포함하는 임상정보 데이터 베이스;

상기 임상데이터로부터 임상 전문가에 의해 정의된 질환 정보를 저장하는 데이터 베이스;

상기 제1항 내지 제9항 중 어느 한 항의 방법을 이용한 임상 의사결정 지원모듈;

상기 임상결정 지원모듈로부터 임상 의사결정 지원 정보를 포함하여 생성한 일시적 지식(Temporary Knowledge)을 저장하는 지식 데이터 베이스;

상기 시식 데이터 베이스를 통해 생성된 임상 의사결정 지원 종합정보를 획득할 수 있는 어플리케이션 인터페이스 모듈을 포함하는 것을 특징으로 하는 통합 임상 의사결정지원 시스템.

청구항 11

제10항에 있어서,

상기 임상 의사결정 지원 모듈에서 생성된 정보와, 임상 전문가에 의해 판단된 임상정보를 통해 핵심 지식을 저장하는 핵심 지식 저장 데이터베이스를 더 포함하는 것을 특징으로 하는 통합 임상 의사결정지원 시스템.

청구항 12

제10항에 있어서,

상기 임상 의사결정 지원모듈은,

상기 제1항 내지 제9항의 방법을 이용하는 주요 파라미터 추출 모듈 및 임상 의사결정모형 설계 모듈을 포함하는 것을 특징으로 하는 통합 임상 의사결정지원 시스템.

청구항 13

제12항에 있어서,

상기 임상 의사결정모형 설계 모듈은,

상기 임상정보 데이터베이스로부터 랜덤 N번 샘플링(random sampling)에 의해서 수집된 N개의 실험군 및 대조군 데이터에서 상기 주요 파라미터 추출 방법에서 적용된 1개의 참조치 또는 2개의 참조치에 의해서 결정된 모든 검사항목을 적용하여 트리형 구조로 설계되는 것을 특징으로 하는 통합 임상 의사결정지원 시스템.

명 세 서

기 술 분 야

[0001] 본 발명은 임상 데이터의 파라미터 추출방법 및 이를 이용한 임상 지원 시스템에 관한 것으로, 엔트로피 러프 근사화 기술에 기반한 감별진단을 위한 주요 파라미터 추출방법 및 이를 이용한 통합 임상 의사결정 지원 시스템에 관한 것이다.

배 경 기 술

[0002] 의료 분야에서, 환자의 상태에 대해 알고, 치료하기 위해 이용가능한 여러 가지 상이한 툴들이 존재한다. 전통적으로, 의사는 환자를 물리적으로 검사하고, 수 년간의 연구로부터 수집된 개인 지식의 광대한 어레이를 얻어서, 환자가 겪고 있는 문제점 및 상태를 식별하여, 적절한 치료법을 결정할 것이다.

[0003] 전통적으로, 지원 정보의 소스는 다른 종사자, 참조 책 및 매뉴얼, 비교적 간단한 검사 결과 및 분석 등을 포함한다. 과거 10년 동안, 특히 최근에, 또 다른 참조 물질의 넓은 어레이가 종사자에게 이용가능하게 되어, 이용가능한 자원을 크게 확장시키고, 의료진 진단 및 환자 간호를 개선 및 향상시키게 되었다.

[0004] 진단 자원들 중에서, 의사 및 다른 간병인에게 현재 이용가능한 것은, 처방되고 제어될 수 있는 자원뿐만 아니라 정보의 데이터베이스이다. 데이터베이스는, 다소 통상적인 참조 라이브러리이며, 많은 소스로부터 이용가능하게 알려져 있으며, 의사에게, 가능한 질병 상태에 대한 상세 정보, 그러한 상태를 인식하는 방법에 관한 정보 및 그러한 상태의 치료법을 수초 이내에 제공한다.

[0005] 물론, 유사한 참조 물질이, 약품의 상호 작용, 질병 및 의료 사건에 대한 경향 등과 같은 고려 사항을 식별하는데 이용가능하다. 이들 참조 물질 중 어떤 것은 간호 제공자에게 무료로 이용가능하고, 다른 것은 가입(subscription) 또는 공동 멤버쉽과 전형적으로 관련된다.

[0006] 잠재적인 환자 상태 및 의료 사건을 조사하고, 잠재적인 의료 문제의 소스를 지적하기 위해 규정 및 제어될 수 있는 특정 데이터 획득 기법이 또한 알려져 있다. 전통적인 처방가능 데이터 소스는 간단한 혈액 테스트, 소변 테스트, 물리적 검사의 수기된 결과 등을 포함한다. 최근 수 십년 동안, 신체의 시스템의 동작 및, 어느 정도까지는, 상황 및 자극에 대한 시스템의 반응을 검출 및 기록하는 다양한 유형의 전기적 데이터 획득을 포함하는 보다 정교한 기법이 개발되어 왔다.

[0007] 보다 정교한 시스템이 개발되어 왔는데, 그것은 그러한 시스템이 개발되기 전에는 수술의 개입을 통해서만 보여지고 분석될 수 있었던 내부 특징을 포함하는 신체의 화상을 제공하고, 소정의 다른 방법으로는 볼 수 없었던 다른 특징 및 기능을 보고 분석하는 것을 가능하게 한다. 이들 모든 기법은 의사에게 이용가능한 광대한 자원의 어레이에 부가되어, 의료 간호의 품질을 크게 향상시켰다.

[0008] 의료 관련 정보의 소스에 있어서의 극적인 증가 및 향상에도 불구하고, 테스트 및 데이터의 처방 및 분석, 의료 사건의 진단 및 치료는 여전히, 훈련된 간호 제공자의 전문적 지식에 크게 의존한다. 사람의 경험에 의해 제공된 입력 및 판단은 그러한 상황에서 대체되지 않을 것이며, 대체되어서는 않된다. 그러나, 의료 정보의 소스가 더 향상되고, 통합될 필요가 있다.

[0009] 다소 자동화된 형태로 통지된 진단 및 분석을 허용하면서 시도가 있었지만, 이들 시도는 신속하고 효율적인 환자 간호에서 가장 유용한 통합 및 상관의 레벨에 접근하지는 못했다. 특징 식별 및 분류 알고리즘에 근거하여

의료 데이터를 분석하기 위해, 증가하는 수의 응용들이 개발되고 있다.

[0010] 그러나,그러한 알고리즘은, 알고리즘에서 전형적으로 수행된 분석의 매우 특정한 본질로 인해, 그리고 분석을 위해 액세스 가능한 정보의 제한된 양으로 인해, 오늘날에는 제한된 정도로만 이용되고 있다. 오늘날, 그러한 알고리즘은 특정의 질병 및 촬상 양식으로 상당히 제한된다. 그러한 활동은 때때로, 이용가능한 데이터의 주기적인 분석에 근거하여 인간 프로그래머에 의해 수행된 특정 프로그램 및 프로젝트를 필요로 한다는 점에서, 알고리즘의 추가의 강화 및 향상에는 문제가 있다.

[0011] 또한, 종래에 사용되는 알고리즘 또는 임상지원 진단 프로그램들은 환자의 문진결과에 대한 증상정보만을 활용하여 해당 증상에 대응된 기초정보를 활용하여 관련 질환의 진단 정보를 제공한다는 점에서, 기초 임상정보 데이터가 제한된 한계로 인해 정확도 내지 정밀도가 떨어지는 문제점이 있다.

[0012] 그리고, 수많은 임상정보 데이터를 이용하는 종래의 방법에서는 일반적인 통계적인 분석방법을 사용하는데, 이와 같은 방법은 해당 검사항목의 임상 데이터가 없거나(null value) 알려지지 않은 데이터(unknown value)의 처리방법에 있어서, 해당 검사 항목을 제거하는 '데이터 전처리'과정을 거치거나 null value를 해당 거사항목의 중위수 혹은 평균값으로 데체하는 과정을 수행하는 분석방법을 채택하기 때문에, null value 비율이 큰 경우 '데이터 전처리' 과정에서 손실될 우려가 있으며, 실제 환자가 검사하지 않은 검사항목에 대해서 대표값으로 치환됨으로 데이터의 왜곡 혹은 신뢰성 문제가 발생할 수 있는 문제점이 있다.

발명의 내용

해결하려는 과제

[0013] 상술한 문제를 해결하기 위한 본 발명의 과제는 데이터의 왜곡 혹은 신뢰성 문제가 발생할 수 있는 전처리 과정의 사용 없이 수집된 임상검사결과의 원 데이터(raw data)를 활용할 수 있게 하고자 함이고, 이를 이용한 임상 의사결정 지원 시스템을 제공하여 특정 질환에 대한 임상 의사결정모형뿐만 아니라, 유사 질환에 대해서 부분적으로 설계된 임상 의사결정모형을 통합하고, 임상 지식을 DB화 함으로써, 유사질환의 감별진단을 위한 통합 임상 의사결정 지원 시스템을 구축하고자 함이다.

과제의 해결 수단

[0014] 상술한 문제를 해결하기 위한 본 발명의 제1 특징은 (a) 다수의 검사항목별 임상 데이터가 저장된 데이터 베이스로부터 추출된 임상 데이터를 엔트로피 최대화 척도(entropy maximization measure)를 이용하여 서로 다른 2개의 데이터 그룹으로부터 임상적 참조치를 산출하는 단계; (b) 상기 서로 다른 2개의 그룹 사이의 임상적 차이를 평가하여, 후보 검사항목을 추출하는 단계; (c) 상기 어느 하나의 그룹에서 산출된 검사항목의 참조치를 기준으로 해당 검사항목의 속성값(attribute values)을 명목형 속성값(nominal attribute values)으로 변환하는 단계; (d) 상기 (b) 단계에서 추출된 후보 검사항목들 사이에서 감별진단을 위한 중요 파라미터를 추출하는 단계를 포함한다.

