

---

# 개선된 ART2 알고리즘을 이용한 자가 질병 진단 시스템

김광백\* · 우영운\*\* · 김주성\*\*\*

## Self Disease Diagnosis System Using Enhanced ART2 Algorithm

Kwang-Baek Kim\* · Young Woon Woo\*\* · Ju Sung Kim\*\*\*

### 요 약

본 논문에서는 개인의 건강 상태를 일련의 과정에 따라 스스로 파악하여 전문 의료 관리에 대한 접근 방향의 결정을 돕고 전문의가 쉽게 새로운 질병 및 증상을 학습 할 수 있도록 하는 자가 질병 진단 시스템을 제안하였다. 제안된 자가 진단은 보건 복지부에 제출된 '한국인이 부담을 가지는 질병' 관련 보고서와 의료 콘텐츠 'Engel Pharm'을 참조하여 선정한 60가지의 질병과 각 질병에 대한 대표 증상 161가지를 이용하여 질병을 도출한다. 개선된 ART2 학습 알고리즘을 적용하여 질병 종류를 군집화하고 각 질병의 증상에 관련된 질의 결과를 입력 벡터로 제시하여 사용자의 건강 상태를 진단함으로써 자신의 건강에 대한 정보를 제공한다.

### ABSTRACT

In this paper, we have proposed a self disease diagnosis system for ordinary persons to help the decision of access methods to a specialized medical management, and for medical specialities to discover new diseases and their symptoms easily, using verification of an individual's health status by a series of processes performed by oneself. In the proposed self disease diagnosis system, illness is decided by 60 kinds of diseases selected using the report called 'Diseases that Koreans take seriously' published by Ministry of Health & Welfare and medical contents called 'Engel Pharm', and also using 161 representative symptoms for the 60 kinds of diseases. An individual's health information is extracted by diagnosis of one's health status by a clustering of the 60 kinds of diseases using enhanced ART2 algorithm and input vectors from the results of questions for symptoms of each disease.

### 키워드

자가 질병 진단, 대표 증상, ART2 알고리즘, 군집화  
Self Disease Diagnosis, Representative Symptoms, ART2 Algorithm, Clustering

## I. 서 론

사회구조의 다변화와 초고령화 사회로의 진입은 단순한 수명의 연장보다는 건강한 삶의 영위에 대한 중요

성을 강조하게 되었고 이는 많은 현대인들이 간편하고 친숙한 방법으로 자신의 건강을 돌볼 수 있는 방법에 대한 사회적 요구를 증대시켜왔다. 특히 건강과 관련된 의료정보교류에서 시간과 공간의 제약을 극복할 수 있는

---

\* 신라대학교 컴퓨터정보공학부  
\*\* 동의대학교 멀티미디어공학과  
\*\*\* 신라대학교 간호학과

방법의 모색이 각종 정보통신매체의 발달과 더불어 그 필요성이 더욱 커지고 있다[1,2]. 따라서 본 논문에서는 자신의 건강 상태를 스스로 평가하고 관리 할 수 있도록 유도하는 자가 질병 진단 시스템을 제안한다.

제안된 자가 질병 진단 시스템은 보건 복지부에 제출된 ‘한국인이 부담을 가지는 질병’ 관련 보고서[3]와 의료 콘텐츠 ‘Engel Pharm’를 참조하여 일반인에게 가장 흔하지만 가장 부담이 큰 질병 60종을 선정하고 각 질병들에 대한 증상 161가지, 그리고 선정된 질병들과 신체와의 연관성을 고려하여 증상들의 특징이 뚜렷하게 나타나는 이상 부위 22가지를 선정한 후, 데이터베이스(Database)로 구축한다. 그리고 본 논문에서 제안한 개선된 ART2 학습 알고리즘을 적용하여 각 질병에 대한 대표 증상들을 분류하고 질병을 도출한다.

