

논문 2011-48CI-2-4

# 결정트리 데이터마이닝을 이용한 족부 임상 진단

## (Podiatric Clinical Diagnosis using Decision Tree Data Mining)

김진호\*, 박인식\*, 김봉옥\*\*, 양윤석\*\*\*, 원용관\*\*\*\*, 김정자\*\*\*\*\*

(Jin-ho Kim, In-sik Park, Bong-ok Kim, Yoon-seok Yang, Yonggwan Won, and Jung-ja Kim)

### 요약

최근 건강에 대한 관심이 고조되면서 발과 다리에 대한 진단, 치료, 예방의 전반적인 진료를 받고 있는 족부의학(Podiatry)이 주목받고 있지만 국내 연구는 미비한 실정이다. 또한 임상 데이터 분석에 있어 대부분의 기존 연구들은 기초 통계적인 방법에 근거한 정량분석만을 수행함으로써, 획득된 정보를 임상에 적용 하는데 있어서는 충분한 신뢰성을 보장할 수 없다. 임상 데이터 마이닝은 데이터마이닝의 다양한 분석 방법론을 이용하여 의료 현장에서 발생한 임상 데이터를 분석함으로써 전문가의 진단과 치료 과정의 결정에 도움을 주고 있다. 결정트리(Decision Tree) 알고리즘은 분석과정의 설명과 표현성이 좋고, 결과에 대한 해석이 편리하여 임상에서 적용하기가 용이하다. 본 연구에서는 신뢰성 있는 족부 임상 진단 평가를 위해 충남대학교병원 재활의학과 신발클리닉에 내원한 환자 1310명(남자:633명, 여자:677명)의 2620족(foot)을 대상으로 수집된 진료 데이터에 결정트리를 적용하여 22개의 족부 질환 인자에 따르는 15개의 족부 질환을 분류하고 그에 대한 64개의 진단 규칙을 탐사 하였다. 또한 5개의 클래스(영유아, 소아, 청소년, 노인, 전체)로 분류된 각 그룹들로부터 생성된 결정 트리를 통해 각 클래스의 질환 특성과 질환 주요 인자, 클래스 간 상관관계를 비교, 분석하였다. 탐사된 결과는 족부 임상 전문가의 의사결정에 더욱 정성적이고 유용한 선행적 지식을 제공할 것이고, 효과적이고 정확한 진단과 예측을 위한 임상 도구로써 사용될 수 있다.

### Abstract

With growing concerns about healthy life recently, although the podiatry which deals with the whole area for diagnosis, treatment of foot and leg, and prevention has been widely interested, research in our country is not active. Also, because most of the previous researches in data analysis performed the quantitative approaches, the reasonable level of reliability for clinical application could not be guaranteed. Clinical data mining utilizes various data mining analysis methods for clinical data, which provides decision support for expert's diagnosis and treatment for the patients. Because the decision tree can provide good explanation and description for the analysis procedure and is easy to interpret the results, it is simple to apply for clinical problems. This study investigate rules of item of diagnosis in disease types for adapting decision tree after collecting diagnosed data patients who are 2620 feet of 1310(males:633, females:677) in shoes clinic (department of rehabilitation medicine, Chungnam National University Hospital). and we classified 15 foot diseases followed factor of 22 foot diseases, which investigated diagnosis of 64 rules. Also, we analyzed and compared correlation relationship of characteristic of disease and factor in types through made decision tree from 5 class types(infants, child, adolescent, adult, total). Investigated results can be used qualitative and useful knowledge for clinical expert's, also can be used tool for taking effective and accurate diagnosis.

**Keywords :** Clinical foot diagnosis, Foot deformity, Classification, Decision Tree, Induction rule

\* 정회원, 전북대학교 헬스케어공학과

(Department of Healthcare Engineering, Chonbuk National University)

\*\* 정회원, 충남대학교병원 재활의학과

(Department of Rehabilitation, Chungnam National University College of Medicine)

\*\*\* 정회원, \*\*\*\*\* 정회원-교신저자, 전북대학교 바이오메디컬공학부

(Division of Biomedical Engineering, Chonbuk National University)

\*\*\*\* 정회원, 전남대학교 공과대학 전자컴퓨터공학과

(School of Electronics and Computer Engineering, Chonnam National University)

※ 이 논문은 2009년도 정부(교육과학기술부)의 재원으로 한국연구재단의 지원을 받아 수행된 기초연구사업(No. 2009-0077586) 및 한국학술진흥재단의 지원을 받아 수행된 연구임.(지방연구중심대학육성사업/헬스케어기술개발사업단)

접수일자: 2011년2월14일, 수정완료일: 2011년3월7일

## I. 서 론

최근 발의 통증 및 불편감을 주소로 내원하는 환자가 점차 늘어나고 있다. 이것은 유병율이 증가한 것이 아니라 생활의 여유가 증가하면서 그동안 등한시하였던 발의 문제에 관하여 관심을 가지게 된 것이다. 이에 족부에 대한 해부학, 병리학, 생리학 등의 기초의학은 물론 인체공학을 심도 있게 다루고, 발과 다리에 대한 진단, 치료, 예방의 전반적인 진료를 맡고 있는 전문적 의학의 한 분야로써 족부의학(podiatry)이 주목 받고 있다.