[0015] 여기서, 상기 서로 다른 2개의 데이터 그룹은, 하나의 질환을 갖는 그룹과 다른 질환을 갖는 그룹, 또는 하나의 질환을 갖는 그룹과 나머지 질환을 갖는 그룹인 것이 바람직하고, 상기 엔트로피 최대화 척도는,

$$Maximize\ to\ H(T) = H_{R1}(T) + H_{R2}(T),$$

$$where\ \ H_{R1}(T) = -\sum_{g=a_{min}}^{T} P_{R1}(g) \cdot log\ P_{R1}(g),$$

$$H_{R2}(T) = -\sum_{g=T+1}^{a_{max}} P_{R2}(g) \cdot log P_{R2}(g),$$

$$P(g) = \sum_{i=a_{min}}^{g} p(i)$$

[0016]

[0017] (여기서, P(g)는 도메인 범위에서의 누적 확률값을 의미하고, HR1(T)과 HR2(T)는 임계값으로, 해당 검사항목의 참조치가 T일때 두 영역 R1과 R2의 엔트로피를 의미하며, H(T)는 이 때의 엔트로피 합을 나타낸다.)와 같은 식으로 계산되는 것이 바람직하다.

[0018] 또한, 바람직하게는 상기 (b) 단계는, 상기 참조치가 한 개인 경우, 하나의 검사항목에 대한 상기 서로 다른 2개의 그룹의 참조치가 서로 다른 경우만을 후보 검사항목을 추출하고, 상기 참조치가 2개인 경우, 어느 하나의 참조치 범위가 다른 하나의 참조치 범위에 포함되지 않는 경우만을 후보 검사항목으로 추출되는 것이 바람직하고, 상기 (c) 단계는, 상기 참조치가 한 개인 경우, 상기 참조치를 기준으로 2개의 영역의 검사항목의 값들을 명목화하고, 상기 참조치가 두 개인 경우, 상기 두 개의 참조치를 기준으로, 3개 영역의 검사항목의 값들을 명목화 하는 것이 바람직하다.

[0019] 더하여, 상기 (d) 단계는, 상기 추출된 후보 검사항목과 상기 검사항목별 상기 명목형 값으로 변환되는 의사결정표를 생성하는 단계; 상기 의사결정표를 근거로 식별가능행렬을 생성하는 단계; 상기 식별가능행렬로부터 식별가능함수를 계산하여 감별진단을 위한 주요 파라미터를 추출하는 단계를 포함하는 것이 바람직하고, 상기 식별가능행렬은,

$$(c_{ij}) = a \in A : a(x_i) \neq a(x_j),\ \exists\, i,j,\ \text{for}\ d_i \neq d_j$$

[0020]

[0021] (여기서, 전체 A는 검사항목을 나타내는 입력변수의 전체집합이고, a는 임의의 입력변수이고, x_i는 i 번째 case이고, d_i는 i 번째 출력 속성값으로, 질환을 나타낸다. 그리고, c_{ij} 는 서로 다른 두 case들 사이에서 속성값들의 차이를 보인 입력변수들을 의미하고, N은 전체 case의 수를 의미한다.)와 같은 식으로 생성하는 것이 바람직하다.

[0022] 그리고, 바람직하게는 상기 식별가능함수는,

$$f(A) = \prod_{(x,y) \in U^2} \sum \delta(x,y) : (x,y) \in U^2 \text{ and } \delta(x,y) \neq \emptyset$$

[0023]

[0024] (여기서, $\sum \delta(x,y)$ 는 (x,y)원소 내에 포함된 속성값들 사이에 논리합(OR) 연산을 의미하고, $\prod_{(x,y) \in U^2} (\cdot)$ 는 해당 case 내에서 서로 다른 원소들 사이에 논리곱(AND) 연산을 나타낸다.)와 같은 식으로 나타내는 것이 바람직하고, 상기 의사결정표에서 상기 어느 하나의 명목형 값이 없거(null)나 알려지지 않은값(unknown value)인 경우에도 대응 가능한 모든 값을 가질 수 있는 것이 바람직하다.

[0025] 그리고, 본 발명의 제2 특징은 통합 임상 의사결정지원 시스템으로서, 다수의 검사항목별 임상 데이터를 포함하는 임상정보 데이터 베이스; 상기 임상데이터로부터 임상 전문가에 의해 정의된 질환 정보를 저장하는 데이터 베이스; 상기 제1항 내지 제9항 중 어느 한 항의 방법을 이용한 임상 의사결정 지원모듈; 상기 임상결정 지원모

듈로부터 임상 의사결정 지원 정보를 포함하여 생성한 일시적 지식(Temporary Knowledge)을 저장하는 지식 데이터 베이스; 상기 지식 데이터 베이스를 통해 생성된 임상 의사결정 지원 종합정보를 획득할 수 있는 어플리케이션 인터페이스 모듈을 포함한다.

[0026] 여기서, 상기 임상 의사결정 지원 모듈에서 생성된 정보와, 임상 전문가에 의해 판단된 임상정보를 통해 핵심 지식을 저장하는 핵심 지식 저장 데이터베이스를 더 포함하는 것이 바람직하고, 상기 임상 의사결정 지원모듈은, 상기 제1항 내지 제9항의 방법을 이용하는 주요 파라미터 추출 모듈 및 임상 의사결정모형 설계 모듈을 포함하는 것이 바람직하다.

[0027] 또한, 바람직하게는 상기 임상 의사결정모형 설계 모듈은, 상기 임상정보 데이터베이스로부터 랜덤 N번 샘플링(random sampling)에 의해서 수집된 N개의 실험군 및 대조군 데이터에서 상기 주요 파라미터 추출 방법에서 적용된 1개의 참조치 또는 2개의 참조치에 의해서 결정된 모든 검사항목을 적용하여 트리형 구조로 설계되는 것일 수 있다.

발명의 효과

[0028] 이와 같은 본 발명에 따른 주요 파라미터 추출방법을 이용하면, 데이터의 왜곡 혹은 신뢰성 문제가 발생할 수 있는 전처리 과정의 사용 없이 수집된 임상검사결과의 원 데이터(raw data)를 활용할 수 있다는 장점이 있고, 다양한 전처리 방법 및 응용분야에 활용될 수 있다.

[0029] 또한, 상기 추출방법을 이용한 본 발명의 통합 임상 의사결정 지원 시스템은 특정 질환에 대한 임상 의사결정 모형뿐만 아니라, 유사 질환에 대해서 부분적으로 설계된 임상 의사결정모형을 통합하고, 임상 지식을 DB화 함으로써, 유사질환의 감별진단을 위한 통합 임상 의사결정 지원 시스템을 구축할 수 있게 된다.

[0030] 더하여, 병원 내 분과별 인턴 혹은 레지던트들의 교육 및 학습용 콘텐츠 제작시 효과적으로 사용할 수 있게 된다.

도면의 간단한 설명

[0031] 도 1은 본 발명에 따른 엔트로피 러프 근사화 기술기반 질환별 감별진단을 위한 주요 파라미터 추출방법의 흐름도를 나타낸 도면,

도 2는 본 발명에 따른 엔트로피 러프 근사화 기술기반 질환별 감별진단을 위한 주요 파라미터 추출방법을 적용하기 위한, 입원환자에 대한 기본 검사항목 중 Total Bilirubin (단위, mg/dL) 검사항목에 대해서 심부전 환자군과 비심인성 호흡곤란 환자군의 검사결과를 나타내고,

도 3은 본 발명에서 적용되는 엔트로피 최대화 척도에 의해서 결정된 검사항목 Total Bilirubin의 참조치를 나타내는 그래프이고,

도 4는 본 발명에 따른 주요 파라미터 추출방법의 단계로서, 명목형 변환 방법의 모식도를 나타낸 도면이고,

도 5는 본 발명에 따른 또 다른 실시예로서, 통합 임상 의사결정지원 시스템의 구성을 나타낸 도면이고,

도 6은 본 발명에 따른 통합 임상 의사결정 지원 시스템에 적용되는 의사결정모형과 종래의 의사결정 모형의 예를 나타내는 모형도이다.

발명을 실시하기 위한 구체적인 내용

[0032] 이하에서 본 발명의 바람직한 실시예를 도면을 참조하여 상세히 설명하기로 한다.

[0033] 감별진단(Differential Diagnosis)은 증상(symptom)의 특징에서 어떤 병명을 생각할 수 있을 때, 그것과 유사한 특징을 가진 다른 질병을 모두 고려하여 상호 간의 비교 및 검토 후 처음에 생각한 병명과 다름이 없는가의 여부를 알기 위한 진단법을 의미함. 예를 들어, 고열, 흉통, 기침, 가래 등의 증상이나 진찰소견, 임상검사 소견 등으로 폐렴이라고 생각할 수 있을 경우에, 이와 비슷한 특징을 가진 질병이 감별진단에서 문제가 되는데, 이 경우 유행성 감기, 급성기관지염, 폐농양, 폐결핵의 급성형, 늑막염 등이 그것이다.

