(19) 대한민국특허청(KR)
(12) 등록특허공보(B1)

(51) 。Int. Cl. [7]
G06F 15/18

(45) 공고일자 2002년06월22일
(11) 등록번호 10-0328119
(24) 등록일자 2002년02월27일

(21) 출원번호 10-1994-0024399
(22) 출원일자 1994년09월28일

(65) 공개번호 특1995-0009256
(43) 공개일자 1995년04월21일

(30) 우선권주장 93-245618 1993년09월30일 일본(JP)

(73) 특허권자 시스멕스 가부시키가이샤
이에츠구 히사시
일본국 고오베시 쥬오구 와키하마 카이간 도오리 1-5-1

(72) 발명자 가나이가즈유끼
일본국효고껜고베시쭈오꾸미나토지마나까마치7쵸메2-1도아이요우덴시가부시끼가이샤내

(74) 대리인 이후동

심사관 : 김준학

(54) 귀속도판별장치

요약

1.청구범위에 기재된 발명이 속한분야

본발명은 2군선형판별의 분석결과를 통합해서 다수군판별을 행하므로서 피검사데이터가 어느군에 귀속하는가의 판별을 행하는 귀속도 판별장치에 관한것이다.

2.발명이해결하려고하는 기술적과제

2군선형판별에의해 피검사데이터가 다군중의 어느군에 귀속되는가를 산출함과동시에 각2군의 조합의 판별마다에 그지지정도를 구하고 그 지지정도를 참조해서 피검사데이터의 귀속도를 판별하므로서 종래보다도 높은 신뢰성으로 귀속도의 판별을 행할수있도록한 귀속도판별장치에 관한것이다.

3.발명의 해결방법의 요지

다수의 군에대해 각군의 특징적인 패턴을 미리 기억한 기억수단과CPU를 구비하고 CPU에의해 기억수단에 기억된 다수의 군중에서 임의의2군을 모든 조합에대해 선택하고 선택한 각2군을 최적으로 2분하는 선형판별함수를 각 2군마다에

각각 설정하고 설정된 각 2군마다의 선형판별함수를 사용해서 각 2군마다에 피검사데이더가 어느쪽군에속하는가의 2군판별결과를 산출하고 지지정도를 각 2군마다에 결정하고 각 2군마다의 2군판별결과와 지지정도와에 기초하여 각군에대한 피검사데이터의 귀속도를 산출한다.

4.발명의 중요한 용도

이장치는 예를들면 질환의 진단 (환자가 어느질환군에 속하는가의 진단)에 유효하게 사용된다.

대표도

도 1

명세서

도면의 간단한 설명

제 1도는, 본 발명의 구성을 나타내는 블록도.

제 2도는, 본 발명에 의한 귀속도 판별장치의 일실시예의 구성을 나타내는 블록도.

제 3도는, 본 발명의 귀속도 판별장치의 처리개요를 나타내는 블록도.

제 4도는, 지식획득 프로세스와 판별진단 프로세스의 상세를 나타내는 블록도.

제 5도는, 2군 판별정도 Rij 의 함수 f(X)를 나타내는 그래프.

제 6도는, 지식획득 프로세스와 판별진단 프로세스의 전체의 처리내용을 나타내는 플로우차아트.

제 7도는, 판별진단 프로세스의 상세 내용을 나타내는 플로우차아트.

제 8도는, 실제로 귀속도의 판별을 행한 경우의 출력장표(지지모드 : 없음)의 예를 나타내는 설명도.

제 9도는, 실제로 귀속도의 판별을 행한 경우의 출력장표(지지모드 : 거리비)의 예를 나타내는 설명도.

제 10도는, 실제로 귀속도의 판별을 행한 경우의 출력장표(지지모드 : 거리비, 상관비)의 예를 나타내는 설명도.

* 도면의 주요부분에 대한 부호의 설명

101. 기억수단, 102. 판별함수 설정수단

103. 2군 판별결과 산출수단, 104. 지지정도 결정수단

105. 귀속도 산출수단

발명의 상세한 설명

발명의 목적

발명이 속하는 기술 및 그 분야의 종래기술

본 발명은, 2군 선형판별의 분석결과를 통합하여 다군(多群)판별을 행하므로서, 피검사데이터가 어느 군에 귀속하는가의 판별을 행하는 귀속도(歸屬度) 판별장치에 관한 것이다.

이 장치는, 예를 들면, 질환의 진단(환자가 어느 질환군에 속하는가의 진단)에 유효하게 사용된다.

종래의 2군 선형판별분석은, 착안하는 대상이 2군내의 어느 군에 귀속하는가를 통계적으로 판별하는 수법으로서, 대단히 유명한 방법이다.

또, 이 2군 선형판별분석은, 지식획득의 관점으로부터도 최종적으로는 2군 선형판별함수의 판별치의 정부(正負)로 2군내의 어느 군에 귀속하는가를 판결하기 때문에, 대단히 정보 압축률이 높다.

한편, 어느 질환데이터에 대해, 생각할 수 있는 다군 중에서 1군을 결정하는 방법(진단시스템)으로서는, 엑스퍼트시스템을 대표로 하는 기호화한 지식과 추론엔진으로 연역적으로 실현하는 방법, 뉴럴네트워크의 학습을 이용하는 방법, 다변량 해석의 응용으로서의 다군판별과 같이 귀납적으로 실현하는 방법 등이 있다.

인간, 예를 들면, 숙련된 전문의가 어떠한 사고 과정에서 다군 중에서 1군을 결정하는가를 생각해 보면, 결코 다군으로부터 단번에 1군을 결정할 수는 없고, 생각할 수 있는 군을 수군(數群)을 추려서, 그들을 상호 비교하므로서, 가장 「그럴듯한」군을 결정하고 있는 것이 아닌가 생각된다.

이와 같은 프로세스는, 일반적으로 「선호 프로세스」라 불리고 있고, 인간이 비교적 만족하고 있는 것이다.

이와 같은 다군패턴의 인식에 대해서는, 전자정보통신학회 논문지 A Vol. J 72-A No.1 pp.41-48(1989년 1월)에, 쌍판별에 의한 다군패턴인식에 관한 기재가 있다.

이 다군패턴의 인식에 있어서는, 모든 쌍에 대해 2군간의 판별을 행하여, 이들 결과를 통합해서 다군에의 판정결과를 내도록 하고 있다.

즉, 전체 군에 공통의 분산공분산행렬을 가정하는 것이 아니고, 쌍마다에 분산공분산행렬을 가정하여, 미지입력패턴 x에 대해 각 군 Πi, Πj로부터의 마하라노비스의 거리 Di^2, DJ^2, 사후확립 P(Πi | x), P(Πj | x)를 계산하고, 마하라노비스의 거리 Di^2, DJ^2의 대소 또는 사후확립 P(Πi | x) P(Πj | x)의 대소로부터 판별결과를 내고, 이것을 기준화해서(기준화대 통계량) 통합한 후, 최종적으로 미지입력패턴 x의 분류결과를 얻도록 하고 있다.