Kevin Deschamps등은 20명의 무지외반증 환자를 대상으로 12대의 동작분석 카메라를 사용하여 무지외반증의 운동학적(kinematics) 분석을 하였다. Yu-Chi Chen등은 11명의 평발 환자를 대상으로 보행 중 맨발인 상태, 신발을 신은 상태, 깔창(inalsole)이 들어간 신발을 신은 상태에서 각각 동작분석을 실시하여 평발 환자들에게 발 보조기가 미치는 영향에 대해 조사 하였다. Joshua Burns등은 52명(남자:19명, 여자:34명)의 피험자를 대상으로 체중 부하 상태에서 Foot Posture Index(FPI)를 이용하여 평발, 요족발 그리고 정상발인 사람으로 구분하고 발목의 배측굴곡(dorsiflexion)에 영향을 미치는 아킬레스건의 길이와 발의 형태 사이의 상관관계를 분석 하였다<sup>[1-3]</sup>.

위와 같이 대다수의 선행 연구들은 족부 질환에 대해 생체 역학이나 통계적 방법에 근거한 간단한 정량분석(quantitative analysis)만을 수행하는 실정이다. 이는 신체적 특성의 개인 차이 또는 질병의 정도에 따른 다양한 특성들을 설명하는데 있어 정성적인(qualitative) 해석을 제공하지 못함으로 인하여 획득된 정보를 임상에 적용하기에는 신뢰성 문제가 야기될 수 있다. 따라서 신체적 특성의 개인 차이와 질병간의 다양한 상호 연관성을 종합적이고 분석적으로 제공할 수 있도록, 관련 데이터의 지능적인 분석 방법론에 대한 연구가 필요하다<sup>[4]</sup>.

데이터 마이닝(data mining)은 대용량의 데이터베이스에서 숨겨진 규칙(rules)이나 패턴(patterns)을 탐사하여 이로부터 새로운 지식(knowledge)을 발견하기 위한 데이터 분석 방법으로서, 최근 다양한 데이터 마이닝 방법들을 응용하여 임상 의료 데이터 분석에 적용하고 있는 추세이다. 의료분야에서 데이터 마이닝이 사용된 초기의 예로서 1980년대 초에 스탠포드 대학교에서 수

행된 RX 프로젝트를 들 수 있다. 연구자들은 50명의 전신홍반루프스 환자들을 50회 이상 추적 관찰한 데이터베이스를 이용하여 여러 임상변수와 결과들 간의 시간적인 선후관계와 통계적 상관성을 연구한 바가 있다. Tiet등은 약물중독자들 중 자살시도를 생각해본 적이 있는 5671명을 대상으로 실제적으로 자살을 시도할 위험이 높은 환자군을 찾아내기 위해서 결정 트리 기법을 사용하였다<sup>[5-6]</sup>. 이아름 등은 신뢰성 있는 자세 제어 능력(Postural control ability) 평가를 위해서 측정된 훈련 데이터에 가중연관규칙(Weighted association rule) 탐사를 적용하여 자세 훈련 유형에 따른 근육 활성 패턴과의 연관성을 분석하였으며, 자세 제어 효율적인 재활 훈련 규칙을 탐사하였다<sup>[7]</sup>.

결정트리(decision tree)란 의사결정규칙을 나무구조 형태로 도표화하여 관심의 대상이 되는 집단을 여러 개의 소집단으로 분류하거나 예측을 수행하는 분석기법이다. 결정 트리 탐색과 모형화라는 두 가지 특성을 모두 가지고 있다고 할 수 있다. 결정트리는 분류 또는 예측의 과정이 나무구조에 의한 추론규칙에 의해서 표현되기 때문에, 신경망이나, 판별분석, 회귀분석 등 다른 방법에 비해서 그 과정을 쉽게 이해하고 설명할 수 있다. 또한 결정트리는 신경망과는 달리 규칙방식으로 표현이 가능하며, 규칙은 SQL과 같은 데이터베이스 언어로 표현될 수 있다는 장점을 가지고 있다<sup>[8-9]</sup>.

본 논문에서는 족부 임상 진단 평가를 위해 진단된 질환의 특성을 예측하고 해석이 용이한 장점을 가진 결정트리 기법인 C5.0 알고리즘을 사용하여 분석했다. 본 논문의 구성은 다음과 같다. I장의 서론에 이어 II장에서는 병원에 내원하는 환자들에게서 가장 흔하게 나타나는 족부 질환들의 특징에 대해 소개하고 결정트리 알고리즘에 대해서 설명하였다. III장에서는 실험구성을 위한 방법들과, 결정트리에 의해 도출된 결과를 분석하였다. IV장에서 결론 및 향후 계획에 대해 설명한다.

## II. 본 론

### 1. 족부 질환별 증상

발은 우리 몸 전체의 1/4에 해당되는 52개의 뼈와 60개의 관절, 214개의 인대, 38개의 근육을 비롯하여 수많은 혈관으로 구성되어 있다. 또한 발은 인체의 모든 중량을 받쳐주는 곳으로써 지면으로 부터의 충격을 흡수하고 분산시키는 역할을 한다. 이러한 발에 질환이 생

기면 정상적인 보행을 할 수 없게 되고 잘못된 보행 습관으로 인한 족압의 불균형은 어깨통, 요통, 무릎통증, 두통 등을 초래한다<sup>[10]</sup>.