[0034] 그러나 이러한 병은 부분적으로 폐렴의 특징과 공통되는 점도 있으나 발병이나 경과의 상태, 폐의 병적 변화의 유무, X선 소견, 세균학적 검사 소견 등에서 서로 다른 특징을 나타내므로 다른 병으로 구별될 수 있으므로, 임

상적 측면에서 질환별 감별진단은 아주 중요한 문제이면서 어려운 문제라는 점에서 감별진단을 위한 주표 파라미터 추출방법과 이를 이용한 임상 의사결정 지원 시스템을 본 발명에서 제안하고자 한다.

[0035] 도 1은 본 발명에 따른 엔트로피 러프 근사화 기술기반 질환별 감별진단을 위한 주요 파라미터 추출방법의 흐름도를 나타낸 도면이다. 도 1에 나타낸 바와 같이, 본 발명의 주요 파라미터 추출방법은 (a) 다수의 검사항목별 임상 데이터가 저장된 데이터 베이스로부터 추출된 임상 데이터를 엔트로피 최대화 척도(entropy maximization measure)를 이용하여 서로 다른 2개의 데이터 그룹으로부터 임상적 참조치를 산출하는 단계(S100); (b) 상기 서로 다른 2개의 그룹 사이의 임상적 차이를 평가하여, 후보 검사항목을 추출하는 단계(S200); (c) 상기 어느 하나의 그룹에서 산출된 검사항목의 참조치를 기준으로 해당 검사항목의 속성값(attribute values)을 명목형 속성값(nominal attribute values)으로 변환하는 단계(S300); (d) 상기 (b) 단계에서 추출된 후보 검사항목들 사이에서 감별진단을 위한 중요 파라미터를 추출하는 단계(S400)를 포함한다.

[0036] 먼저 다수의 검사항목별 임상 데이터가 저장된 데이터 베이스로부터 추출된 임상 데이터를 엔트로피 최대화 척도 (entropy maximization measure)를 이용하여 서로 다른 두 그룹들로부터 임상적 검사결과의 참조치를 산출(Calculate references of clinical laboratory tests from two different groups) 한다.(S100)

[0037] 여기서 '서로 다른 두 그룹' 은 질환 A와 질환 B가 될 수도 있으며, 또는 질환 A를 가진 군(abnormal group)과 그렇지 않은 정상군(normal group)이 될 수도 있다. 즉, 서로 다른 질환을 갖는 환자의 임상 데이터일 수 있고, 질환을 갖는 환자(abnormal group)와 질환을 갖지 않은자(normal group)의 임상 데이터 그룹으로 나눌 수 있음을 의미한다.

[0038] 예를 들어, 병원정보시스템(HIS: Hospital Information System)으로부터 추출된 "Data Mart"(임상전문가에 의해 질환이 정의된 임상 데이터 베이스)가 특정질환, 즉 '급성 심부전 증상을 가진 환자군(I50)과 병원 내원 당시 호흡곤란 증상으로 내원 하였으나 임상적으로 심부전 환자의 소견을 보이지 않은 '비심인성 호흡곤란 환자군'(No)으로 구성될 수 있다. 여기서 I50은 '국제질병분류코드 ICD-10'에서 지정된 질병분류코드를 의미하며, '비심인성 호흡곤란 환자군은 특정 질병으로 분류되지 않음으로 ' No'로 표기된다. 또한 병원정보시스템으로부터 추출된 " Data Mart"는 다음과 같은 임상검사항목들로 정의되어 있다.

[0039] 전혈구 검사(CBC & Differential Count), 프로트롬빈 시간(Prothrombin Time, PT), 활성화 부분 트롬보플라스틴 시간(Activated Partial Thromboplastin Time, APTT), 혈청 전해질(Serum Electrolytes), 입원환자에 대한 기본 검사 (Routine Admission), 아밀라제(Amylase), 동맥혈 가스 분석(Blood pH and Gas), 리파아제(Lipase), CK-MB, Troponin-I, CK, LDH, CRP, Fibrinogen, Ca^{2+}, Mg^{2+}, Pro BNP 등이 있다.

[0040] 도 2는 본 발명에 따른 엔트로피 러프 근사화 기술기반 질환별 감별진단을 위한 주요 파라미터 추출방법을 적용하기 위한, 입원환자에 대한 기본 검사항목 중 Total Bilirubin (단위, mg/dL) 검사항목에 대해서 심부전 환자군과 비심인성 호흡곤란 환자군의 검사결과를 나타낸다.

[0041] 여기서 도 2의 (a)는 Total Bilirubin 검사항목의 검사결과를 나타낸 표이고, 도 2의 (b)는 심부전 환자에서 Total Bilirubin 검사결과의 빈도 분포를 나타낸 그래프이고, 도 2의 (c)는 비심인성 호흡곤란 환자군에서 Total Bilirubin 검사 결과의 빈도 분포를 나타낸 그래프이다.

[0042] 도 2의 (a)에서 'Attribute values'는 Total Bilirubin 검사항목의 결과값들을 의미하고, ' CHF'는 심부전 (Congestive Heart Failure) 환자군을, 이때의 각 행의 값은 해당 속성값(attribute value)을 가진 환자의 수를 의미하고, 'Non-C.D'는 비심인성 호흡곤란 (Non-Cardiogenic Dyspnea) 군을 나타내며, 이때의 각 행의 값은 해당 속성값(attribute value)을 가진 환자의 수를 나타낸다.

[0043] 또한 그림 2(b)와 (c)는 각각 심부전 환자군에서 해당 속성값(attribute value)을 가진 환자들의 분포와 비심인성 호흡곤란을 가진 군에서 해당 속성값(attribute value)을 가진 환자들의 분포를 나타낸다. 도 2의 (b)와 (c)의 분포를 근거로, 엔트로피 최대화 척도를 이용하여 검사항목 Total Bilirubin에 대한 각 그룹의 임상적 참조치를 계산하면 도 3과 같은 결과를 얻을 수 있다. 도 3은 본 발명에서 적용되는 엔트로피 최대화 척도에 의해서 결정된 검사항목 Total Bilirubin의 참조치를 나타내는 그래프이다.

[0044] [수학식 1]은 본 발명에 따른 주요 파라미터 추출방법에 적용되는 엔트로피 최대화 척도를 나타내는 수학식이다.

수학식 1

$$Maximize\ to\ H(T) = H_{R1}(T) + H_{R2}(T),$$

$$where\quad H_{R1}(T) = -\sum_{g=a_{min}}^{T} P_{R1}(g)\cdot \log P_{R1}(g),$$

$$H_{R2}(T) = -\sum_{g=T+1}^{a_{max}} P_{R2}(g)\cdot \log P_{R2}(g),$$

$$P(g) = \sum_{i=a_{min}}^{g} p(i)$$

[0045]

[0046] 여기서, 해당 검사항목의 도메인 범위가 a_{min} ~ a_{max} (즉 도 2의 (a)에서 0.2 ~ 4.5)일 때, P(g)는 도메인 범위의 최소값 0.2에서 g까지의 누적 확률 값 (cumulative probability value)을 나타내고, $H_{R1}(T)$과 $H_{R2}(T)$는 임계값 (threshold value), 즉 해당 검사항목의 참조치가 T일 때 두 영역 R1과 R2의 엔트로피(entropy)를 의미하며, 이때의 엔트로피의 합을 H(T)로 표현하며, 검사항목 값 g가 a_{min}에서 a_{max}까지 변화될 때 최대 엔트로피의 값을 가진 임계값 T가 해당 검사항목의 참조치가 된다.

[0047] 이와 같은 방법으로 결정된 검사항목 Total Bilirubin에서 심부전 환자군과 비심인성 호흡곤란 환자군의 참조치는 각각 도 3의 (a)와 (b)과 같다. 도 3의 (a)는 심부전 환자군에서 Total Bilirubin의 참조치, (b) 비심인성 호흡곤란 환자군에서 Total Bilirubin의 참조치를 나타낸다.

[0048] 도 3의 (a)와 (b)에서 심부전 환자군과 비심인성 호흡곤란 환자군에서 Total Bilirubin의 임상적 참조치가 각각 0.8과 0.6임을 알 수 있으며, 심부전 환자군의 참조치가 비심인성 호흡곤란 환자군에 비해서 상대적으로 높은값을 가짐을 알 수 있다.

[0049] 본 발명에서는 각 검사항목의 임상적 참조치를 한 개(T) 그리고 두 개인 경우 (T_1과 T_2)를 모두 추출하였는데, 이는 임상적 측면에서는 임상검사 항목의 참고치를 특정 검사항목에서는 한 개, 대부분의 검사항목에서는 두 개인 경우가 존재하기 때문이다. 만약 해당 검사항목의 참조치를 두 개로 결정할 경우, [수학식 1]에서 H_{R1}, H_{R2}, H_{R3}과 같이 3개의 영역으로 분할되어야 한다.

[0050] 2번째 단계는 서로 다른 두 그룹 사이에 임상적 차이 (즉 참조치의 변화)를 평가(Evaluate clinical difference (i.e., variation of reference values) between the two groups.)<참조치의 평가 방법> 하는 단계이다.(S200) 임상 검사의 참조치가 한 개인 경우, 검사항목 Total Bilirubin에 대해서 심부전 환자군의 참조치 CHF = α이고, 비심인성 호흡곤란 환자군의 참조치 Non-C.D = β라고 할 때, 만약 α = β이면, 해당 검사항목 Total Bilirubin은 이들 두 환자군에서 차이를 보이지 않으므로 제거되고, 참조치가 서로 다른 값을 가진다면 감별진단을 위한 후보 검사항목으로 남겨둔다.