## II. 본 론

### 2.1. 질병 및 증상 조사

본 논문에서는 보건복지부에 제출된 ‘한국인의 질병 부담 2005년 보고서’에 근거하여 가장 흔하지만 의료관리 부담이 큰 질병과 의료 콘텐츠 ‘Engel Pharm’를 참조하여 질병 60종을 선정하였다.

표 1은 질병 장애로 인한 부담 순위에 대해서 상위 10위까지 나타낸 것이고 표 2는 연합뉴스에서 보도된 한국인의 10대 사망 원인과 지난 10년간의 증감추세를 나타낸 것이다[4,5].

각 질병에 대한 수집과정은 그림 1과 같다. 질병 증상에 대한 질의는 가장 뚜렷하게 나타나는 증상들을 대표 증상으로 분류하여 구현하고 이상 부위는 증상과 신체 부위의 연관성을 고려하여 채택한다. 이상 부위와 질의를 선정하는 과정은 그림 2와 같다.

표 1. 질병 장애로 인한 부담 상위 10위  
Table 1. Top 10 Serious Diseases

| 순위  | 질병          |
|-----|-------------|
| 1위  | 당뇨병         |
| 2위  | 위·십이지장 궤양   |
| 3위  | 천식          |
| 4위  | 뇌졸중 등 뇌혈관질환 |
| 5위  | 휴머티즘성 관절염   |
| 6위  | 우울증         |
| 7위  | 심근경색증위협심증   |
| 8위  | 피부질환        |
| 9위  | 정신분열증       |
| 10위 | 중이염         |

표 2. 한국인의 10대 사망 원인  
Table 2. Top 10 Causes for Death of Koreans

|         | 1992 |       | 2002 |       | 증감    |       |
|---------|------|-------|------|-------|-------|-------|
|         | 순위   | 사망률   | 순위   | 사망률   | 순위    | 사망률   |
| 암       | 1    | 110.7 | 1    | 130.7 | 20.0  | 35.3% |
| 뇌혈관질환   | 2    | 80.3  | 2    | 77.2  | -3.1  | 20.8% |
| 심장질환    | 3    | 43.0  | 3    | 37.2  | -5.8  | 10%   |
| 당뇨병     | 7    | 13.5  | 4    | 25.1  | 11.6  | 6.7%  |
| 만성기도 질환 | 8    | 12.9  | 5    | 22.6  | 9.7   | 6.1%  |
| 간질환     | 5    | 31.6  | 6    | 22.0  | -9.6  | 5.9%  |
| 자살      | 10   | 9.7   | 7    | 19.13 | 9.4   | 5.1%  |
| 교통사고    | 4    | 34.4  | 8    | 19.12 | -15.3 | 5.1%  |
| 고혈압성 질환 | 6    | 27.4  | 9    | 10.6  | -16.8 | 2.8%  |
| 호흡기결핵   | 9    | 9.9   | 10   | 6.6   | -3.3  | 1.8%  |

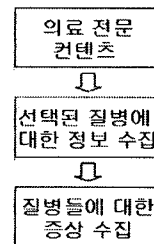


그림 1. 질병과 증상의 수집 과정  
Fig. 1. Process for Collecting Diseases and Symptoms

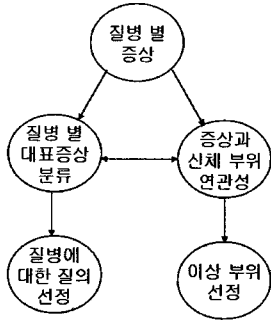


그림 2. 이상 부위와 질의 선정 과정  
Fig. 2. Process for Selecting Abnormal Parts and Questions

2.2. 데이터베이스 구조

본 논문에서 제안한 데이터베이스 구조는 그림 3과 같다. 질병과 증상의 수집 과정에서 선정된 60종의 질병과 161가지 증상에 대해 각 테이블을 설계하고, 각 테이블의 일차 키로 구성된 질병 관련 증상 테이블을 설계한다. 이 테이블은 각 질병들이 가지는 대표 증상들을 저장한다.