#### 가. 평발

평발(pes planus)은 발바닥의 안쪽 아치가 비정상적으로 낮아지거나 소실되는 변형이다. 이차적으로 발뒤꿈치와 발 앞쪽이 바깥쪽으로 향하게 된다. 평발은 체중부하가 걸릴 때에는 발바닥이 편평해지지만 체중부하가 없으면 안쪽 아치가 나타나는 유연성 편평족과 체중부하와 관계없이 편평한 강직성 편평족으로 나눌 수 있다. 평발은 대부분 원인 불명으로 나타난다. 가장 흔한 원인은 관절이 과도하게 유연하여 발생하는 것이며 뇌성마비와 같은 신경 근육성 질환, 족근골 결합 같은 선천성 질환, 골절을 포함한 외상, 류마티스 관절염 및 후경골근기능장애로 발생하기도 한다. 평발의 증상으로는 외관상 발 안쪽 아치가 소실되고 발뒤꿈치가 바깥쪽으로 기울어진다. 신발 안쪽이 주로 닳으며, 장시간 보행 및 운동 시 통증을 느낄 수 있다. 소아의 평발은 발의 통증보다는 변형이 주된 증상으로 통증이 없는 것이 보통이지만, 청년기가 지나서 통증을 호소하는 경우가 있다. 소아의 유연성 편평족은 그대로 방치하여도 대부분의 경우 문제가 되지 않는다. 또한 통증이 있던 환자도 성장함에 따라 증세가 없어지는 것이 일반적이다. 통증은 장시간의 보행 및 운동 후 근육 피로, 족저근막의 과도한 신장 등과 관련하여 나타난다. 그러나 강직성 편평족인 경우에는 그 원인 질환에 의한 통증이 발생하거나 발을 자주 빠는 증상이 생기기도 하며, 특히 후경골근기능장애와 관련한 통증 및 족부 기능 장애가 초래되기도 한다.

#### 나. 요족

까치발 혹은 요족(pes cavus)은 평발과 반대로 아치가 높아서 발바닥 중에서 바닥과 닿지 않는 면이 넓고, 발등이 높으며 발 내측의 아치가 움푹 패인 형태로서, 보행 시 발의 외측에 몸무게를 실어 보행한다. 요족은 아킬레스건의 경직이 주요 원인으로 나타나며 여성의 경우 습관적으로 하이힐을 많이 신음으로서 근육의 변형에 의해 나타나기도 한다. 신경학적으로 소아에게서 가장 흔히 요족을 일으키는 원인으로는 척추 융합부전, 뇌성마비, 소뇌질환, 관절 구축증, 외상성 요족 변형등이 있다. 이러한 발은 충격 흡수 기능이 떨어지기 때문



그림 1. 발의 형태에 따른 분류 (A:평발, B:요족)

Fig. 1. Foot type classification.  
(A:Pes planus, B:Pes cavus)

에 하지, 발, 허리에 부담을 주게 되며 갈퀴족지변형(claw toe deformity)과 동반되는 경우가 많고 범위가 넓고 통증을 유발하는 발바닥의 굳은 살을 유발하기도 한다. 요족 증상을 치료하지 않게 되면 발 모양뿐 아니라 점차적으로 종아리 근육이 경직되어 발 전체에 심한 통증과 변형을 일으킬 수 있으며 심한 경우 통증으로 인해 정상적인 보행이 불가능해 질 수 있다. 그림 1에서 발의 형태에 따른 평발과 요족의 모습을 볼 수 있다.

#### 다. 경골 내염전

경골 내염전(tibial internal torsion)은 무릎의 대퇴골 내과(medial condyle)와 외과(lateral condyle)를 잇는 축에 비해서 발목의 경골 내과(medial malleolus of tibia)와 비골 외과(lateral malleolus of fibula)를 잇는 축이 내측으로 비틀어진 상태를 말한다. 이러한 경골 내염전은 소아에서 내족지 보행(in-toe gait)을 일으키는 가장 흔한 원인이다. 발생 원인으로는 무릎을 꿇고 앉거나 W자로 앉을 경우 나타나며 경골 내염전 뿐만 아니라 대퇴골 내회전, 중족골 내전까지 복합적으로 나타날 수 있다. 경골 내염전은 정상적인 성장 발달 과정에서 대부분 저절로 좋아지는 것으로 알려져 있으나 전통적으로 좌식 생활을 많이 하는 동아시아 국가에서는 성장 후에서도 경골 내염전이 많은 것으로 알려져 있다<sup>[11-13]</sup>. 경골 내염전이 있으면 내족지 보행을 하기 때문에 보행 시 자신의 발에 걸려 넘어져 상해의 위험이 있다. 또한 내족지 보행을 보상하기 위해 발가락 끝을 바깥쪽으로 걸으려 무리하게 노력하기 때문에 발의 회내를 일으켜 평발을 유발할 수 있다.