[0051] 임상 검사의 참조치가 두 개인 경우, 위의 경우와 유사한 방법으로, 심부전 환자군의 참조치 CHF = [α, β](a_{min} ≤ α ≤ β ≤ a_{max})이고, 비심인성 호흡곤란 환자군의 참조치 Non-C.D = [γ, δ](a_{min} ≤ γ ≤ δ ≤ a_{max})이라고 할 때, 만약 조건을 만족한다면, 다시 말해서 심부전 환자군의 참조치의 하한과 상한값이 비심인성 호흡곤란 환자

군의 참조치의 하한과 상한값의 범위 내에 포함될 때, 해당 검사항목 Total Bilirubin은 제거되고, 그렇지 않으면 감별진단을 위한 후보 검사항목으로 남겨둔다.

[0052] 이러한 방법으로 본 발명에서는 임상 검사의 참조치가 한 개 혹은 두 개일 때의 모든 가능한 후보 검사항목들을 추출한다. 도 2에서 본 발명에 따른 주요 파라미터 추출방법의 (1)단계 및 (2)단계를 를 나타내는 "엔트로피 최대화 척도 (첫번째 필터링 방법)"이 이에 해당된다.

[0053] 세번째 단계는 계산된 정상군의 각 검사항목의 참조치를 기준으로 해당 검사항목의 속성값 (attribute values)을 명목형 속성값으로 변환(Convert mixed attribute values into nominal attribute values using the reference of normal group obtained from (1))는 단계이다.(S300)

[0054] 도 4는 본 발명에 따른 주요 파라미터 추출방법의 과정으로서, 명목형 변환 방법의 모식도를 나타낸 도면이다. 도 4의 (a)는 임상검사의 참조치가 한 개인 경우, 해당 검사항목 별 명목형 속성값 변환하는 방법의 모식도이고, 도 4의 (b)는 임상검사의 참조치가 두 개인 경우, 해당 검사항목 별 명목형 속성값 변환하는 방법의 모식도이다.

[0055] 도 4에 나타낸 바와 같이, 만약 임상검사의 참조치가 한 개의 값(0.6)으로 결정된다면 그 참조치를 기준으로 해당 검사항목을 두 개의 부분 공간 (즉 normal과 abnormal)으로 분할하고, 검사항목의 값들을 normal과 abnormal로 변경한다.(도 4의 (a))

[0056] 참조치가 한 개의 경우와 유사한 방법으로, 만약 비심인성 호흡곤란 환자군에 대한 Total Bilirubin의 검사항목에서 두 개의 참조치가 각각 {0.6, 1.4}로 결정되었다면, 0.2 ~ 0.6은 lower abnormal, 0.6 ~ 1.4는 normal, 1.4 ~ 4.5는 upper abnormal과 같이 해당 검사항목을 3개의 부분 공간으로 분할한 후, 검사항목 값들을 명목화 시킨다.(도 4의 (b))

[0057] 네번째 단계는, 추출된 후보 검사항목들 사이에서 감별진단을 위한 중요 파라미터를 추출하는 단계로서, 러프집합의 근사화 척도를 이용하여 두번째 단계에 의해서 필터링된 후보 검사항목들 사이에서 감별진단을 위한 중요 파라미터들을 추출하는 단계이다.(S400)

[0058] 두번째 단계에 의해서 추출된 후보 검사항목들과 이때에 명목형 값으로 변환된 의사결정표를 다음과 같다고 가정한다.

표 1

[0059]

Case	Input variable					Output variable
	WBC	RBC	Total Bilirubin	Troponin I	Pro BNP	
1	U_abnormal	U_abnormal	U_abnormal	L_abnormal	Normal	I50
2	U_abnormal	-	U_abnormal	U_abnormal	U_abnormal	I50
3	Normal	Normal	L_abnormal	L_abnormal	-	No
4	Normal	Normal	U_abnormal	-	-	I50
5	-	-	-	U_abnormal	Normal	No
6	Normal	Normal	-	-	Normal	No

[0060] [표 1]에서 WBC (White Blood Cell), RBC (Red Blood Cell), Total Bilirubin, Troponin I, Pro BNP는 입력변수 (Input variable), 즉 검사항목을, 'Output variable'은 심부전 (CHF) 환자군과 비심인성 호흡곤란 환자군 (No)을 나타낸다.

[0061] 입력변수 WBC, RBC, Total Bilirubin, Troponin I, Pro BNP의 값들 중에서, 'U_abnormal', 'L_abnormal'은 각각 Upper Abnormal과 Lower Abnormal을 의미한다. (도 4 참조) 또한 Case 2 ~ 6에서 '-' 은 null value 혹은 unknown value를 나타내고, 이들 값은 병원 내원 시 검사하지 않은 임상검사항목을 의미한다. 다시 말해서, 대부분의 환자들은 병원에 내원 시 해당 진료과에서 필요한 부분에 대한 임상 검사만을 실시하기 때문에 이러한 null 또는 unknown value들은 항상 존재한다.

[0062] 그리고, [표 1]의 의사결정표를 근거로 다음과 같은 [수학식 2]를 이용하여 식별가능행렬 (discernibility matrix)을 구성한다.

수학식 2

[0063] $$(c_{ij}) = a \in A : a(x_i) \neq a(x_j),\ \exists\, i,j,\ \text{for}\ d_i \neq d_j$$

[0064] 여기서 A는 [표 1]에서 입력변수 {WBC, RBC, Total Bilirubin, Troponin I, Pro BNP}의 전체집합을 의미하고, a는 전체 입력변수의 집합에서 임의의 입력변수를 나타낸다. 또한 x_i와 x_j는 i-번째와 j-번째 case를 나타내고, d_i와 d_j는 각각 i-번째와 j-번째의 출력 속성값 (즉 I50 혹은 No)을 의미한다.

[0065] 이때 $\{a \in A : a(x_i) \neq a(x_j)\}$의 의미하는 바는 a가 WBC라고 할 때, i-번째 case와 j-번째 case에서 서로 다른 값을 가진 변수(즉 속성)를 의미한다. 따라서 c_{ij} (i, j = 1,2,...,N)는 서로 다른 두 case들 사이에서 속성값들의 차이를 보인 입력변수들을 의미하고, N은 전체 case의 수를 의미한다.

[0066] 또한 본 발명에서는 null 혹은 unknown value를 나타내는 '-' 을 통계적인 어떠한 전처리 과정 없이 사용하기 위해서 'don't care' 조건으로 정의하여 사용함. (여기서 ' don't care' 조건은 해당 null or unknown value가 대응 가능한 모든 값을 가질 수 있음을 의미한다.)

[0067] 즉, 이와 같은 처리 방법은 일반적으로 특정 질환의 감별진단 시 해당 검사항목의 null value의 비율이 큰 경우 '데이터 전처리' 과정에서 손실될 우려가 있으며, 실제 환자가 검사하지 않은 검사항목에 대해소 대표값으로 치환됨으로써 데이터의 왜곡 혹은 신뢰성 문제가 발생할 수 있기 때문에 , 상기 전처리 과정의 사용없이 수집된 임상검사결과의 원 데이터(raw data)를 활용할 수 있다는 장점을 가지며, 다양한 전처리 방법 및 응용분야에 활용될 수 있다.

표 2

[0068]

Case	1	2	3	4	5	6
1			WBC, RBC, Total Bilirubin		Troponin I	WBC, RBC
2			WBC, Total Bilirubin, Troponin I		Pro BNP	WBC, Pro BNP
3	WBC, RBC, Total Bilirubin	WBC, Total Bilirubin, Troponin I		Total Bilirubin		
4			Total Bilirubin			
5	Troponin I	Pro BNP				
6	WBC, RBC	WBC, Pro BNP				

[0069] [수학식 2]의 식별가능행렬의 정의에 따라 c_{ij}는 전체 case의 수가 6 ([표 1] 참조)이므로, 6 × 6 행렬이 구성되며, 대각행렬 ({1,1}, {2,2}, {3,3}, {4,4}, {5,5}, {6,6})을 기준으로 상삼각 행렬과 하삼각 행렬은 서로

대칭인 구조를 가지고, 빈칸(□)은 동일한 출력 속성값 (즉 I50 혹은 No 간의 비교)을 가지거나 (ii) 서로 다른 출력 속성값에 대해서 동일한 입력변수의 명목값을 가진 경우를 나타낸다.

[0070] 다시 말해서, 동일한 출력 속성값을 가지는 경우는 대각행렬 이외에 {(1,2), (1,4), (2,4), (2,1), (4,1), (4,2)}, {(3,5), (3,6), (5,6), (5,3), (6,3), (6,5)}가 되며, (ii) 서로 다른 출력 속성값에 대해서 동일한 입력변수의 값을 가지는 경우는 {(4,5), (4,6), (5,4), (6,4)}에 해당된다.

[0071] [표 2]의 식별가능행렬로부터 전제 Case에 대한 식별가능함수 다음과 같은 [수학식 3]을 계산하고, 감별진단을 위한 주요 파라미터 (즉 중요 검사항목들의 리스트)들을 추출한다.