단, 사후확립 P(Πi | x), P(Πj | x)에 대해서는, 구제적인 산출방법의 기재는 없다.

또, 여기에는, 최종결과를 얻는 방법으로서, 다수결법, 미니막스법, 및 기대치에 의한 방법이 제시되어 있다.

이 다수결법이란, 2군 판별결과를 집계해서 다수결에 의해 정해를 얻고자 하는 것이다.

미니막스법이란, 군수가 많은 경우에 유효하며, 마하라노비스의 거리 중에서 최대치를 구하며, 거리가 커졌을 때, 그 군에 속하는 것을 부정하는 것이다.

기대치에 의한 방범이란, 군 Πi에 관한 기준화대 통계량의 기대치를 취하며, 최량의 결과를 부여하는 군을 해답으로 하는 것이다.

그러나, 이와 같은 종래의 다군판별에 있어서는, 예를 들면, 실제로 각 질환을 판별하는 시스템을 구축하는 때에는, 지식획득(성형판별함수의 산출등)에 사용하는 데이터의 질의 차이에 의해 얻어진 선형판별함수의 신빙성이 다른 것이 보통이다.

또, 실제의 질환의 판별에 있어서는, 대단히 중요한 의미를 갖는 조합의 2군 판별과, 별로 의미를 갖지 않는 조합의 2군 판별이 있고, 모든 조합의 2군 판별을 평등히 생각하는 것은 바람직하지 않다.

전문의는 경험을 바탕으로 하여 무의식 중에 이와 같은 처리를 행하고 있다고 생각된다.

이상의 것들은, 현실의 다군판별시스템 특히 질환의 판별을 행하는 시스템에 있어서는, 대단히 중요한 점이 된다.

본 발명은, 이와 같은 사정에 감안하여 이루어진 것으로서, 2군 선형판별에 의해 피검사데이터가 다군 중의 어느 군에 귀속되는가를 산출함과 동시에, 각 2군의 조합의 판별마다에 그 지지정도를 구하고, 그 지지정도를 참조해서 피검사데이터의 귀속도를 판별하므로서, 종래보다도 높은 신뢰성으로 귀속도의 판별을 행할 수 있도록 한 귀속도 판별장치에 관한 것이다.

제 1도는 본 발명의 구성을 나타내는 블록도이며, 도면에 나타내는 바와같이, 본 발명은, 다수의 군에 대해 각 군의 특징적인 패턴을 미리 기억한 기억수단(101)과, 기억수단(101)에 기억된 다수의 군 중에서 임의의 2군을, 모든 조합에 대해 선택하고, 선택한 각 2군을 최적으로 2분하는 선형판별함수를, 각 2군마다에 각각 설정하는 판별함수 설정수단(102)과, 판별함수 설정수단(102)에 의해 설정된 각 2군마다의 선형판별함수를 사용해서, 각 2군마다에 피검사데이터가 어느 쪽의 군에 속하는가의 2군 판별결과를 산출하는 2군 판별결과 산출수단(103)과, 2군 판별결과 산출수단(103)에 의한 2군 판결결과의 지지정도를 각 2군마다에 결정하는 지지정도 결정수단(104)과, 각 2군마다의 2군 판별결과와 지지정도에 기초하여, 각 구에 대한 피검사데이터의 귀속도를 산출하는 귀속도 산출수단(105)을 구비하여 이루어진 귀속도 판별장치이다.

본 발명에 있어서의 기억수단(101)으로서는, 플로피디스크장치나 자기디스크장치 등의 각종의 외부기억장치를 사용할 수가 있다.

판별함수 설정수단(102), 2군 판별결과 산출수단(103), 지지정도 결정수단(104) 및 귀속도 산출수단(105)으로서는, CPU, ROM, RAM, I/O포트로 이루어진 마이크로컴퓨터를 사용할 수가 있다.

본 발명에 있어서는, 지지정도 결정수단(104)에 의한 지지정도는, 근소한 차이로 판별된 것인가 큰 차이로 판별된 것인가의 정도를 수량화한 2군 판별정도와, 선형판별함수의 확실성의 정도를 수량화한 판별함수의 확신도와, 판별하고자 하는 2군의 중요성의 정도를 수량화한 중요도의 3개의 요소를 고려하므로서 결정되는 것이 바람직하다.

또, 지지정도 결정수단(104)의 지지정도에 있어서의 2군 판별정도는, 피검사데이터의 각 2군에 대한 마하라노비스 거리의 2승을 구하고, 이들 수치의 거리비를 구하여, 이 거리비를 정규화 한 것을 2군 판별정도로 하므로서 구해지는 것이 바람직하다.

또한, 지지정도 결정수단(104)의 지지정도에 있어서의 판별함수의 확신도는, 판별함수 설정수단에 의해 각 2군마다에 각각 선형판별함수가 설정될 때의 판별성능의 평가지표의 하나인 상관비에 의해 결정되는 것이 바람직하다.

본 발명에 의하면, 판별함수 설정수단에 의해, 기억수단에 기억된 다수의 군중에서 임의의 2군를 선택하며, 선택된 2군을 최적으로 2분하는 선형판별함수를 설정한다.

다음에, 이 선형판별함수를 사용해서, 2군 판별결과 산출수단에 의해, 피검사데이터가 2군 중의 어느 군에 속하는가의 2군 판별결과를 얻음과 동시에, 지지정도 결정수단에 의해, 그 2군 판별결과의 지지정도를 얻어, 이것을 모든 2군 쌍의 조합에 대해 행한다.

그리고, 얻어진 2군 판별결과와 지지정도에 기초하여, 피검사데이터의 각 군에 대한 귀속도를 산출한다.

이에 의해, 획득지식의 실상에 맞는 정확한 질환군의 판별을 할 수가 있다.

(실시예)

이하, 도면에 나타내는 실시예에 기초하여 본 발명을 상세히 설명한다.

또한, 이에 의해 본 발명이 한정되는 것은 아니다.

본 실시예에 있어서는, 혈액분석장치로부터 얻어지는, WBC(백혈구 수), RBC(적혈구 수), HGB(헤모글로빈 량), MCV(평균적혈구 용적), MCHC(평균적혈구 헤모글로빈 농도), PLT(혈소판 수), RDW(적혈구 분포폭)의 7종의 측정데이터를 사용하여, 7차원의 피검사데이터 X(= X1, X2, X3, X4, X5, X6, X7)가, β-중증성 지중해 빈혈, 철 결핍성 빈혈, 2차성 빈혈①, 2차성 빈혈②, 2차성 빈혈③, 재생불량성 빈혈+MDS, 용혈성 빈혈①, 용혈성 빈혈②, 거적아구성(巨赤芽球性) 빈혈, 철 결핍성 빈혈치료증, 제대혈(臍帶血), 다혈증, 정상(성인), 정상(소아)의 14군의 어느 것인가의 군(질환)에 속하는가의 판별을 행하는 경우에 대해 설명한다.