#### 라. 척추측만증

척추측만증(scoliosis)은 척추가 정면에서 보았을 때 옆으로 휘는 것을 지칭하나, 실제로는 단순한 2차원적인 기형이 아니라 추체 자체의 회전 변형과 동반되어 옆에

서 보았을 때에도 정상적인 만곡 상태가 아닌 3차원 기형 상태이다. 척추 측만증을 일으키는 원인으로는 척추 생성 과정에서 이상이 생겨 발생하는 선천성 척추 측만증이 있으며, 그 외에 중추 신경계나 신경학적으로 발생하는 신경 근육성 척추 측만증, 신경 섬유종에 의한 척추 측만증, 마르팡(marfan) 증후군 등 여러 증후군에 동반된 척추 측만증이 있다. 또한 구조적 척추 측만증으로써 평발, 경골내염전, 내·외반슬, 다리길이 차이와 같은 우리의 몸 전체를 지탱하는 발과 다리의 부정렬로 인해 골반이 기울게 되면 척추 측만증이 나타나게 된다. 척추 측만증이 있으면 척추가 부적절하게 하중을 전달하여 허리와 다리에 통증을 유발하여 성장 장애를 보이기도 한다.

2. 결정 트리

데이터 마이닝에서 연구된 결정 트리는 나무의 구조에 근거하여 의사결정규칙을 찾아내는 방법으로서 분류(classification)와 예측(prediction)을 수행하는 분석 방법이다. 이 방법은 분류 또는 예측의 과정이 나무구조에 의한 추론 규칙(induction rule)에 의해서 표현되기 때문에 분석자가 그 과정을 쉽게 이해하고 설명할 수 있다는 장점이 있다. 분석 결과는 ‘조건 A’ 이고 ‘조건 B’ 이면 ‘결과 집단 C’ 이다 라는 형태의 if-then 규칙으로 표현되므로 이해가 쉽다. 일반적으로 결정트리의 분석 절차는 다음의 4단계를 따른다.

- [단계 1] 의사결정나무의 형성 : 분석 목적과 자료의 구조에 따라 적절한 분리기준과 정지규칙을 고려하여 의사결정나무를 형성한다.
- [단계 2] 가지치기 및 축소 : 부적절한 추론규칙을 가지고 있거나 분류오류를 크게 할 위험이 있는 가지를 제거한다. 또한 가지들간의 관계를 조절하는 가지축소 과정을 거친다.
- [단계 3] 타당성 평가 : 위험도표, 이익도표, 또는 교차타당성 평가 등을 통하여 의사결정나무를 평가한다.
- [단계 4] 모형의 해석 및 예측 : 얻어진 의사결정나무 모형을 해석하고 예측모형을 설정한다.

결정트리를 유도하는 기본적인 알고리즘의 기본 개념은 하향식(top-down)으로 재귀적인 분할과 정복(divide-and-conquer)방법으로 트리를 구축하는 탐욕

**Algorithm : Generate\_decision\_tree.**

**Input :** The training sample, sample, represented by discrete-valued attributes; the set of candidate attributes attribute-list

**Output :** A decision tree

**Method :**

- (1) create a node N; (N: single node)
- (2) if sample are all of the same class, C then (C: each class)
- (3) return N as a leaf node labeled with the class C;
- (4) if attribute-list is empty then
- (5) return N as a leaf node labeled with the most common class in sample;
- (6) select test-attribute, the attribute among attribute-list with the highest information gain;
- (7) label node N with test-attribute;
- (8) for each known value  $a_i$  of test-attribute ( $a_i$ : partition the sample)
- (9) grow a branch from node N for the condition test-attribute =  $a_i$
- (10) let  $s_i$  be the set of sample for which test-attribute =  $a_i$
- (11) if  $s_i$  is empty then ( $s_i$ : a partition)
- (12) attach a leaf labeled with the most common class in samples;
- (13) else attach the node returned by Generate\_decision\_tree ( $s_i$ , attribute-list-test-attribute);

그림 2. 훈련된 샘플로부터 의사결정 트리를 유도하는 기본 알고리즘  
 Fig. 2. The basic algorithm for decision tree induction from trained samples.

(greedy) 알고리즘이다. 그림 2에 요약된 알고리즘은 가장 기본적인 결정트리 유도알고리즘인 ID3이다. 이 알고리즘은 C4.5와 C5.0의 기초가 된다<sup>[8-9]</sup>. 그림 2를 살펴보면, (1)단계에서 트리는 훈련 샘플들을 나타내는 단일 노드로 시작하여 만약 샘플들이 모두 같은 클래스라면 (2)단계와 (3)단계에서 노드는 잎이 되고, 해당 클래스로 레이블을 부여한다. 샘플들이 모두 같지 않은 클래스 이면 (6)단계와 같이 정보 이득(information gain)이라는 엔트로피(entropy) 기반 척도를 사용하여 샘플들을 각각 클래스로 가장 잘 분리하는 속성을 선택하기 위해 휴리스틱(heuristic)한 방법을 사용한다. 이 속성은 “검사” 또는 “의사결정” 속성이 된다. (8)단계~(10)단계에서 가지(node)는 검사 속성이 갖는 각각의 값에 대해 각각 생성되고 샘플들은 적절하게 분할된다. 분할된 각 샘플들에 대해 의사결정트리를 형성하기 위해서 앞의 과정들이 재귀적으로 수행되며 각 속성이 한 노드에서 발생하면 그 노드의 자손(descendent) 노드에서는 고려

될 필요가 없게 된다. 재귀적 분할은 다음 세 가지 조건들 중 하나가 참이 되면 정지한다. 첫 번째로 (2)단계와 (3)단계에서 주어진 노드의 모든 샘플들이 같은 클래스에 속하거나 두 번째로 (4)단계에서 샘플들을 더 분할하고 싶어도 더 남아있는 속성이 없는 경우, 세 번째로 (1)단계에서 가지에 대한 검사  $test\text{-}attribute = a_i$  를 만족하는 샘플이 없는 경우이다.