또한 그림 2에서와 같이 이상 부위와 질의 선정에서 선정된 22가지의 이상 부위를 저장하기 위해 증상 부위 테이블을 설계한다.

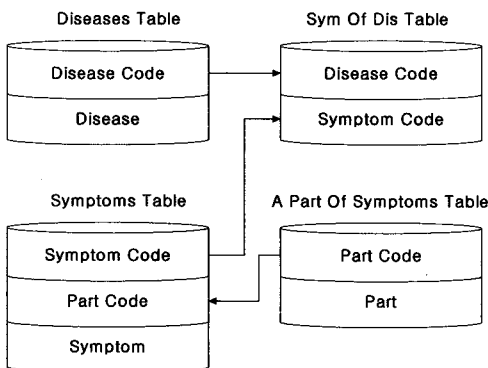


그림 3. 질병과 증상에 대한 데이터베이스 구조  
Fig. 3. Database Structure for Diseases and Symptoms

2.3. 제안된 자가 질병 진단 시스템

본 논문에서 제안된 자가 질병 진단 시스템은 이상 부위와 그에 따른 증상을 질의로 선택한다. 선택된 질의 결과에 따라 각 질병들에 대해 유사도를 측정하고 측정된

유사도를 이용하여 최종 질병을 도출한다. 제안된 자가 진단 시스템의 구성은 그림 4와 같다.

질의 선택에 따른 질병 도출은 총 60종의 질병에 대해서 각 질병들의 증상에 대한 질의 결과를 입력 벡터로 제시한 후, 본 논문에서 제안한 개선된 ART2 알고리즘을 적용하여 질병을 도출한다.

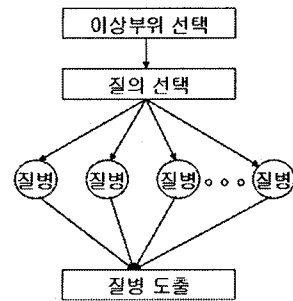


그림 4. 제안된 자가 질병 진단 시스템  
Fig. 4. The Proposed Self Disease Diagnosis System

질병 도출 과정은 그림 5와 같다. 유사한 질병을 나타내는 층은 음영이 어두운 색상 일수록 입력되는 질의들 중 많은 질의를 포함하는 군집을 의미한다.

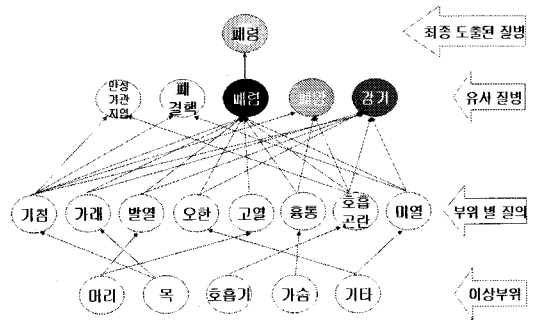


그림 5. 최종 질병 도출 과정  
Fig. 5. Process for Extracting Final Disease

표 3은 질병들 중에서 폐렴과 감기에 대한 부위별 증상을 나타 낸 것이다.

표 3. 폐렴과 감기의 부위별 증상  
Table 3. Symptoms for Pneumonia and Cold

| 부위 \ 질병 | 폐렴             | 감기             |
|---------|----------------|----------------|
| 머리      | 발열(미열)         | 발열(미열)         |
|         | 고열             | .              |
| 기타      | 오한             | 오한             |
| 호흡기     | 호흡곤란           | 호흡곤란           |
| 가슴      | 흉통             | .              |
| 목       | 통증<br>(붓고 따가움) | 통증<br>(붓고 따가움) |
|         | 기침             | 기침             |
|         | 가래(객담)         | 가래(콧물)         |