제 2도는 본 발명에 의한 귀속도 판별장치의 일실시예의 구성을 나타내는 블록도이다.

이 도면에 있어서, 10은 플로피디스크장치나 자기디스크장치 등의 각종의 외부기억장치로 구성되는 데이터베이스, 12는 2군 판별분석장치, 14는 예를 들면 RAM과 같은 기억장지, 16은 제어장치, 18은 CRT 디스플레이장치와 같은 표시장치나 도트프린터와 같은 인자장치로 이루어진 출력장치이다.

2군 판별분석장치(12) 및 제어장치(16)는, CPU, ROM, RAM, I/O포트와 같은 마이크로컴퓨터에 의해 구성되어 있다.

제 3도는 본 발명의 귀속도 판별장치의 처리개요를 나타내는 블록도이며, 이 블록도에 기초하여 본 장치의 처리의 개요를 설명한다.

우선, 지식획득 프로세스에 대해 설명한다.

본 장치에 의한 귀속도의 판별에 있어서는, 상기 각 군(질환) i : 1, 2, …, 14 마다에 다수의 양질의 지식획득 데이터 Di : D1, D2, …, D14를 확보해 둘 필요가 있다.

이 지식획득 데이터를 Di는, β-중증성 지중해 빈혈, 철 결핍성 빈혈, 2차성 빈혈①, 2차성 빈혈②, 2차성 빈혈③, 재생불량성 빈혈+MDS, 용혈성 빈혈①, 용혈성 빈혈②, 거적아구성 빈혈, 철결핍성 빈혈치료중, 제대혈, 다혈증, 정상(성인), 정상(소아)의 14군에 소속하는 사람의 혈액을 실제로 분석하여, 각 군에 속하는 사람의 WBC, RBC, HGB, MCV, MCHC, PLT, RDW의 7종의 혈액측정데이터를 각각 기록한 것이다.

이 지식획득 데이터군 Di는, 데이터베이스(10)에 격납되어 있다.

2근 판별분석장치(12)는, 데이터베이스(10)로부터 이 지식획득 데이터군 Di를 판독하여, 이 지식획득 데이터근 Di의 모든 2군 쌍의 조합에 대해 2군 판별분석을 행하여, 선형판별함수 Fij 등의 지식을 획득한다.

획득된 지식은 기억장치(14)에 기억된다.

다음에, 판별진단 프로세스에 대해 설명한다.

제어장치(16)는, 미지의 피검사데이터 X를 수용하여, 피검사데이터 X와 획득된 지식으로부터 피검사데이터 X가 각 질환에 속할 가능성을, 2군 판별결과 Aij 및 그 지지정도 Sij의 수치로서 산출한다.

이 피검사데이터 X란, 어떤 질환에 속하는가를 조사하려고 하는 환자의 실제의 WBC, RBC, HGB, MCV, MCHC, PLT, ROW의 7종의 혈액데이터이다.

수치로서 산출된 각 질환에 속할 가능성의 결과는, 각 군마다에 집계되고, 귀속도 Ki로서 출력장치(18)에 출력된다.

본 장치에 있어서는, 미지의 피검사데이터 X와 상기한 획득된 지식으로부터 질환의 판별을 행한다.

제 4도는 상기한 지식획득 프로세스와 판별진단 프로세스의 상세한 것을 나타내는 블록도이다.

동 도면에 있어서, 지식획득 프로세스는 블록 20으로 나타내고, 판별진단 프로세스는 블록 22로 나타낸다.

이하, 이 블록도를 참조하면서 본 장치의 처리에 대해 설명한다.

2군 판별분석장치(12)에서는, 구체적으로는 14군, 즉, 지식획득 데이터 D1, D2, …, D14의 모든 2군 쌍에 대해, 2군 선형판별분석을 행한다.

2군 쌍의 조합(i, j)은 전부가 $_{14}C_2 = (14 \times 13)/2 = 91$ 가지가 있다.

그 조합에 대응해서, 2군을 분획하기 위해, 선형판별함수 Fij(X1, X2, …, X7)도 91가지가 필요하게 된다.

선형판별함수 Fij(X1, X2, …, X7)은 다음 식과 같이 된다.

$$Fij(X1, X2, \cdots, X7) = aij0 + aijX1 + aij_2 X2 + \cdots + aij_7 X7 \quad 식(1)$$

단, aij_0, aij_1, aij_2, …, aij_7는 계수, X1, X2, …, X7은 변수로 한다.

지식획득 데미터의 질환군 Di , Dj를 최량으로 판단하는 선형판별함수 Fij(X1, X2,…, X7)를 구하는 방법, 즉, 상기한 계수 aij_0, aij_1, aij_2, …, aij_7을 구하기 위해서는, 공지의 방법을 사용할 수가 있다.

또, 지식획득 데이터의 질환군 Di , Dj의 각 평균치행렬 Yi, Yj , 각 분단공분산행렬 Si, Sj는, 상관비를 구하는 방법도 공지의 것이다.

이들 행렬 Yi, Yj, Si, Sj는, 후술하는 피검사데이터의 군 i, j에의 마하라노비스 거리의 2승 Mi , Mj를 산출하는 때에 사용된다.

이상의 2군 쌍(i , j)마다에 얻어진 선형판별함수식의 계수 aij_0, aij_1, aij_2, …, aij_7 평균치행렬 Yi, Yj, 분산공분산행렬 Si,Sj, 및 상관비가, 본 장치에서 사용되는 획득된 지식이 된다.

제어장치(16)에서는, 피검사데이터 X와 획득된 지식으로부터 피검사데이터 X가 각 질환에 속할 가능성을 수치로서 산출한다.

즉, 질환의 2군 쌍(i , j)를 선택하며, 피검사데이터 X와 그 2군 쌍에 대응하는 선형판별함수 Fij(X1, X2, …, X7)로부터, 피검사데이터 X가 질환 i에 속하는가, 질환 j에 속하는가를 구한다.

이 경우, 질환 i에 속한다고 판별된 때에는 Aij=1로 하고, 질환 j로 판별된 때에는 Aij=0으로 한다.

이 2군 판별결과 Aij는, 피검사데이터 X : (X1, X2, …, X7)을 상기한 식 (1)에 대입해서 얻어지는 판별함수치의 정부에 의해 결정할 수 있다.

질환 i에 속할 가능성의 정도 Ki를 산출하기 위해서는, 상기한 2군 판별결과 Aij의 정보만을 사용하여 가능성 정도 Ki를 구하는 것이 가능하다.