수학식 3

[0072] $$f(A)=\prod_{(x,y)\in U^2}\left(\sum\delta(x,y):(x,y)\in U^2 \text{ and } \delta(x,y)\neq\emptyset\right)$$

[0073] 여기서 δ(x,y)≠φ은 [표 2]의 식별가능행렬에서 빈칸(□)을 제외한 전체 원소들을 의미하고, $\sum\delta(x,y)$는 (x,y)원소 내에 포함된 속성값들 사이에 '논리합 (OR) 연산'을 의미하고, $\prod_{(x,y)\in U^2}(\cdot)$는 해당 case 내에서 서로 다른 원소들 사이에 '논리곱 (AND) 연산'을 나타낸다. 이는 식별가능함수를 불 대수 (Boolean algebra)에서 논리곱 정규형 (conjunctive normal form)으로 표현하는 것과 동일한 방법이다.

[0074] 위의 [표 2]의 식별가능행렬로부터 다음과 같은 식별가능함수 f(A)를 구성할 수 있으며, 불 대수의 2가지 법칙, 즉 분배법칙 (distributive law)과 흡수법칙 (absorptive law)을 이용하여 간소화된 최종 식을 도출할 수 있다.

[0075] f(A) = (WBC + RBC + Total Bilirubin) * Troponin I * (WBC + RBC) * (WBC + Total Bilirubin + Troponin I) * Pro BNP * (WBC + Pro BNP) * Total Bilirubin

[0076] = (WBC + RBC) * Troponin I * Pro BNP * Total Bilirubin

[0077] = WBC * Total Bilirubin * Troponin I * Pro BNP + RBC * Total Bilirubin * Troponin I * Pro BNP

[0078] a) (WBC+RBC+Total Bilirubin) * (WBC+RBC) = (WBC+RBC) <= 흡수법칙

[0079] b) (WBC+Total Bilirubin+Troponin I) * Troponin I = Troponin I <= 흡수법칙

[0080] c) (WBC+Pro BNP) * Pro BNP = Pro BNP <= 흡수법칙= A

[0081] 따라서 최종 식별가능함수 식 f(A) = WBC * Total Bilirubin * Troponin I * Pro BNP + RBC * Total Bilirubin * Troponin I * Pro BNP가 됨을 알 수 있으며, 감별진단을 위해서 2가지 형태의 주요 파라미터 (즉 중요 검사항목의 리스트)를 도출할 수 있다.

[0082] ①첫 번째 주요 파라미터 : {WBC, Total Bilirubin, Troponin I, Pro BNP}

[0083] ②두 번째 주요 파라미터 : {RBC, Total Bilirubin, Troponin I, Pro BNP}

[0084] 위의 2가지 주요 파라미터의 집합들 사이에서 "Total Bilirubin, Troponin I, Pro BNP"는 심부전 환자군과 비심인성 호흡곤란 환자의 감별진단에 있어서 필요 불가결한 검사항목임을 알 수 있다.

[0085] 따라서 본 발명에서는 파라미터의 길이가 최소인 한 개의 주요 파라미터 집합을 추출함으로써 최종 중요 검사항

목들의 집합을 선택한다. 그리고, 위의 예에서처럼, 파라미터의 길이가 동일한 주요 파라미터가 2개 이상인 경우에는 임의로 한 가지 중요 파라미터의 집합을 선택하여 최종 중요 검사항목을 선별할 수 있다.

[0086] 도 5는 본 발명에 따른 또 다른 실시예로서, 통합 임상 의사결정지원 시스템의 구성을 나타낸 도면이다. 도 5에 나타낸 바와 같이, 본 발명의 임상 의사결정지원 시스템은 병원정보시스템(HIS)에서 추출된 다수의 검사항목별 임상 데이터를 포함하는 임상정보 데이터 베이스(10); 상기 임상데이터로부터 임상 전문가에 의해 정의된 질환정보를 저장하는 데이터 베이스(20); 상술한 상기 엔트로피 러프 근사화 기술기반 질환별 감별진단을 위한 주요 파라미터 추출방법을 이용한 임상 의사결정 지원모듈(30); 상기 임상결정 지원모듈로부터 임상 의사결정 지원정보를 포함하여 생성한 일시적 지식(Temporary Knowledge)을 저장하는 지식 데이터 베이스(60); 상기 시식 데이터 베이스를 통해 생성된 임상 의사결정 지원 종합정보를 획득할 수 있는 어플리케이션 인터페이스 모듈(70)을 포함한다.

[0087] 여기서 임상 의사결정 지원모듈은 의사결정지원모델을 포함하는 것으로, 본 발명의 실시예에서는 심부전 환자군의 의사결정지원모형의 설계방법에 대해 이하에서 설명하기로 한다. (대상: 심부전 vs. 비심인선 호흡곤란 환자군)

[0088] [표 3]은 심부전 환자군과 비심인성 호흡곤란 환자군의 기본 임상 특성 (72가지 임상검사 항목)을 나타낸 표이다.

표 3

Variables	CHF (N = 71)	Non-CD (N = 88)	P value
Age, (yrs)	72.39 ± 9.36	65.13 ± 15.14	< 0.000
Gender, n (%)	M: 26 (36.6); F: 45 (63.4)	M: 49 (55.7); F: 39 (44.3)	< 0.017
Urinary tests			
Color, n (%)	Amber: 3 (4.2); Straw: 68 (95.8)	Amber: 3 (3.4); Straw: 85 (96.6)	0.788
S.G.,	1.02 ± 0.01	1.02 ± 0.01	0.719
pH,	6.08 ± 0.94	6.14 ± 0.87	0.030
Albumin, n (%)	Neg.: 43 (63.4); Pos.: 26 (36.6)	Neg.: 68 (77.3); Pos.: 20 (22.7)	0.055
Glucose, n (%)	Neg.: 55 (77.5); Pos.: 16 (22.5)	Neg.: 64 (72.7); Pos.: 24 (27.3)	0.494
Ketone, n (%)	Neg.: 65 (91.5); Pos.: 6 (8.5)	Neg.: 79 (89.8); Pos.: 9 (10.2)	0.703
O.B., n (%)	Neg.: 33 (46.5); Pos.: 38 (53.5)	Neg.: 55 (62.5); Pos.: 33 (37.5)	< 0.043
Urobilinogen,	0.30 ± 0.80	0.40 ± 1.18	0.545
Bilirubin, n (%)	Neg.: 66 (93.0); Pos.: 5 (7.0)	Neg.: 82 (93.2); Pos.: 6 (6.8)	0.956
Nitrite, n (%)	Neg.: 67 (94.4); Pos.: 4 (5.6)	Neg.: 85 (96.6); Pos.: 3 (3.4)	0.497
WBC1, n (%)	Neg.: 51 (71.8); Pos.: 20 (28.2)	Neg.: 72 (81.8); Pos.: 16 (18.2)	0.135
RBC, n (%)	Neg.: 13 (18.3); Pos.: 58 (81.7)	Neg.: 34 (38.6); Pos.: 54 (61.4)	< 0.005
WBC2, n (%)	Neg.: 3 (4.2); Pos.: 68 (95.8)	Neg.: 9 (10.2); Pos.: 79 (89.8)	0.154
Ep. Cell, n (%)	Neg.: 16 (22.5); Pos.: 55 (77.5)	Neg.: 19 (21.6); Pos.: 69 (78.4)	0.836
Cast, n (%)	Neg.: 71 (100.0); Pos.: 0 (0.0)	Neg.: 88 (100.0); Pos.: 0 (0.0)	-
Other, n (%)	Neg.: 67 (94.4); Pos.: 4 (5.6)	Neg.: 84 (95.5); Pos.: 4 (4.5)	0.755
Crystal, n (%)	Neg.: 71 (100.0); Pos.: 0 (0.0)	Neg.: 87 (98.9); Pos.: 1 (1.1)	0.368
CBC			
WBC, $\times 10^3/\mu L$	9.12 ± 3.62	9.42 ± 4.12	0.633
RBC, $\times 10^6/\mu L$	3.98 ± 0.63	4.31 ± 0.63	< 0.025
HGB, g/dL	12.16 ± 2.19	13.67 ± 2.01	< 0.030
HCT, %	36.31 ± 6.41	37.63 ± 5.68	0.210
MCV, fl	91.23 ± 5.88	89.34 ± 5.50	< 0.040
MCH, pg	30.87 ± 2.42	30.79 ± 2.12	0.745
MCHC, g/dL	33.73 ± 1.21	34.52 ± 1.13	< 0.000
PLT, $\times 10^3/\mu L$	255.39 ± 109.80	281.06 ± 103.03	0.134
NEUT, %	78.08 ± 11.77	71.98 ± 14.44	0.800
LYMP, %	20.00 ± 11.10	19.15 ± 12.93	0.745
MONO, %	5.14 ± 2.33	5.32 ± 2.27	0.627
EOS, %	2.60 ± 3.67	2.91 ± 3.52	0.846
BASO, %	0.57 ± 0.33	0.54 ± 0.38	0.684
LUC, %	1.85 ± 0.89	1.67 ± 0.76	0.170
MPV, fl	8.43 ± 1.06	7.89 ± 0.80	< 0.000
APTT, sec	32.05 ± 8.66	29.25 ± 5.28	< 0.019
PT1, sec	1.19 ± 0.35	1.05 ± 0.22	< 0.003
PT2, sec	13.45 ± 3.96	11.91 ± 2.41	< 0.003
Fibrinogen,	348.25 ± 90.68	379.71 ± 111.72	0.057
Serum Electrolytes			
Na, mmol/L	142.61 ± 6.01	142.72 ± 4.35	0.901
K, mmol/L	4.76 ± 0.93	4.46 ± 0.57	< 0.008
Cl, mmol/L	105.62 ± 7.48	105.30 ± 5.18	0.742
LDH, U/L	740.87 ± 466.11	571.35 ± 172.63	< 0.002
Lipase, U/L	31.97 ± 17.47	30.62 ± 16.58	0.696
CK,	163.55 ± 150.15	401.33 ± 823.56	0.277
CK-MB,	3.67 ± 4.78	2.45 ± 3.53	0.071
Amylase, U/L	50.10 ± 37.53	47.98 ± 21.86	0.696
Routine Admission			
Calcium, mg/dL	8.74 ± 0.80	8.90 ± 0.67	0.113
Inorg. Phos., mg/dL	4.08 ± 1.19	3.34 ± 0.82	< 0.000