귀납적 학습방법인 C5.0은 Ross Quinlan<sup>[14]</sup>이 ID3에 이어 개발한 C4.5를 영국 ISL사의 데이터마이닝 툴인 Clementine에서 제일 먼저 채택하여 C4.5를 바탕으로 개발한 귀납적 학습방법이다. C5.0은 생성된 결정트리가 지나치게 많은 단계와 잎 노드를 가질 경우 학습된 결정트리의 일반화 능력을 제고하기 위하여 잎 노드를 제거하는 가지치기(pruning)를 하는데 유용하다. 이 방법은 예측 정확도를 향상시킬 수 있으나 오류율도 증가시키므로 오류기반 가지치기로 오류율의 증가를 통제한다. Clementine에서는 부스팅(Boosting)방법으로 잘못 분류된 개체들에 집중하여 분류 규칙을 새로 만드는 단계를 반복하고 각 단계의 분류 규칙을 결합하여 최종 모형을 산출한다. 부스팅 알고리즘은 다음의 3단계를 따른다.

[Step 1] 개체  $i$  에 초기가중치

$$w_i = 1/n (i = 1, \dots, n) \text{ 을 부여한다.}$$

[Step 2] 단계  $m (= 1, \dots, m)$  에서의 반복과정은 다음과 같다.

- 분류규칙  $G_m(x)$  를 만들어 개체  $x$  를 분류 한다.
- 오분류율  $err_m$  을 산출 한다.

$$a_m = \log(1 - err_m) / err_m$$

- 오분류개체  $i$ 에 대하여 가중치  $w_i$  를 다음과 같이 바꾼다.

$$w_i \leftarrow w_i \cdot \exp(a_m)$$

- 부스팅 시도 수만큼 Step2를 반복 한다.

[Step 3] 최종 분류규칙을 산출 한다

$$G(x) \leftarrow a_1 \cdot G_1(x) + \dots + a_m \cdot G_m(x)$$

부스팅 옵션의 사용은 시도 수만큼의 선택할 수 있는 규칙모형을 생성하고 C5.0의 예측정확도를 보다 월등하게 제시할 수 있다. 또한 귀납적 학습방법인 C5.0은 자동적으로 If-then rule을 생성해주는 규칙들의 집합 형태인 명제형 지식(Propositional Knowledge)으로 해석

및 분석이 용이하기 때문에 일반 사용자들도 쉽게 이해할 수 있다.

### III. 실험 방법 및 결과

#### 1. 데이터 수집 및 변수정의

본 연구에서는 충남대학교병원 재활의학과 신발클리닉에 내원한 1310명(남자:633명, 여자:677명)의 오른쪽과 왼쪽 2620족(foot)에 대한 진료 데이터를 수집하여 환자 데이터베이스를 구축 하였다. 진료 항목은 나이(age), 성별(sex), 골반(pelvis), 무릎(knee), 발목(ankle), 거골하관절(subtalar joint)의 내번(inversion)과 외번(eversion), 가동 범위(range of motion) 그리고 휴지기 종골각도(resting calcaneal stance position : RCSP)등의 오른쪽과 왼쪽 각각에 대해 총 50개의 진단 항목으로 구성되어 있으며 이 중에서 결정나무 모형 구성에 오류가 있는 항목을 제외한 총 22개의 진단 항목이 데이터 분석을 위한 변수로써 사용되었으며 각 진단 항목에 대한 설명은 표 1과 같다.

표 1. 족부 진단 인자

Table 1. Foot diagnostic feature.

| 항목                       | 내용               |
|--------------------------|------------------|
| Sex                      | 성별               |
| Age                      | 나이               |
| (Lt)Ankle Dorsiflexion   | 왼쪽 발목 배측굴곡 정도    |
| (Rt)Ankle Dorsiflexion   | 오른쪽 발목 배측굴곡 정도   |
| (Lt)STJ Inversion        | 왼쪽 거골하관절 내번각     |
| (Lt)STJ Eversion         | 왼쪽 거골하관절 외번각     |
| (Lt)STJ ROM              | 왼쪽 거골하관절 가동범위    |
| (Rt)STJ Inversion        | 오른쪽 거골하관절 내번각    |
| (Rt)STJ Eversion         | 오른쪽 거골하관절 외번각    |
| (Rt)STJ ROM              | 오른쪽 거골하관절 가동범위   |
| (Lt)FFtoRF               | 왼쪽 전·후족 각도       |
| (Rt)FFtoRf               | 오른쪽 전·후족 각도      |
| (Lt)RCSP                 | 왼쪽 휴지기종골각도       |
| (Rt)RCSP                 | 오른쪽 휴지기 종골각도     |
| (Lt)Pelvic Tilting       | 왼쪽 골반 기울임 방향     |
| (Rt)Pelvic Tilting       | 오른쪽 골반 기울임 방향    |
| (Lt)Pelvic Rotation      | 왼쪽 골반 회전 방향      |
| (Rt)Plevic Rotation      | 오른쪽 골반 회전 방향     |
| (Lt)Pelvic Elevation     | 왼쪽 골반 거상 각도      |
| (Rt)Pelvic Elevation     | 오른쪽 골반 거상 각도     |
| (Lt)Pelvic Trendelenburg | 왼쪽 골반 트렌델렌버그 각도  |
| (Rt)Pelvic Trendelenburg | 오른쪽 골반 트렌델렌버그 각도 |