그러나, 일반적으로 정밀도는 별로 좋지 않다.

왜냐하면, 큰 차이로 판별되는 경우와 근소한 차이로 판별되는 경우가 있으나, 그 점이 고려되지 않기 때문이다.

여기서, 본 발명에서는, 군 i와 군 j의 2군 판별결과에 대한 지지정도 Sij의 개념을 도입하여, 같은 2군 판별결과라도, 어느 정도의 판별정도인가를 고려할 수 있게 했다.

2군 판별결과의 지지정도는, 2군 판별결과의 정도를 나타내는 것이며, 0 1의 값을 취한다.

이 지지정도가 1에 가까울수록 2군 판별결과에 대한 지지정도가 높고, 결국 큰 차이로 판별된 것을 나타내고, 지지점도가 0에 가까울수록 2군 판별결과에 대한 지지정도가 낮고, 결국 근소한 차이로 판별된 것을 나타낸다.

이하, 구체적인 2군 판결결과의 지지정도 Sij의 산출예를 설명한다.

또한, 이 설명에 있어서는, 판별함수의 확신도 Cij , 증요도 Wij, 2군 판별정도 Rij로 나누어서 각각 설명한다.

(1) 판별함수의 확신도 Cij의 산출방법

판별함수의 획득은, 당연히 현실의 데이터로부터 행해지기 때문에, 각 군의 지식획득 데이터의 질이 지식인 판별함수에 크게 영향을 준다.

2군 판별분석에서는, 얻터진 2군 선형판별함수 Fij(X1, X2, …, X7)의 판별성능을 나타내는데 상관비를 사용할 수가 있다.

이 상관비는 군 i와 군 j의 2군을 판별하는 선형판별식에, 피검사데이터 X를 대입해서 얻어지는 판별함수치 Z의 양 군의 도수분포를 구하며, 각 군마다의 판별함수치의 평균치 Yi, Yj , 분단 σi^2, σi^2 (견본수 : ni, nj)와, 양은 전체의 평균치 Y_T 분산 σ_T^2 [본수 : $n_T (=n_i+n_j)$]를 산출한다.

여기서, 급내분산(級內分散)을 σ_W^2 으로 하면, 급내분산 σ_W^2 은

$$\sigma_W^2 = (n_i \sigma_i^2 + n_j \sigma_j^2)/n_T$$

급간(級間)분산을 σ_B^2 이라 하면, 급간분산 σ_B^2 는

$\sigma_B^2 = [n_i (Y_i - Y_T)^2 + n_j (Y_j - Y_T)^2]/n_T$ 가 된다.

또, $\sigma_T^2 = \sigma_W^2 + \sigma_B^2$ 이 되기 때문에, 상관비는, 다음과 같이 산출된다.

상관비 $Cij = \sigma_B^2/\sigma_T^2$ (0 < 상관비 < 1)

이 상관비는, 판별능력이 높으면 급간분산이 크고, 급내분산이 적게되기 때문에, 1 에 가까운 값이 된다.

이 상관비를 판별함수의 확신도 Cij 로 한다.

(2) 중요도 Wij 의 설정

피검사데이터 X에 대한 i군에의 귀속도는, 전체 14군 중, 군 i와 군 i를 제외한 13군의 2군 쌍의 선형판별의 결과로서 구해진다.

군 i에의 귀속도를 산출하는 때에, 일반적으로는 13조(본 실시에에서는, 하나의 군에 관해 13조의 2군 쌍이 존재한다)의 2군 쌍에 대해 중요도에 차이가 있는 경우가 보통이다.

예를 들면, 철 결필성 빈혈군에의 귀속도를 산출하는 때에, β - 사라세미아군, 정상군 등의 유사군과의 판별은, 임상적으로 중요하지만, 다혈증군등과의 판별은 정보적으로도 별로 가치가 없고, 임상적 중요도도 낮다.

이와 같은 관계를 13점 만점(본 실시예에 있어서는, 하나의 군에 대해 13조의 2군 쌍이 존재하기 때문에)으로 각 2군 쌍에 중요도 Wij로서 할당한다.

(3) 2군 판별점도 Rij의 산출방법

피검사데이터 X의 질환군 i, j에의 마하라노비스 거리의 2승 Mi, Mj를 산출한다.

질환 i에의 마하라노비스 거리의 2승 Mi는, 잘 알려져 있는 바와 같이, 피검사데이터 X, 지식획득으로 얻어진 질환 i군의 평균치행렬 Yi, 분산공분산행렬 Si로부터 다음 식으로 구해진다.

$$Mi={}^{t}(X-Yi)Si^{-1}(X-Yi)$$

단, ^{t}는 전치행렬, $^{-1}$ 은 역행렬을 나타낸다.

질환 j에의 마하라노비스 거리의 2승 Mj도 같은 모양으로 구해진다.

2개의 마하라노비스 거리의 2승으로부터 Mi의 Mj에 대한 상대적인 마하라노비스 거리의 2승비 MRij를 설정하며, 그 값 MRiJ를 함수 f에 의해 0 1의 값으로 변환하여 2군 판별정도 Rij로 한다.

예를 들면, MRij(=Mj/Mi)를 x로 해서

$$Rij=f(x)=\frac{1-\exp[-0.366x(x-1)]}{1+\exp[-0.366x(x-1)]}$$

단, $x\leq 1$인 때에는, Rij=0으로 한다.

로 산출한다.

제 5도는 상기한 함수 f(x)의 그래프이다.

이 함수 f(x)는 정규화를 위한 함수의 일예로서, 1 ∞의 값을 취하는 MRij를, 0 1로 연속적으로 변환할 수가 있다.

정수 -0.366의 근거는, 피검사데이터 X의 i, j군에의 마하라노비스 거리의 비가 2인 때(i군으로의 거리에 대해, j군에의 거리가 2배인 때 : MRij는 2승이기 때문에 4로 한다)에, 2군 판별정도 Rij가 0.5가 되도록 설정되어 있다.

이 정수 -0.366에 관해서는, 판별하는 군의 상대적인 가깝기에 따라 변경할 필요가 있다.

또한, 그 외에, 지식획득 데이터에 2군 성형판별식을 적용한 때의 판별적중율을 사용하는 방법도 있다.

이상의 결과로부터, 판별결과에 대한 지지점도 Sij를

$$Sij=Rij\times Cij\times Wij$$

로 구해서, 대음 식으로 i군에 대한 귀속도 Ki를 산출한다.

$$Ki=\sum_{j=1}^{n} Aij \times Sij \times 1/(n-1)$$

단 n=14, j≠i로 한다.

또, Aij=1인 경우는 질환 i측으로 판별된 때, Aij=0의 경우는 질환 j측으로 판별된 때이다.

상기한 귀속도 Ki는, 0≤Ki≤1로 된다.