2.4. 개선된 ART2 학습 알고리즘

제안된 자가 질병 진단 시스템은 사용자의 증상과 질병의 유사도를 측정하기 위하여 개선된 ART2 학습 알고리즘을 적용한다. ART2 학습 알고리즘은 입력 패턴에 대하여 목표값 없이 자율 학습을 통해 군집화하는 클러스터링 방법이다. ART2 모델은 임의의 입력 패턴에 대해 이미 학습된 패턴들을 잊지 않고 새로운 패턴들을 학습할 수 있는 안정성(stability)과 적응성(plasticity)을 가지면서 실시간 학습이 가능하여 저속 및 고속을 지원할 뿐만 아니라 지역 최소화(local minima) 문제가 발생하지 않는 장점을 갖는다[6]. ART2의 연결 가중치 변화는 모든 입력 패턴의 평균값을 취함으로써 클러스터 생성에 고르게 반응하게 된다. 그러나 ART2 알고리즘은 경계 변수 설정에 따라 클러스터의 수가 달라진다는 단점이 있다. ART2 알고리즘에서 경계 변수는 패턴들을 클러스터링하기 위한 반지름 값으로서 임의의 패턴과 저장된 패턴과의 불일치 허용도를 결정하는 정적인 값이다. 그러나 자가 진단에 적용하기 위해서는 경험적으로 경계 변수를 설정해야하므로 어려운 점이 많다[7]. 특히 질병과 같이 유사한 특징을 갖는 서로 다른 패턴들이 많은 경우에는 경계 변수를 설정하는 것이 매우 어렵다.

따라서 본 논문에서는 ART2의 단점인 경계 변수를 동적으로 조정하기 위해 총체화된 값이 모든 입력 중 가장 작은 값보다 크지 않은 성질을 갖는 퍼지 교점속 연산자 중에서 식 1과 같은 Yager의 교점속 연산자[8]를 적용하여 경계 변수를 동적으로 조정하는 개선된 ART2 학

습 방법을 제안하여 자가 진단의 질병 군집화에 적용한다.

$$\mu_{x1 \cap x2} = 1 - \min \{1, ((1 - \mu_{x1})^p + (1 - \mu_{x2})^p)^{\frac{1}{p}}\} \quad (1)$$

여기서, 식 (1)은  $p \in (0, \infty)$  값에 따라 단조 감소하는 함수이다.

자가 진단에서 질병을 군집화 하는데 적용된 개선된 ART2 알고리즘의 학습 과정은 그림 6과 같다.

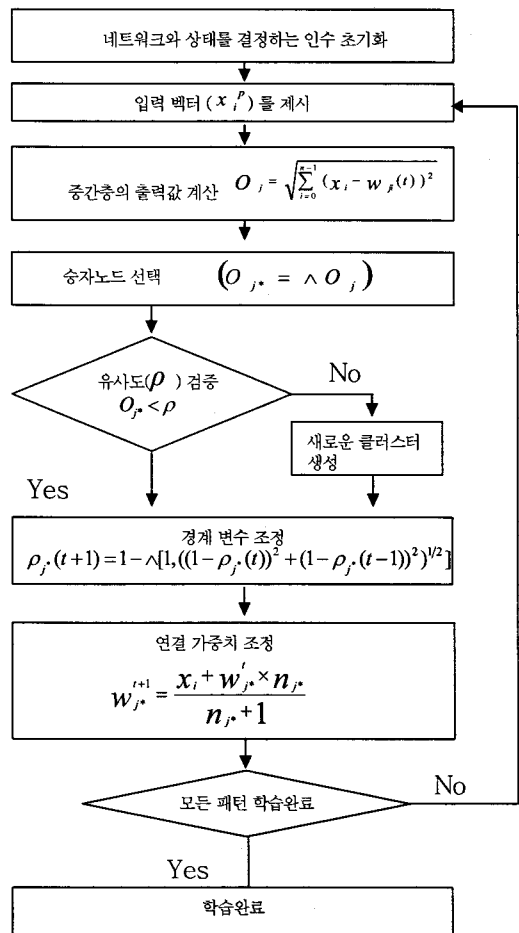


그림 6. 개선된 ART2 알고리즘의 학습 과정  
Fig. 6. Process for Learning of Enhanced ART2 Algorithm