[0089]

Glucose, mg/dL	174.38 ± 78.27	156.08 ± 60.56	0.108
BUN, mg/dL	30.83 ± 21.37	19.60 ± 13.53	**< 0.000**
Creatine, mg/dL	1.62 ± 1.07	1.14 ± 0.72	**< 0.001**
Cholesterol, mg/dL	172.61 ± 48.65	168.66 ± 43.71	0.596
Protein, g/dL	6.85 ± 0.69	6.91 ± 0.72	0.630
Albumin, g/dL	3.82 ± 0.36	3.89 ± 0.44	0.327
Bilirubin(T), mg/dL	1.04 ± 0.71	0.73 ± 0.39	**< 0.001**
Bilirubin(D), mg/dL	0.37 ± 0.25	0.26 ± 0.17	**< 0.002**
ALP, U/L	100.24 ± 31.94	94.74 ± 47.48	0.386
AST, U/L	106.80 ± 231.56	41.33 ± 67.67	**< 0.013**
ALT, U/L	71.58 ± 152.29	30.52 ± 30.66	**< 0.015**
Ca^{2+}, mEq/L	2.25 ± 0.15	2.29 ± 0.16	0.061
Mg^{2+}, mg/dL	2.33 ± 0.35	2.16 ± 0.30	**< 0.002**
ABGA			
pH,	7.46 ± 0.06	7.46 ± 0.07	0.977
pCO2, mmHg	38.47 ± 13.29	38.56 ± 10.68	0.963
pO2, mmHg	77.64 ± 14.69	83.11 ± 19.77	**< 0.047**
HCO3,	23.96 ± 5.20	25.16 ± 4.08	0.114
BE,	−0.26 ± 5.67	1.47 ± 3.35	**< 0.018**
O2CT,	14.74 ± 3.58	16.36 ± 3.26	**< 0.004**
O2SAT, mmHg	95.88 ± 2.85	96.75 ± 2.03	**< 0.026**
TCO2,	25.06 ± 5.46	26.33 ± 4.36	0.114
Hb,	11.62 ± 2.48	12.43 ± 2.27	**< 0.035**
HCT,	35.84 ± 7.12	37.75 ± 6.53	0.083
CRP,	1.77 ± 2.12	4.15 ± 6.08	**< 0.002**
Pro BNP,	11,720.87 ± 12,416.65	2,133.80 ± 4,649.93	**< 0.000**
Troponin I,	0.42 ± 1.20	0.11 ± 0.20	**< 0.017**

[0090]

[0091] (이하 약어 설명-A P value < 0.05 was considered significant. Abbreviations: CHF, patients with a congestive heart failure; Non-C.D., patients without a congestive heart failure; M, males F, females S.G., specific gravity; O.B., occult blood; WBC, white blood cell; RBC, red blood cell; Ep. Cell, epithelial cell; HGB(Hb), hemoglobin; HCT, hematocrit; MCV, mean corpuscular volume; MCH, mean corpuscular hemoglobin; PLT, platelet count; NEUT, neutrophil; LYMP, lymphocyte; MONO, monocyte; EOS, eosinophil; BASO, basophil; LUC, large unstained cell; MPV, mean platelet volume; APTT, activated partial thromboplastin time; PT, prothrombin time; Cl, chloride; LDH, lactate dehydrogenase; CK, creatine kinase; CK-MB, creatine kinase MB fraction Inorg. Phos., inorganic phosphorus; BUN, blood urea nitrogen; Bilirubin(T), total bilirubin; Bilirubin(D), direct bilirubin; ALP, alkaline phosphatase; AST, aspartate aminotransferase; ALT, alanine aminotransferase; Ca2+, actual calcium; Mg2+, magnesium; ABGA, arterial blood gas analysis; O2CT, oxygen content; O2SAT, oxyhemoglobin saturation; TCO2, total carbon dioxide; CRP, c-reactive protein.)

[0092] [표 4]는 도 1에서 Train 1 ~ Train 10 (10-fold 교차검증) 동안에 본 발명에 따른 주요 파라미터 추출방법의 단계 1에서 단계 4에 의해서 결정된 중요 검사항목들의 리스트와 빈도(검사항목의 값들이 2개의 명목형 값으로 변환된 경우)를 나타낸 표이다.

표 4

Selected feature lists		Fold 1	Fold 2	Fold 3	Fold 4	Fold 5	Fold 6	Fold 7	Fold 8	Fold 9	Fold 10	Frequency
Urinalysis	pH								O			1
Common Blood Cell & Differential Count	WBC								O			1
	RBC					O						1
	HGB	O	O									2
	MCV			O		O	O					3
	MCH				O	O		O		O	O	5
	MCHC			O	O		O	O		O		5
	PLT	O	O									3
	NEUT	O							O			2
	MONO	O			O		O	O				4
	EOS	O	O	O	O	O	O	O	O	O	O	10
	MPV			O			O			O		3
Routine Admission	Calcium								O			1
	Inorganic Phosphorus			O								1
	Glucose		O									1
	BUN	O	O		O		O	O				5
	Cholesterol				O			O				2
	Bilirubin(T)										O	1
	Bilirubin(D)	O	O		O	O	O	O		O	O	8
	ALP	O	O			O						3
	ALT			O								1
Serum Electrolytes	K		O									1
	Cl								O	O		3
Arterial Blood Gas Analysis	BE								O			1
	O2SAT			O	O	O	O	O		O	O	7
	TCO2										O	1
PT1										O		1
Fibrinogen									O			1
LDH				O		O						2
Amylase						O					O	3
Troponin I		O	O	O	O	O	O	O	O	O	O	10
CK-MB											O	1
Length of feature lists		9	9	9	8	14	9	9	9	8	9	91

[0093]

[0094] [표 4]에서 Fold 1 ~ Fold 10은 각각 Train 1 ~ Train 10을 의미하고, 'O'은 각 Fold에서 단계 1에서 단계 4를 수행하였을 때, 선택된 검사항목들을 나타낸다. 예를 들어 Fold 1(Train 1)의 경우, {HGB, PLT, NEUT, MONO, EOS, BUN, Direct Bilirubin, Troponin I}가 중요 검사항목으로 선택된 것을 의미한다. 또한 'Length of feature lists'는 각 Fold에서 선택된 중요 검사항목들의 수를 의미하고, 'Frequencies'는 Fold 1 ~ Fold 10 동안에 각 검사항목에서 선택된 총 빈도수를 나타낸다.

[0095] [표 5]는 도 1에서 Train 1 ~ Train 10 (10-fold 교차검증) 동안에 본 발명에 따른 주요 파라미터 추출방법의 단계 1에서 단계 4에 의해서 결정된 중요 검사항목들의 리스트와 빈도(검사항목의 값들이 3개의 명목형 값으로 변환된 경우)를 나타낸 표이다.

표 5

Selected feature lists		Fold 1	Fold 2	Fold 3	Fold 4	Fold 5	Fold 6	Fold 7	Fold 8	Fold 9	Fold 10	Frequency
Urinalysis	RBC	O	O		O	O				O		5
Common Blood Cell & Differential Count	WBC										O	1
	RBC		O									1
	HGB							O				1
	HCT	O			O	O	O	O	O		O	7
	MCV				O				O			2
	MCHC	O		O		O	O		O		O	6
	NEUT			O								1
	MONO									O		1
	EOS	O					O		O	O	O	5
	MPV							O				1
Serum Electrolytes	Inorganic Phosphorus	O				O						2
	Cholesterol				O							1
	Bilirubin(T)										O	1
	AST							O				1
	Na			O	O					O		3
	K	O										1
	Cl		O						O			2
Arterial Blood Gas Analysis	pH		O		O	O	O					4
	BE									O		1
	O2SAT		O	O		O						3
	TCO2										O	1
APTT				O				O				3
Fibrinogen									O			1
LDH										O		1
Lipase			O									1
Pro_BNP		O	O	O	O	O	O	O	O	O	O	10
Ca^{2+}				O			O	O				3
Length of feature lists		7	7	7	7	7	6	7	7	7	7	69

[0096]

[0097] 이처럼, 본 발명에서는 10-fold 교차검증 동안에 1개의 참조치 (2개의 명목형 값 변환)와 2개의 참조치 (3개의 명목형 값 변환)에 의해서 결정된 모든 검사항목들을 고려하여 심부전과 비심인성 호흡곤란 환자군의 감별진단을 위한 임상 의사결정모형을 설계한다.