지지정도 Sij는, 2군 판별정도 Rij만으로도 종래보다도 판별정밀도를 양호하게 할 수 있으나, 판별함수의 확신도까지 고려하면(Sij = Rij × Cij), 지식획득 데이터의 실상에 맞는 것이 된다.

그리고, 다시 또 중요도 Wij까지 고려하면(Sij = Rij × Cij × Wij), 더 한층 인간의 판단에 근접시키는 것이 가능해진다.

다음에, 이와 같은 처리동작의 내용을 제 6도 및 제 7도에 나타내는 플로우차아트에 기초해서 설명한다.

제 6도는 지식획득 프로세스와 판별진단 프로세스의 전체의 처리내용을 나타내는 플로우차아트이다.

이 도면에 나타내는 바와 같이, 본 장치의 처리에서는, 우선, 지식획득 프로세스에 있어서, 지식획득 데이터를 Di로부터, 각 2군 쌍마다에, 선형판별함수 Fij, 평균치행렬 Yi, 분산공분살행렬 Si, 상관에 Cij를 획득한다(단계31).

다음에, 각 2군 쌍마다의 중요도 Wij를 설정한다(단계 32).

또한, 이 중요도 Wij는, 지식획득 데이터군 Di와 함께 대리 설정해 두어도 된다.

그리고, 판별진단 프로세스에 있어서, 이들 선형판별함수 Fij, 평균치행렬 Yi, 분단공분산행렬 Si, 상관비 Cij, 중요도 Wij를 사용해서, 피검사데이터 X에 대해 다군판별을 행한다(단계33).

제 7도는 판별진단 프로세스의 상세내용을 나타내는 플로우차아트이다.

판별진단 프로세스의 처리에 있어서는, 우선, 피검사데이터 X가 입력되면(단계41), 임의의 군 i를 지정하고(단계42), 피검사데이터 X를 선형판별함수식 Fij에 대입하여, 2군 판별결과 Aij를 결정한다(단계43).

이어서, 피검사데이터 X와, 평균치행렬 Yi 및 분산공분산행렬 Si로부터, 마하라노비스의 거리 Mi, Mj를 산출하며(단계44), 마하라노비스의 거리 Mi, Mj의 2승비 MRij에 기초해서, 2군 판별정도 Rij를 산출한다(단계45).

다음에, 상관비 Cij, 중요도 Wij, 2군 판별정도 Rij에 의해, 2군 판별결과 Aij의 지지정도 Sij를 산출한다(단계46).

여기서, 지정군 i에 대해서의 전체조합 ij(단, j≠i)가 종료되었는가 아닌가를 조사하여(단계47), 종료되었으면, 2군 판별결과 Aij와 그 2군 판별결과의 지지정도 Sij로부터, 지정군 i의 귀속도 Ki를 산출한다(단계 48).

그리고, 모든 군 i에 대해 종료되었는가 아닌가를 조사하여(단계 49), 모든 군 i에 대해 종료되였으면, 귀속도 Ki를 값이 높은 순으로 나란히 바꾸어서(단계 50), 가장 가능성이 높은 군 i를 결정한다(단계 51).

이상의 처리를 실제로 행한 경우의 출력 장표의 예를 제 8도 내지 제 10도에 나타낸다.

이들 도면에 나타낸 것은, 2차성 빈혈 ③의 질환을 갖는 검사체를 피검사데이터로 하고, 상기한 프로세스를 경유하며 각 군의 귀속도 Ki를 구한것이다.

이들 도면 중, 제 8도는 2군 판별결과만을 사용해서 2군 판별결과의 지지정도 Sij를 고려하지 않은 경우(지지모드 : 없음)의 예를 나타내고, 제 9도는 2군 판별결과의 지지정도 Sij 중, 2군 판별정도 Rij만을 고려한 경우(지지모드 : 거리비)를 나타내고, 제 10도는 2군 판별결과의 지지정도 Sij 중, 2군 판별정도 Rij와 판별함수의 확신도 Cij를 고려한 경우(지지모드 : 거리비, 상관비)를 나타내고 있다.

각 모드에 있어서의 귀속도 Ki의 산출식은 다음과 같이 된다.

지지모드 : 없는 경우

$$Ki = \sum_{j=1}^{n} Aij \times 1/(n-1)$$

(단, j≠1)

지지모드 : 거리비의 경우

$$Ki = \sum_{j=1}^{n} Aij \times Rij \times 1/(n-1)$$

(단 j≠1)

지지모드 : 거리에, 상관비의 경우

$$Ki = \sum_{j=1}^{n} Aij \times Rij \times Cij \times 1/(n-1)$$

(단 j≠1)

판별결과의 지지정도 Sij를 고려하지 않는 경우에는, 제 8도에 나타내는 바와 같이, 2차상 빈혈 ③군과 용혈성 빈혈 ①의 귀속도 Ki가 동일했으나, 판별결과의 지지정도 Sij를 고려한 경우에는, 제 9도의 2군 판별정도 Rij만을 고려한 경우, 및, 제 10도의 2군 판별정도 Rij와 판별함수의 확신도 Cij의 쌍방을 고려한 경우에도, 2차성 빈혈 ③의 귀속도 Ki가 제 1위로 되어 있다.

또, 지지정도 Sij에 대해, 판별함수의 확신도 Cij를 고려하지 않는 경우(제 9도)와 고려한 경우(제 10도)를 비교하면, 고려한 경우(제 10도)의 쪽이 1위, 2위의 차이가 상대적으로 크게 되어 있다.

이와 같이해서, 지식획득 데이터군 D1, …, Dn으로부터, 모든 2군 쌍의 조합에 대해 2군 판별분석을 행하며 지식이 되는 판별함수 Fij를 얻는다.

그리고, 이 판별함수 Fij를 사용해서 미지의 피검사데이터 X에 대해 전체 조합의 2군 판별분석을 행하여, 2군 판별결과 Aij, 지지정도 Sij를 얻어, 이 2군 판별결과 Aij와 지지정도 Sij로부터 피검사데이터 X의 각 군에 대한 귀속도 Ki를 산출하므로서, 획득지식의 실상에 맞는 정확한 질환군의 판별을 행할 수가 있다.

본 발명에 의하면, 다수의 군 중에서 임의의 2군을 선택하며, 선택한 2군을 최적으로 2분하는 선형판별함수를 설정하며, 그 선형판별함수를 사용하여 피검사데이터가 2군 중 어느 군에 속하는가의 2군 판별결과를 얻음과 동시에, 그 2군 판별결과의 지지정도를 얻어, 얻어진 2군 판별결과와 지지정도에 기초해서, 피검사데이터의 각 군에 대한 귀속도를 산출하도록 했기 때문에, 획득지식에 맞는 정확한 질환군의 판별이 가능해진다.