### III. 실험 및 결과분석

실험을 위하여 Intel Pentium-IV 2GHz CPU와 256MB RAM이 장착된 IBM 호환 PC상에서 JAVA JDK 1.6.0과 Oracle을 연동하여 제안된 시스템을 구현하였다

제안된 자가 질병 진단 시스템은 각 질병에 대한 대표 증상을 학습하는 단계와 질병을 진단하는 단계로 구분되며 증상을 학습하는 시스템을 구현한 화면은 그림 7과 같다.

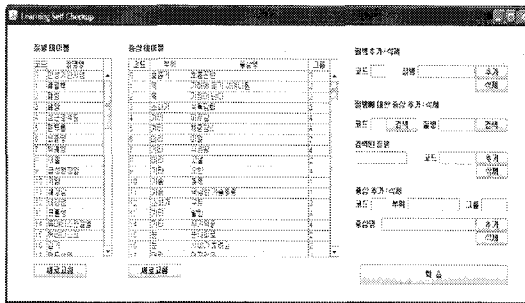


그림 7. 제안된 자가 진단 학습 시스템 화면  
Fig. 7. Window of the Proposed Self Disease Diagnosis System

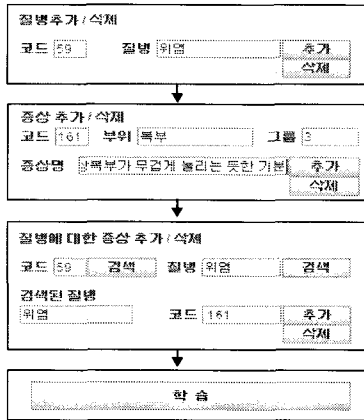


그림 8. 새로운 질병에 대한 학습 과정  
Fig. 8. Process for Learning of New Disease

개선된 ART2 학습 알고리즘은 이미 학습된 패턴들을 잃지 않고 안정성 있게 새로운 패턴을 학습할 수 있기 때문에 새로운 질병에 대한 추가 및 관리가 용이한 장점을 가지고 있다. 따라서 본 논문에서는 전문가가 새로운 질병에 대해 쉽게 질병을 추가 및 수정 할 수 있는 각 질병

에 대한 증상을 학습하는 시스템을 제안함으로써, 사용자 하여금 정확하고 최신의 질병까지 진단할 수 있다. 새로운 질병에 대한 학습 과정은 그림 8과 같고 학습이 완료된 이후 추가된 질병, 증상, 질병과 증상과의 관계, 증상부위는 구축된 데이터베이스에 저장된다.

본 논문에서 제안한 자가 질병 진단 시스템의 구성도는 그림 9와 같고 구현한 자가 질병 진단 시스템은 그림 10과 같다.

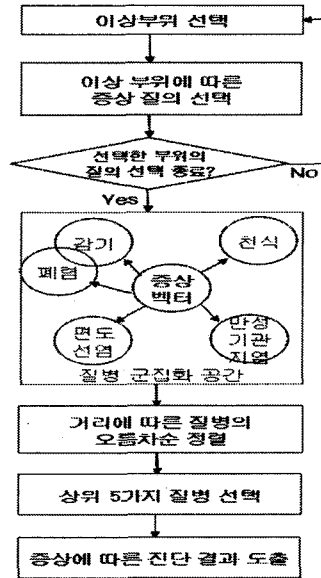


그림 9. 제안된 자가 질병 진단 시스템의 구성도  
Fig. 9. Diagram of the Proposed Self Disease Diagnosis System