2. 데이터 처리

본 연구는 각 항목에 의해서 진단된 결과, 즉 진단명에 대한 특성을 예측하는데 그 목적이 있다. 따라서 진단명을 목표 변수로, 그 외 진단 항목을 입력변수로 선택하여 결정트리 학습방법을 이용 분석을 실시하였다. 목표 변수가 되는 질환명의 항목들을 SPSS(SPSS, Inc., Chicago, USA)를 이용하여 1310명의 환자들의 2610족(foot)에 대한 족부 질환명에 대해 빈도 분석을 실시하였으며, 그 결과 유효퍼센트가 1.0%를 넘는 항목들만을 추출하여 진단명과 빈도수 그리고 유효 퍼센트를 표 2와 같이 정리 하였다.

자료 분석 결과 전체 1310명중 849명(남자:433명, 여자:416명)의 데이터가 최종적으로 분석을 위해 사용되었으며 모형을 구축하기 위해 훈련용 데이터와 테스트용 데이터를 7:3의 비율로 랜덤하게 추출 후 훈련 집합(training set)을 이용하여 모형을 구축하고 테스트 집합(test set)을 이용하여 모형의 비교와 최종적인 평가를 수행 하였다. 자료의 평가는 전체 849명과 연령별로 구분된 데이터에 대해 각각 평가 하였으며 Clementine 11.1(SPSS, Inc., Chicago, USA) 버전을 이용하여 C5.0 귀납 규칙(Induction Rule) 알고리즘을 적용 하였다. 표

표 2. 최종 선정된 특징 질환  
Table 2. The last selected feature disease.

|                                 | 내용              | 빈도  | 유효퍼센트(%) |
|---------------------------------|-----------------|-----|----------|
| (Lt)Pes cavus                   | 왼발 요족           | 29  | 3.4      |
| (Lt)Pes planus                  | 왼발 평발           | 45  | 5.3      |
| (Lt)Pes planus<br>(Rt)Pes cavus | 왼발 평발<br>오른발 요족 | 23  | 2.7      |
| (Lt)Pes planus<br>GCM tightness | 왼발 평발<br>비복근경직  | 14  | 1.6      |
| (Rt)Pes cavus                   | 오른발 요족          | 20  | 2.4      |
| (Rt)Pes planus                  | 오른발 평발          | 38  | 4.5      |
| (Rt)Pes planus<br>GCM tightness | 오른발 평발<br>비복근경직 | 14  | 1.6      |
| GCM tightness                   | 비복근경직           | 14  | 1.6      |
| Pes cavus                       | 양발 요족           | 74  | 8.7      |
| Pes cavus<br>GCM tightness      | 양발 요족<br>비복근경직  | 30  | 3.5      |
| Pes planus                      | 양발 평발           | 372 | 43.8     |
| Pes planus<br>GCM tightness     | 양발 평발<br>비복근경직  | 97  | 11.4     |
| Pes planus<br>Intoe             | 양발 평발<br>안짱걸음   | 24  | 2.8      |
| Pes planus<br>Overpronation     | 양발 평발<br>과회내증   | 26  | 3.1      |
| Pes planus<br>Scoliosis         | 양발 평발<br>척추측만증  | 29  | 3.4      |
| 합계                              |                 | 849 | 100.0    |

표 3. 나이별 연령구분

Table 3. Age categories.

| 나이(year) | 구분  | 데이터 수(n) |
|----------|-----|----------|
| 0~6      | 영유아 | 228      |
| 7~12     | 소아  | 313      |
| 13~18    | 청소년 | 156      |
| 19 이상    | 성인  | 152      |

표 4. 수정된 애시워스 척도

Table 4. Modified Ashworth scale(MAS).

|          | 내용  |
|----------|---|
| grade 0  | 근 긴장의 증가 없음   |
| grade 1  | 가벼운 근 긴장의 증가 있음<br>사지를 굴곡 또는 신전운동을 시키면 걸림 또는 가동 범위의 마지막에 약간의 저항이 있음 |
| grade 1+ | 가벼운 근 긴장도가 있음<br>걸림이 명백하고, 가동범위의 1/2이하의 범위에서 약간의 저항이 있음             |
| grade 2  | 명백한 근 긴장의 증가가 거의 전체 가동범위를 통해 인지 되지만 사지는 쉽게 움직일 수 있음                 |
| grade 3  | 상당한 근 긴장의 증가가 있고 수동 운동은 곤란함   |
| grade 4  | 근육이 경직되어 굴곡이나 신전을 할 수 없음  |

3에서 연령별 구분과 데이터 수를 확인 할 수 있다. 예상 정확도의 향상을 위해 전정강도(pruning severity)는 75%, 가지 당 최소 레코드는 2, 부스팅 시도 수는 10으로 설정하였다.