(57) 청구의 범위

청구항 1.

혈액분석장치로부터 얻어진 피(被)검사데이터가 다수의 질환(疾患)군 중 어느 것인가에 속하는가를 판별하기 위한, 귀속도 판별장치로서,

상기 다수의 질환군에 대해 각 군의 특징적인 패턴을 미리 기억한 기억수단(101)과,

상기 기억수단(101)에 기억된 다수의 질환군 중에서 임의의 2군을, 모든 조합에 대해 선택하고, 선택한 각 2군을 최적으로 2분하는 선형판별함수를, 각 2군마다에 각각 설정하는 판별함수 설정수단(102)과,

상기 판별함수 설정수단(102)에 의해 설정된 각 2군마다의 선형판별함수를 사용하여, 각 2군마다에 피검사데이터가 어느 쪽 군에 속하는가의 2군 판별결과를 산출하는 2군 판별결과 산출수단(103)과,

상기 2군 판별결과 산출수단(103)에 의한 2군 판별결과의 지지정도를 각 2군마다예 결정하는 지지정도 결정수단(104)과,

상기 각 2군마다의 2군 판별결과와 상기 지지정도에 기초하여, 각 군에 대한 피검사데이터의 귀속도를 산출하는 귀속도 산출수단(105)을 구비하여 이루어진 것을 특징으로 하는 귀속도 판별장치.

청구항 2.

제 1항에 있어서,

상기 지지정도 결정수단(104)에 의한 지지정도가, 근소한 차이로 판별된 것인가 큰 차이로 판별된 것인가의 정도를 수량화한 2군 판별정도와, 선형판별함수의 확실성의 정도를 수량화한 판별함수의 확신도와, 판별하고자 하는 2군의 중요성의 정도를 수량화한 중요도의, 3가지의 요소를 고려하므로서 결정되는 것은 특징으로 하는 귀속도 판별장치.

청구항 3.

제 2항에 있어서,

상기 지지정도 결정수단(104)의 지지정도에 있어서의 2군 판결정도가, 피검사데이터의 각 2군에 대한 마하라노비스의 거리의 2승을 구하고, 이들 수치의 거리비를 구하여, 이 거리비를 정규화 한 것을 2군 판별정도로 하므로서 구해지는 것을 특징으로 하는 귀속도 판별장치.

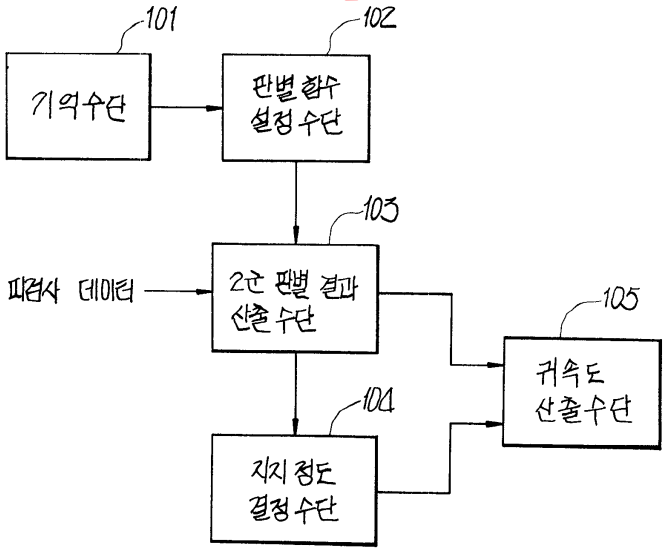
청구항 4.

제 2항에 있어서,

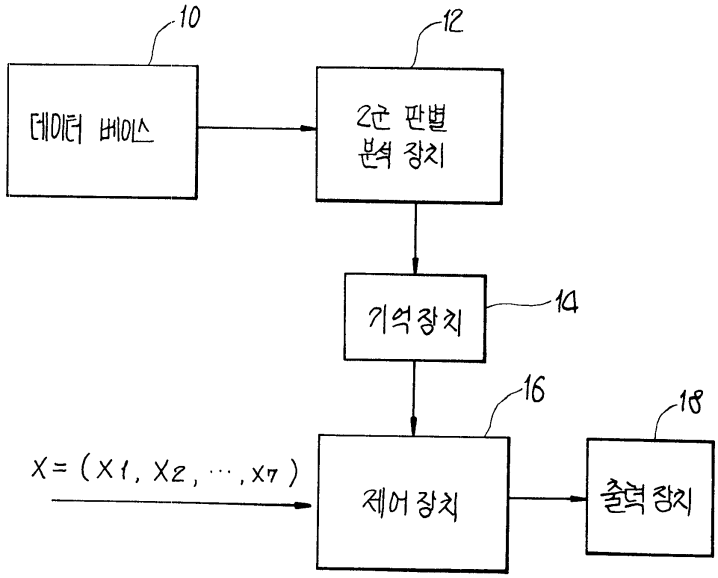
상기 지지정도 결정수단(104)의 지지정도에 있어서의 판별함수의 확신도가, 판별함수 설정수단에 의해 각 2군마다에 각각 선형판별함수가 설정되는 때의 판별성능의 평가지표의 하나인 상관비에 의해 결정되는 것을 특징으로 하는 귀속도 판별장치.