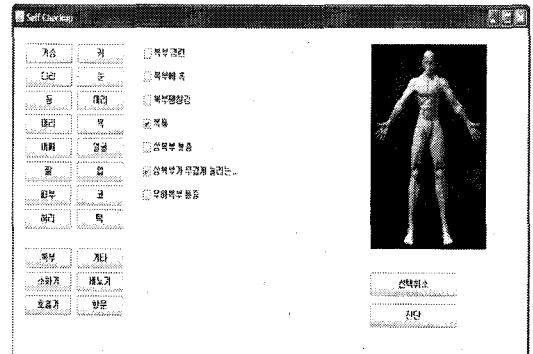


그림 10. 제안된 자가 질병 진단 시스템의 구현  
Fig. 10. Implementation of the Proposed Self Disease Diagnosis System

본 논문에서 구축한 데이터베이스로부터 질병, 증상, 증상 부위를 검색하여 각 부위에 대응하는 증상들을 사용자에게 제공하고, 사용자는 자신이 느끼는 불편한 증상들을 각 부위별로 선택한다. 질의가 선택되면 그 결과를 특징 벡터로 제시하여 최종적으로 사용자가 가질 수 있는 질병 중 가장 가능성이 높은 5종의 질병을 도출한다.

제안된 자가 질병 진단 시스템은 증상에 따른 질병을 도출하기 위해 개선된 ART2 학습 알고리즘을 적용하였다. 기존의 FCM과 개선된 FCM 알고리즘을 이용한 자가 진단 시스템[9]은 그 질병의 수와 질병에 대한 증상의 수가 현저히 적었다. 따라서 학습 과정에서 각 패턴에 대한 군집화가 부정확하고, 인식 결과 또한 정확성이 낮았다.

이러한 단점을 개선하기 위하여 본 논문에서는 질병 60종, 증상 161가지를 학습하였고, 그림 11, 12, 13은 같은 증상에 대해 기존의 FCM 알고리즘, 개선된 FCM 알고리즘 그리고 본 논문에서 제안한 개선된 ART2 알고리즘을 적용한 결과 화면이다. 실험에 적용한 증상은 가래, 기침, 오한, 호흡곤란이다.

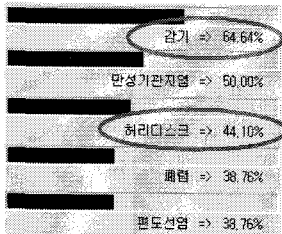


그림 11. 기존의 FCM 알고리즘  
Fig. 11. Conventional FCM Algorithm

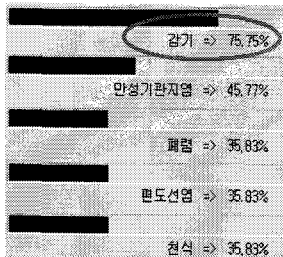


그림 12. 개선된 FCM 알고리즘  
Fig. 12. Enhanced FCM Algorithm

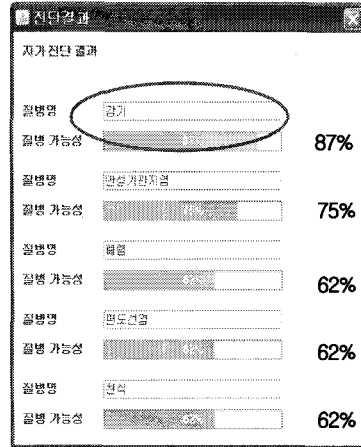


그림 13. 개선된 ART2 알고리즘을 적용한 자가 질병 진단 시스템 결과 화면 #1  
Fig. 13. Result Window of the Self Disease Diagnosis System Using Enhanced ART2 Algorithm #1

그림 11에서는 증상들과의 연관성이 적은 질병인 ‘허리디스크’가 3번째로 높은 유사도를 나타냈다.