[0098] 도 6은 본 발명에 따른 통합 임상 의사결정 지원 시스템에 적용되는 의사결정모형과 종래의 의사결정 모형의 예를 나타내는 모형도이다. 심부전과 비심인성 호흡곤란 환자의 감별진단을 위한 임상 의사결정모형의 모식도를 나타낸다.

[0099] 도 6의 (a)에 나타낸 바와 같이, '타원형의 노드'는 검사항목을 나타내고, '직사각형의 노드'는 최종 의사결정(즉 Yes이면 심부전 환자군, No이면 비심인성 호흡곤란 환자군) 값을 의미한다. 또한 위의 의사결정모형은 도 1에서 'Clinical Decision Support Model (Using Decision Tree)' 모듈에 해당된다.

[0100] 도 6의 (a)에서 '타원형의 노드'는 검사항목을 나타내고, '직사각형의 노드'는 최종 의사결정(즉 Yes이면 심부전 환자군, No이면 비심인성 호흡곤란 환자군) 값을 의미한다. 또한 위의 의사결정모형은 도 1에서 'Clinical Decision Support Model (Using Decision Tree)' 모듈에 해당된다. 도 6의 (b)는 종래의 단계적 특징선택 기법(Stepwise)을 고려한 다중 회귀분석 후 의사결정트리에 의해서 생성된 모형을 나타낸다. 그리고, 본 발명의 실시예에서는 심부전 환자군의 감별진단을 위한 임상 의사결정모형의 성능평가는 도 1에 나타낸 바와 같이, 'Evaluation' 모듈에 수행된다.

[0101] [표 6]은 종래의 의사결정모형과 본 발명에 따른 통합 임상 의사결정 지원 시스템에 적용되는 임상 의사결정모형의 성능평가결과의 비교를 나타낸 표이다.

표 6

[0102]

평가척도	본 발명에 의해서 설계된 의사결정모형	다중회귀분석 후 의사결정모형
평균 민감도	72 %	68 %
평균 특이도	74 %	74 %
평균 기하평균	73 %	71 %
평균 지식의 수	16 개	11 개

[0103] [표 6]에서 평균 지식의 수는 도 6에서 음영으로 표시된 직사각형 노드 (즉 리프노드)의 수를 나타내며, 다음과 같이 심부전 환자군을 감별진단하기 위한 임상적 지식을 도출할 수 있다.

[0104] 예1) 본 발명에 의해서 설계된 의사결정모형에서 유도된 임상적 지식

[0105] If Pro BNP is <= 2,799 and Troponin I is <= 0.09 and BUN is <= 16

[0106] Then Diagnosis is No (Support = 37)

[0107] If Pro BNP is > 2,799 and Bilirubin(D) is > 0.3

[0108] Then Diagnosis is I50 (Support = 25 / 1)*

[0109] '25'의 값은 I50으로 정확히 분류된 환자 수 (True Negative, TN), '1'의 값은 No로 잘못 분류된 환자 수 (False Positive, FP)를 의미한다.

[0110] 예2) 다중회귀분석 후 의사결정모형에서 유도된 임상적 지식

[0111] If Pro BNP is <= 2,799 and Bilirubin(D) is <= 0.6

[0112] Then Diagnosis is No (Support = 79 / 10)

[0113] If Pro BNP is > 2,799 and Bilirubin(D) is > 0.3

[0114] Then Diagnosis is I50 (Support = 25 / 1)

[0115] [표 6]에서 평균 기하평균 (geometric mean)은 10-fold 교차검증 동안에 각 fold에서 다음 수식에 의해서 평가된 결과들의 평균값을 의미한다. 여기서 *sensitivity*와 *specificity*는 각각 민감도와 특이도 평가척도를 의미한다. 즉, 본 발명에 따라 설계된 의사결정모형이 민감도와 평균 지식의 수가 높게 나타나 정확도와 신뢰성이 높음을 알 수 있다.

[0116] 이처럼 본 발명에 따른 통합 임상 의사결정 원 시스템에서는 병원정보시스템(HIS)에서 다수의 임상 데이터 베이스(10)로부터 임상 전문가에 의해 정의된 질환별 데이터 베이스(20)로 구성하고(또는 질환별 데이터 마트), 상기 질환별 DB로부터 다수의 질환별 임상 데이터를 이용하여 본 발명에 따른 임상 의사결정 모듈(30)을 통하여 임상 의사결정 지원 모델을 설계한다.

[0117] 그리고 나서, 지식 데이터 베이스(60)에서 상기 임상결정 지원모듈로부터 임상 의사결정 지원 정보를 포함하여 생성한 일시적 지식(Temporary Knowledge)을 저장하고, 어플리케이션 인터페이스 모듈(70)에서 상기 시식 데이터 베이스(60)를 통해 생성된 임상 의사결정 지원 종합정보를 획득하게 된다.

[0118] 또한, 핵심 지식 저장 데이터 베이스(50)는 상기 임상 의사결정 지원 모듈(30)에서 생성된 정보와, 임상 전문가(40)에 의해 판단된 임상정보를 통해 핵심 지식을 저장하는 것도 역시 가능하다. 이처럼 임상 전문가에 의해 추가적인 핵심 지식을 추출하게 되면, 보다 신뢰성을 높일 수 있는 장점이 있다.

[0119] 이처럼, 본 발명의 통합 임상 의사결정 지원 시스템은 특정 질환에 대한 임상 의사결정모형뿐만 아니라, 유사질환에 대해서 부분적으로 설계된 임상 의사결정모형을 통합하고, 임상 지식을 DB화 함으로써, 유사질환의 감별진단을 위한 통합 임상 의사결정 지원 시스템을 구축할 수 있게 된다.

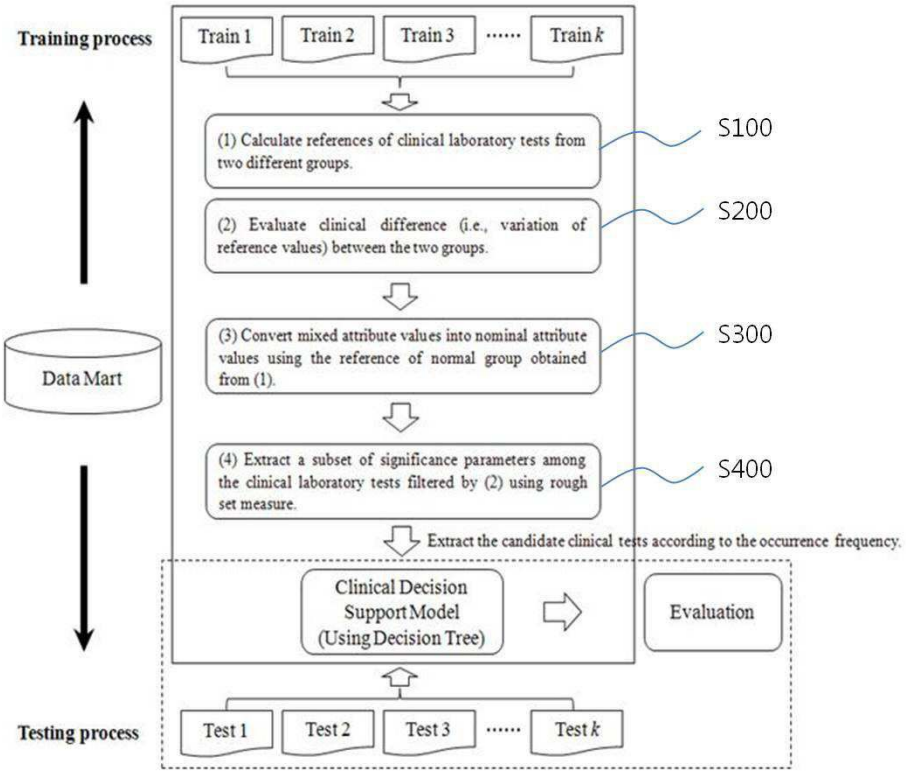
[0120] 또한, 도 5에서 일시적 지식 DB(Temporary Knowledge DB)를 추가적으로 고려하기 때문에, 임상 전문가에 의해서 검증된 지시 저장소(Core Knowledge Repository DB) 이외에 임상적 사례에 대해서 추론할 수 있는 부가적 기능을 제공할 수 있게 된다.

[0121] 더하여, 병원 내 분과별 인턴 혹은 레지던트들의 교육 및 학습용 콘테츠 제작시 효과적으로 사용할 수 있다는 점에서, 새로운 임상적 증례 혹은 사례에 대한 판단을 교과서 혹은 문헌 위주의 사고방식에서 벗어나, 병원정보시스템(HIS) 내에 수년간 누적된 실제 임상결과정보를 종합적으로 활용하여 '근거중심 의학적 판단'이 가능한 임상도구로서 활용할 수 있는 큰 장점이 있다.

[0122] 이상의 설명에서 본 발명은 특정의 실시 예와 관련하여 도시 및 설명하였지만, 특허청구범위에 의해 나타난 발명의 사상 및 영역으로부터 벗어나지 않는 한도 내에서 다양한 개조 및 변화가 가능 하다는 것을 당 업계에서 통상의 지식을 가진 자라면 누구나 쉽게 알 수 있을 것이다.