도면

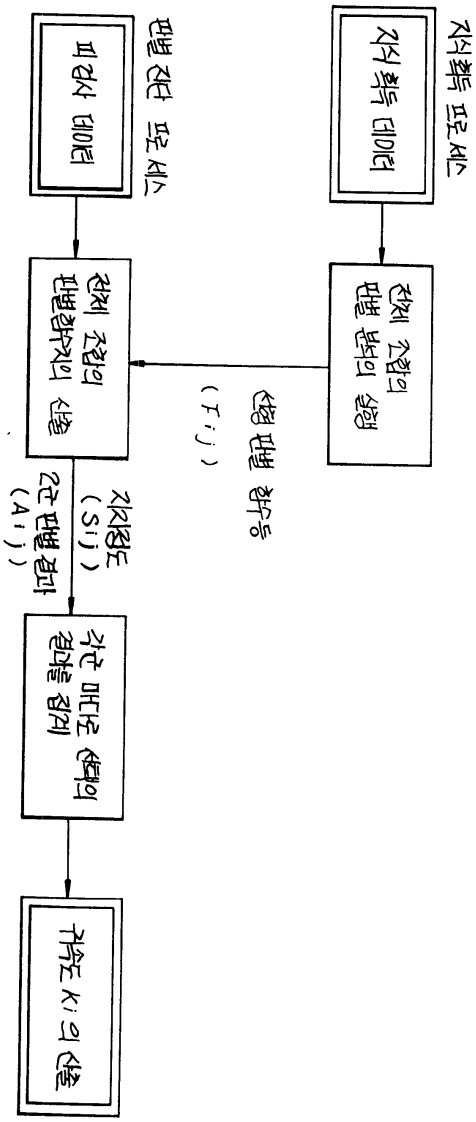
도면 1


101
기억수단
102
판별함수
설정수단
103
피검사 데이터
2군 판별 결과
산출수단
105
귀속도
산출수단
104
자치정도
결정수단

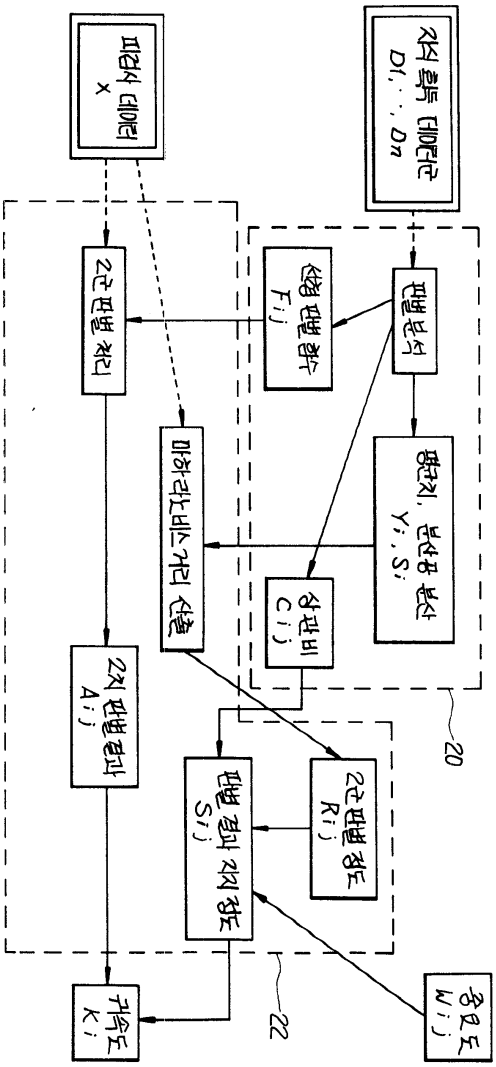
도면 2


10
데이터 베이스
12
2군 판별
분석 장치
14
기억장치
16
제어장치
18
출력장치
X = (X1, X2, ··· , X7)

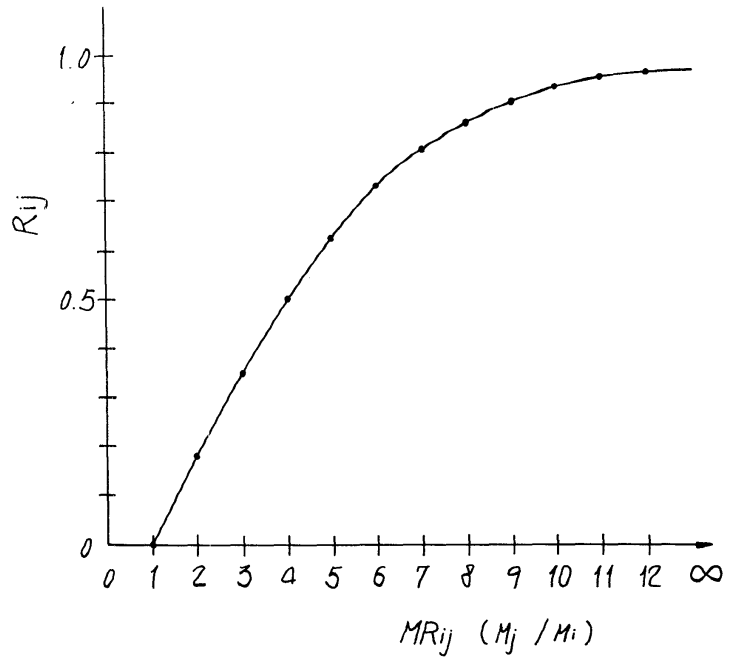
도면 3


지식 획득 프로세스
지식 획득 데이터
전제 조건의 판별 변수의 설정
판별 전단 프로세스
피 검사 데이터
전제 조건의 판별함수치의 산출
선형 판별 함수 (F i j)
자치정도 (S i j)
판정 결과 (A i j)
상관 매트릭스 선택의 결과를 입력
귀속도 Ki의 산출

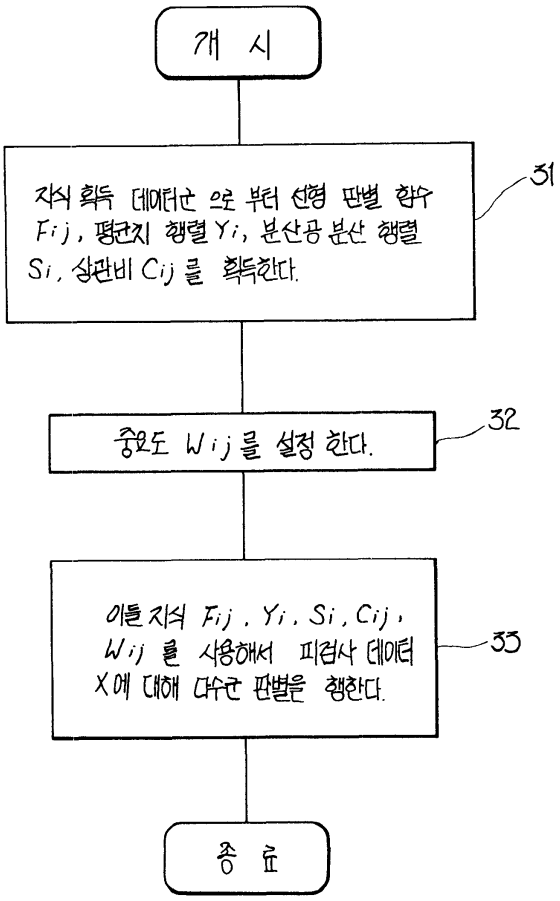
도면 4


자식 획득 데이터군
Di, ..., Dn
피검사 데이터
x
판별 분석
산출 판별 함수
Fij
2군 판별 처리
마하라노비스 거리 산출
평균치, 분산용 분산
Yi, Si
상관비
Cij
20
2차 판별 결과
Aij
판별 결과 격지 정도
Sij
2군 판별 정도
Rij
22
귀속도
Kij
중요도
Wij