반면 개선된 FCM 알고리즘은 같은 증상에 대해 연관성이 큰 질병들만 도출하였지만 정확도가 현저히 낮아, 진단 결과를 신뢰하기 어렵다. 그러나 그림 13에서와 같이 본 논문에서 제안한 개선된 ART2 알고리즘은 연관성이 적은 질병들은 도출되지 않았고, 특히 다른 학습 알고리즘에 비해 그 질병의 정확성이 개선되어 진단 결과를 신뢰할 수 있고 학습 과정에서의 군집화가 기존의 방식에 비해 개선되었음을 확인할 수 있다.

그림 14, 15는 만성 설사, 탈수 상태, 구토, 복통 증상에 대한 결과를 도출한 화면이다. 기존의 개선된 FCM 알고리즘과 본 논문에서 제안한 개선된 ART2 알고리즘을 이용한 자가 질병 진단 시스템은 단순히 군집 중심과 증상들 간의 거리만을 이용하기 때문에 연관성이 적은 질병이 포함될 가능성이 있다. 그러나 본 논문에서는 질병 및 증상 수를 두 배로 늘려 학습함으로써 이를 개선하였다.

그러나 학습 과정에서 질병에 대한 증상의 수가 현저히 적고, 진단 과정에서 증상의 수를 현저히 적게 입력하여 진단할 경우에는 그림 15와 같이 질병들이 각각 다른 대표 증상들을 가짐에도 불구하고 질병 가능성이 같은 값을 가지는 신뢰할 수 없는 진단 결과를 초래하는 단점이 있다.

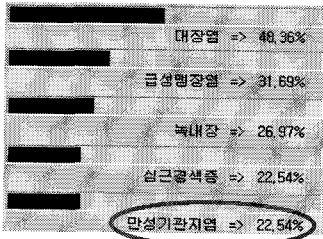


그림 14. 개선된 FCM 알고리즘의 연관성이 적은 결과를 도출한 화면

Fig. 14. Low Relevant Result of Enhanced FCM Algorithm

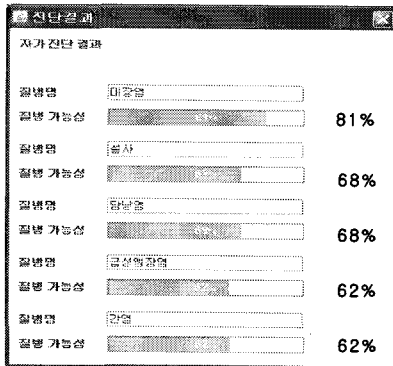


그림 15. 개선된 ART2 알고리즘을 적용한 자가 질병 진단 시스템 결과 화면 #2

Fig. 15. Result Window of the Self Disease Diagnosis System Using Enhanced ART2 Algorithm #2

#### IV. 결론

본 논문에서는 전문 의료 지식이 부족한 개인이 시공간의 제약에서 벗어나 의료인과의 직접 대면에 앞서 자신의 건강 상태를 파악하여 추후 전문 의료 관리에 대한 접근 방향의 결정을 돕는 자가 질병 진단 시스템과 이를 이용하기 위해 전문가가 쉽게 새로운 질병 및 증상을 관리할 수 있는 학습 시스템을 제안하였다.

제안된 자가 질병 진단 시스템은 개선된 ART2 알고리즘을 적용하여 군집화된 질병들과 사용자에게 제공한 증상 질의 결과간의 유사도를 측정하여 질병을 도출하였다.

기존의 개선된 FCM 알고리즘은 질병과 증상의 수가 현저히 적고 진단 결과 또한 현저히 낮았다. 그러나 본 논문에서 제안한 개선된 ART2 알고리즘을 이용한 자가 질병 진단 시스템은 질병과 질병에 대한 증상의 수를 두 배로 늘려 질병 정확성을 향상시켰고, 개선된 ART2 알고리즘의 장점인 임의의 입력 패턴에 대해 이미 학습된 패턴들을 잊지 않고 새로운 패턴들을 학습할 수 있는 안정성과 적응성, 그리고 실시간 학습이 가능한 점을 이용하여 효율적으로 학습하는 자가 진단 학습 시스템을 제안하였다.