비복근경직(gastrocnemius tightness:GCM tightness)의 진단은 근의 긴장도를 임상적으로 정량화 하는 방법인 수정된 애시워스 척도(Modified Ashworth scale: MAS) 방법이 임상에서 주로 이용된다[15]. 이것은 대상 환자를 충분히 이완시켜 수의적 근 수축을 배제한 상태에서 검사자가 한손으로 경골 내과 바로 위를 잡고 한손은 족장 원위부를 잡은 후 최대 저축굴곡(plantarflexion)에서 최대 배축굴곡(dorsiflexion)까지 최대한 빨리 수동 관절운동을 가하면서 느껴지는 저항을 평가 하는 방법이다. 표 4는 MAS 척도에 관해 나타낸다.

3. C5.0 귀납 규칙 결과 및 분석

C5.0 모형을 도출하기 위한 전체적인 수행 절차는 그림 3과 같으며, 표 5는 연령대별(영유아, 소아, 청소년, 성인)과 전체 데이터에 대한 결정 트리 수행 결과 나타난 모델의 예상 정확도를 나타낸다. 보다 우수한 예측 결과를 위해 부스팅 옵션을 사용하였고 부스팅 시도 수

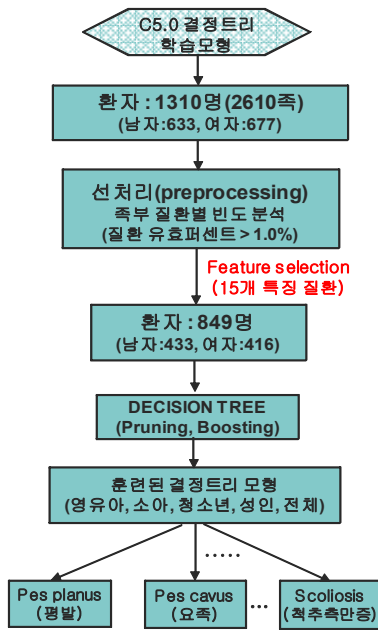


그림 3. C5.0 결정트리 학습모형 수행 절차  
Fig. 3 Procedure of C5.0 decision tree for learning.

표 5. C5.0 예상 정확도(%)  
Table 5. C5.0 estimated accuracy(%).

|          | 영유아  | 소아   | 청소년  | 성인   | 전체   |
|----------|------|------|------|------|------|
| Rule #1  | 87.7 | 76   | 83.3 | 87.5 | 85.5 |
| Rule #2  | 79.4 | 68.1 | 75.6 | 88.8 | 77.6 |
| Rule #3  | 82.5 | 71.2 | 75   | 79.6 | 76.7 |
| Rule #4  | 82   | 70.9 | 71.8 | 80.9 | 80.6 |
| Rule #5  | 85.1 | 73.2 | 73.1 | 88.2 | 76.1 |
| Rule #6  | 79.8 | 71.9 | 75   | 79.6 | 77.3 |
| Rule #7  | 85.5 |      | 74.4 | 83.6 | 79.6 |
| Rule #8  | 76.8 |      | 76.9 | 79.6 | 77.7 |
| Rule #9  | 80.3 |      | 78.8 | 78.9 | 79   |
| Rule #10 | 82.5 |      | 71.2 | 77   | 76.8 |

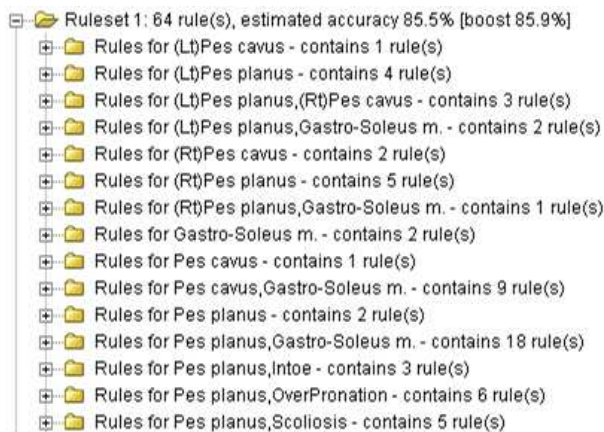


그림 4. 생성된 결정트리 결과(전체 데이터:15개 질환)  
Fig. 4. The result of the generated decision tree total data:15 foot deformity).