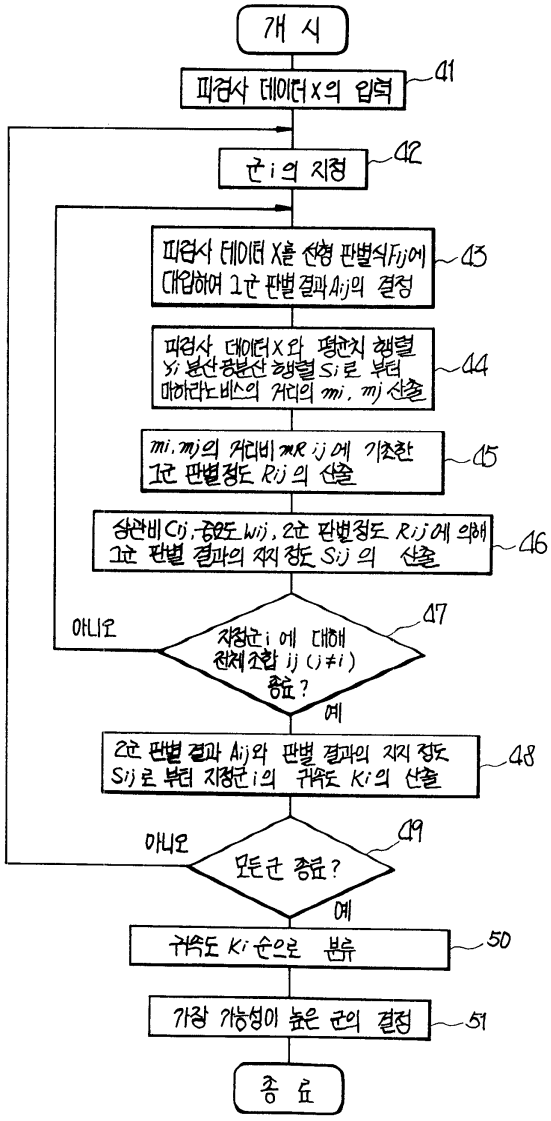
도면 5



도면 6


개 시
지식 획득 데이터군 으로 부터 선형 판별 함수 Fij, 평균치 행렬 Yi, 분산공 분산 행렬 Si, 상관비 Cij 를 획득한다.
31
중요도 Wij를 설정 한다.
32
이들 지식 Fij, Yi, Si, Cij, Wij 를 사용해서 피검사 데이터 X에 대해 다수군 판별을 행한다.
33
종 료

도면 7


개 시
피검사 데이터 x의 입력
41
군 i의 지정
42
피검사 데이터 X를 선형 판별식 Fij에 대입하여 1군 판별 결과 Aij의 결정
43
피검사 데이터 X와 평균치 행렬 Yi 분산공분산 행렬 Si로 부터 마하라노비스의 거리의 mi, mj 산출
44
mi, mj의 거리비 mRij에 기초한 1군 판별정도 Rij의 산출
45
상관비 Cij, 중요도 Wij, 2군 판별정도 Rij에 의해 1군 판별 결과의 지지 정도 Sij의 산출
46
아니오
지정군 i에 대해 전체조합 ij (j≠i) 종료?
47
예
2군 판별 결과 Aij와 판별 결과의 지지 정도 Sij로 부터 지정군 i의 귀속도 Ki의 산출
48
아니오
모든군 종료?
49
예
귀속도 Ki 순으로 분류
50
가장 가능성이 높은 군의 결정
51
종 료

도면 8

< 2군 판별 분석에 의한 진단 >

데이터 베이스 파일: mda 3. dat
기록 No. 153 ID No. 982
진단 분류 5 2차성 빈혈 ③

WBC	: 4.70	[x10^3/ul]
RBC	: 3.15	[x10^6/ul]
HGB	: 10.60	[g/dl]
HCT	: 33.40	[%]
MCV	: 106.00	[fl]
MCH	: 33.60	[pg]
MCHC	: 31.70	[g/dl]
PLT	: 241.00	[x10^3/ul]
RDW-SD	: 68.90	[fl]

< 가능성이 높은순 >
1. 혼합성 빈혈 ① : 0.846
2. 2차성 빈혈 ③ : 0.846
3. 재생 불량성 빈혈+MDS : 0.769

사람 세미아
혼합성 빈혈 ①
2차성 빈혈 ③
재생불량성 빈혈+MDS
혼합성 빈혈 ②
거대아구성 빈혈 제외
정상(성인)
소아 (정상)

지지 모드 : 예측

k:	0.077
	0.154
	0.231
	0.769
	0.846
	0.769
	0.846
	0.000
	0.462
	0.385
	0.615
	0.000
	0.538
	0.308

도면 9

< 2군 판별 분석에 의한 결과 >
데이터 베이스 파일: mda3.dat
개체 No. 133
검사 분류 : 5
ID No. 982
2차성 빈혈 ③
WBC : 4.70 [×10^3/㎕]
RBC : 3.15 [×10^6/㎕]
HGB : 10.60 [g/dl]
HCT : 33.40 [%]
MCV : 106.00 [fl]
MCH : 33.60 [pg]
MCHC : 31.70 [g/dl]
PLT : 241.00 [×10^3/㎕]
RDW-SD : 68.90 [fl]
< 가능성의 높은순 >
1. 2차성 빈혈 ③ : 0.623
2. 용혈성 빈혈 ① : 0.563
3. 재생 불량성 빈혈 + MDS : 0.339
지지 모드 : Kij
Ki
0.000
0.103
0.036
0.294
0.623
0.339
0.563
0.000
0.210
0.210
0.334
0.000
0.144
0.062
사람 세이아
철결핍성 빈혈 ①
2차성 빈혈 ②
2차성 만성 빈혈 ③
재생 불량성 빈혈 + MDS
용혈성 빈혈 ①
골수이상성 증후군 빈혈 ②
골수 섬유증
정상 (성인)
순아 (정상)

도면 10

< 2단 판별 분석에 의한 리턴 >
데이터 베이스 파일: mda3.dat
개별 No : 1533 ID No. : 982
진단 분류 : 5 2차성 빈혈 ③
WBC : 4.70 [×10^3/ul]
RBC : 3.15 [×10^6/ul]
HGB : 10.60 [g/dl]
HCT : 33.40 [%]
MCV : 106.00 [fl]
MCH : 33.60 [pg]
MCHC : 31.70 [g/dl]
PLT : 241.00 [×10^3/ul]
RDW-SD : 68.90 [fl]
< 가능성이 높은순 >
1. 2차성 빈혈 ③ : 0.563
2. 응혈성 빈혈 ① : 0.491
3. 재생 불량성 빈혈+MDS : 0.351
지지 모드 : Rij × Cij
Ki
: 0.000
: 0.098
: 0.253
: 0.563
: 0.351
: 0.491
: 0.000
: 0.207
: 0.195
: 0.328
: 0.000
: 0.136
: 0.052
사람 세미아
철결핍성 빈혈
2차적 빈혈 ①
2차 성 빈혈 ②
2차성 불량성 빈혈 ③
응혈성 빈혈 ①
재생불량성 빈혈 ②+MDS
간질이 구성 빈혈
출혈관성 빈혈
재대출중
다출중
정상 (심인)
소아 (정상)