그러나 증상에 대한 질병 결과를 도출하는 과정에서 증상과 질병간의 유사도를 단지 유클리드 거리만을 적용하여 측정하기 때문에 증상과 연관성이 적은 질병도 결과 중의 하나로 도출될 가능성이 발생하였으며, 학습 과정에서 질병에 대한 증상의 수가 현저히 적고, 진단 과정에서 증상의 수를 현저히 적게 입력하여 진단 할 경우에는 정확한 질병을 도출할 수 없는 경우가 발생하였다.

따라서 향후 연구과제로 각각의 질병과 유사한 질병들 간의 군집화를 개선하기 위해 소속 정도를 계산하고 퍼지추론 규칙을 적용하여 질병 도출의 정확성을 개선할 것이다. 또한 신체 부위와 증상들의 상관관계를 심도 있게 연구하여 제안된 자가 질병 진단 시스템을 실용화할 것이다.

#### 참고문헌

- [1] Berry D. L., Trigg L. J., Lober W. B., Karras B. T., Galligan M. L., Austin-Seymour M., and Martin S., "Computerized symptom and quality-of-life assessment for patients with cancer part I; development and pilot testing," *Oncology nursing forum*, Vol. 31, No. 5, pp.75-83, 2004.
- [2] Mullen K. H. Berry D. L., and Zierler B. K., "Computerized symptom and quality-of-life assessment for patients with cancer part II acceptability and suability," *Oncology nursing forum*, Vol. 31, No. 5, pp.84-89, 2004.

- [3] <http://www.mohw.go.kr>
- [4] <http://www.yonhapnews.co.kr>
- [5] 정태준, 질병 자가진단, 장원출판사, 1995.
- [6] Gail A. Carpenter and Stephen Grossberg, "ART2: self-organization of stable category recognition for analog input patterns," OPTICS, Vol.26, No23, pp.4919-4930, 1987.
- [7] K. B. Kim and S. Park, "Passport Recognition Using Enhanced ART2-based RBF Neural Networks," International Journal of Computer Science and Network Security, Vol.6, No.7A, pp.12-17, 2006.
- [8] H. J. Zimmerman, Fuzzy Set Theory and Its Applications, Kluwer Academic Publishers, 1991.
- [9] 정병희, 전태률, 김광백 "개선된 FCM 알고리즘을 이용한 자가 진단 시스템," 한국해양정보통신학회 인공지능 및 지능정보시스템 학술대회 논문집, pp.143-149, 2006.

저자소개



김 광 백(Kwang-Baek Kim)

1999년 부산대학교 전자계산학과 (이학박사)

1997년~현재: 신라대학교 컴퓨터정보공학부 부교수

2005년~현재 한국멀티미디어학회 이사 및 논문지 편집위원

2005년~현재 한국해양정보통신학회 이사 및 논문지 편집부위원장

※ 관심분야 : Neural Networks, Image Processing, Fuzzy Logic, Medical Imaging and Biomedical System, Support Vector Machines



우 영 운(Young Woon Woo)

1989년 2월 연세대학교 전자공학과 (공학사)

1991년 8월 연세대학교 본대학원 전자공학과(공학석사)

1997년 8월 연세대학교 본대학원 전자공학과(공학박사)

1997년 9월~현재 동의대학교 멀티미디어공학과 교수

2007년~현재: 한국해양정보통신학회 국제이사

※ 관심분야 : 인공지능, 패턴인식, 퍼지이론, 의료정보



김 주 성(Ju Sung Kim)

2002년 부산대학교 간호학과 (간호학박사)

2004~2006년 Postdoctoral fellow,

University of Illinois at Chicago, USA

2006년~현재 신라대학교 간호학과 전임강사

※ 관심분야 : family resemblance of bone mineralization, care burden of family caregivers