도면

도면1

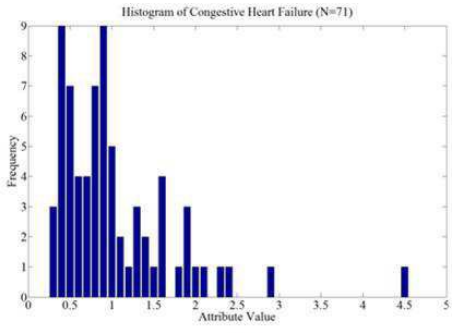


도면2

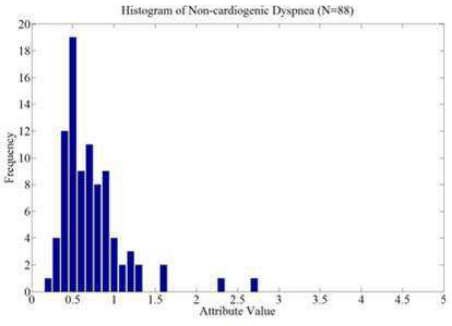
Lab. Test: Total Bilirubin

Attribute Values	CHF	Non-C.D	Attribute Values	CHF	Non-C.D
0.2	0	1	1.4	2	0
0.3	3	4	1.5	1	0
0.4	9	12	1.6	4	2
0.5	7	19	1.8	1	0
0.6	4	9	1.9	3	0
0.7	4	11	2.0	1	0
0.8	7	8	2.1	1	0
0.9	9	9	2.3	1	1
1.0	5	4	2.4	1	0
1.1	2	2	2.7	0	1
1.2	1	3	2.9	1	0
1.3	3	2	4.5	1	0

(a)

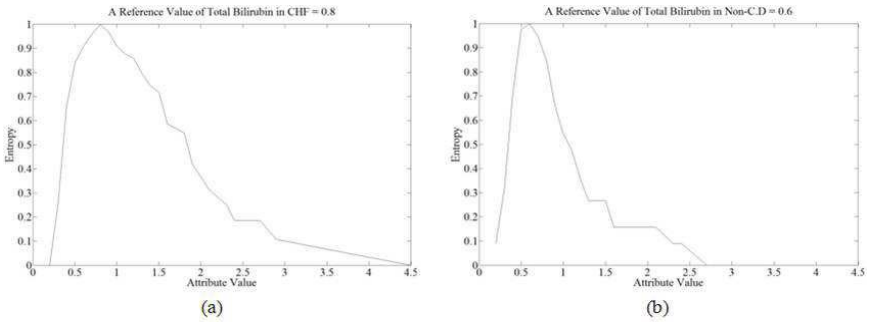


(b)

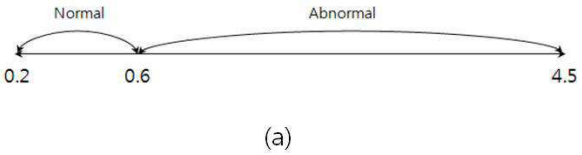


(c)

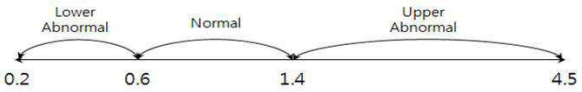
도면3



도면4

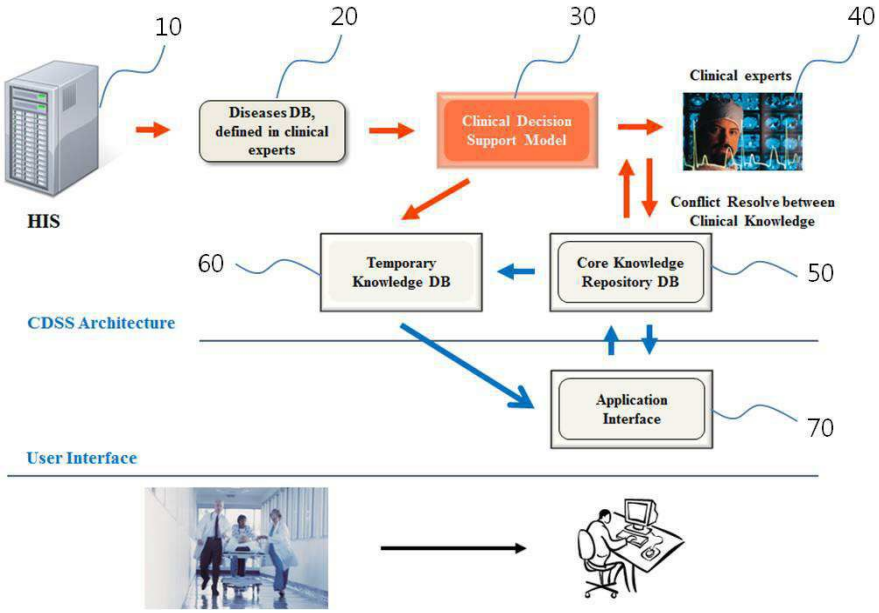


(a)

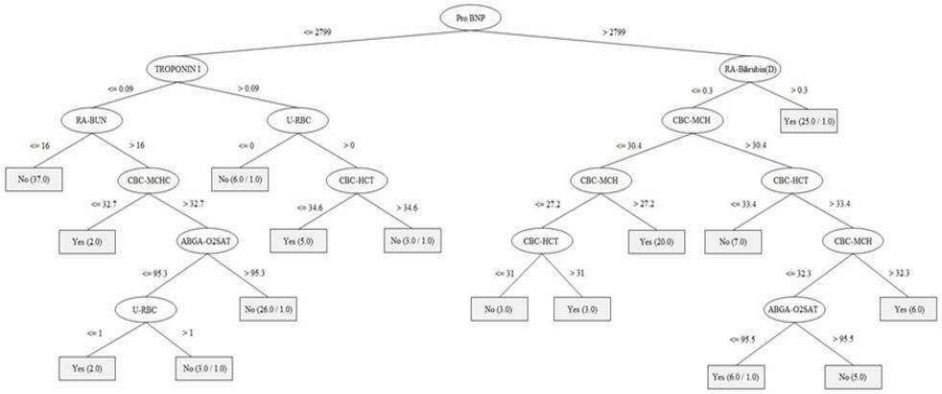


(b)

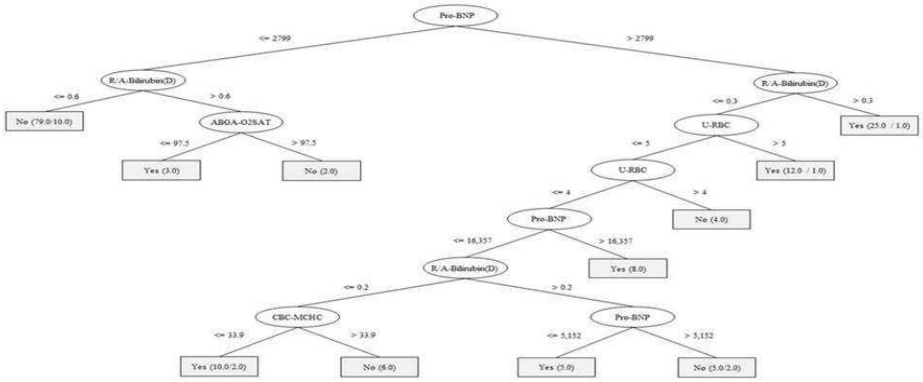
도면5



도면6


Pro BNP
<= 2799
> 2799
TROPONIN I
RA-Bilirubin(D)
<= 0.09
> 0.09
<= 0.3
> 0.3
RA-BUN
U-RBC
CBC-MCH
Yes (25.0 / 1.0)
<= 16
> 16
<= 0
> 0
<= 30.4
> 30.4
No (37.0)
CBC-MCHC
No (6.0 / 1.0)
CBC-HCT
CBC-MCH
CBC-HCT
<= 32.7
> 32.7
<= 34.6
> 34.6
<= 27.2
> 27.2
<= 33.4
> 33.4
Yes (2.0)
ABGA-O2SAT
Yes (5.0)
No (3.0 / 1.0)
CBC-HCT
Yes (20.0)
No (7.0)
CBC-MCH
<= 95.3
> 95.3
<= 31
> 31
<= 32.3
> 32.3
U-RBC
No (26.0 / 1.0)
No (3.0)
Yes (3.0)
ABGA-O2SAT
Yes (6.0)
<= 1
> 1
<= 95.5
> 95.5
Yes (2.0)
No (3.0 / 1.0)
Yes (6.0 / 1.0)
No (5.0)

(a)


Pro-BNP
<= 2799
> 2799
R/A-Bilirubin(D)
R/A-Bilirubin(D)
<= 0.6
> 0.6
<= 0.3
> 0.3
No (79.0/10.0)
ABGA-O2SAT
U-RBC
Yes (25.0 / 1.0)
<= 97.5
> 97.5
<= 5
> 5
Yes (3.0)
No (2.0)
U-RBC
Yes (12.0 / 1.0)
<= 4
> 4
Pro-BNP
No (4.0)
<= 16,357
> 16,357
R/A-Bilirubin(D)
Yes (8.0)
<= 0.2
> 0.2
CBC-MCHC
Pro-BNP
<= 33.9
> 33.9
<= 5,152
> 5,152
Yes (10.0/2.0)
No (6.0)
Yes (3.0)
No (5.0/2.0)

(b)