만큼 총 10개의 규칙(rule)에 대한 예상 정확도를 확인할 수 있다. 소아의 경우 마지막 분류자가 부정확하여 시도 수가 6까지만 증가 하였다. 가장 높은 정확도는 전체에서 85.5%, 영유아 87.7%, 소아 76%, 청소년 83.3%, 성인 89.2%로 각각 나타났다. 그림 4는 전체 데이터에서 생성된 결정 트리 결과를 나타내며 총 64개의 규칙이 생성된 것을 확인할 수 있다. 규칙집합 1(Ruleset 1)은 전체 15개의 족부 질환에 대해서 64개의 진단 규칙을 탐사하였고, 예측 정확도는 85.5% 임을 의미한다. 하위 규칙들은 각 15개의 족부 질환에 대해 생성된 진단 규칙들을 의미하며 각 규칙에 대한 세부적인 표현은 그림 5에서 보이고 있다. 그림 4가 족부 질환의 전체적인 특징을 설명한다면 그림 5의 세부 규칙은 각 질환의 진단 과정과 그 결과에 초점이 맞춰져 있기 때문에 각각의 질환에 대한 진단 특징을 쉽게 이해하고 예측하는데 매우 용이하다.

생성된 규칙들은 그림 5와 같이 분류된 질환에 대한 규칙의 세부 내용을 If-then 규칙들의 집합 형태로 표현된다. 그림 5의 규칙 3의 경우 만약 오른쪽 발목 배측 굴곡(dorsiflexion)의 MAS 정도가 grade 0이고, 왼쪽 거골하관절(subtalar joint:STJ)의 내반 각도가 31° 보다 작거나 같으며, 왼발의 RCSP가 -2° 보다 작거나 같고, 오른발의 RCSP가 -2° 보다 크고 1°보다 작거나 같으면 왼족 골반(pelvic)의 트렌델렌버그(trendelenburg)각도가 1°보다 작거나 같으면 왼발이 평발(pes planus)이다 로 진단함을 의미한다. 마찬가지로 그림 4의 규칙 1

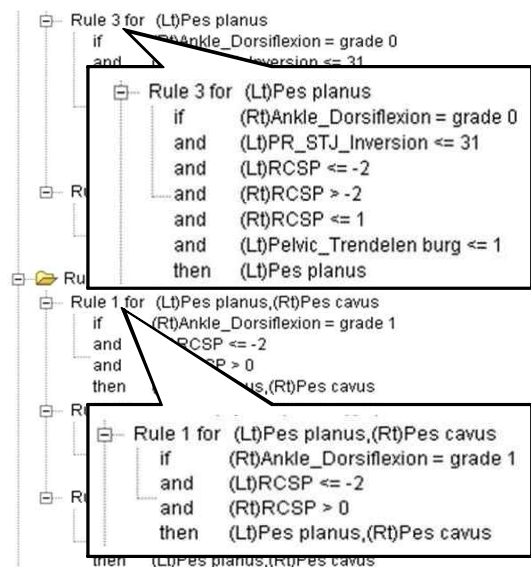


그림 5. 세부 규칙(if-then rule)  
Fig. 5. Detailed rules(if-then rule).

의 경우 오른발의 배측굴곡의 MAS정도가 grade 1이고 왼발의 RCSP가  $-2^{\circ}$  보다 작거나 같고 오른발의 RCSP가  $0^{\circ}$ 보다 크면 왼발을 평발 그리고 오른발을 요족발(pes cavus)로 진단함을 의미한다.

4. 실제 값과 예측 값의 일치율 비교

실제 진단내용과 C5.0에 의해 예측된 결과의 비교를 위해 Analysis노드를 사용하였으며, 표6은 훈련 집합(training set)과 테스트 집합(test set)으로 나뉜 연령대별(영유아, 소아, 청소년, 성인)과 전체에 대해 생성된 질환 진단 규칙 수와 일치율을 나타낸다. 전체 데이터의 훈련 집합 일치율은 78.83%, 테스트 집합 일치율은 72.79%로 나타났다. 연령대별 일치도는 훈련 집합에서 성인, 청소년, 영유아 그리고 소아 순으로 높게 나타났으며, 테스트 집합에서는 성인, 영유아, 청소년 그리고 소아 순으로 높게 나타났다. 표 6의 결과를 보면 성인의 경우 신체발달의 변화가 거의 정지된 시기이므로 질환에 대한 진단률이 99.08%로 정확하며, 영유아와 특히 소아의 경우 발달에 있어 변화와 다양성이 많은 시기임에 따라 더 낮은 진단율을 보이는 것으로 사료된다.

표 6. 생성된 규칙 수와 일치율 비교  
Table 6. A number of generated rules and comparison of agreement.

|     |     | 일치율(%) | 생성 규칙수 |
|-----|-----|--------|--------|
| 영유아 | 훈련  | 69.48  | 14     |
|     | 테스트 | 72.46  |        |
| 소아  | 훈련  | 67.91  | 18     |
|     | 테스트 | 62.92  |        |
| 청소년 | 훈련  | 77.67  | 21     |
|     | 테스트 | 70.21  |        |
| 성인  | 훈련  | 99.08  | 21     |
|     | 테스트 | 89.36  |        |
| 전체  | 훈련  | 78.83  | 64     |
|     | 테스트 | 72.79  |        |

IV. 결 론

의료 데이터 분석의 경우 대부분 통계를 이용한 유의성 검증 또는 상관계수나 상관 분석등과 같은 데이터의 연관성 분석 방법의 경우 제시된 요인들의 상관관계를 수치 값으로만 제공함으로써 정량 평가만이 가능하다. 그러나 모든 의료적 행위에 있어서 질환에 대한 진료 방법과 그에 대한 관련성 파악은 효과적이고 정확한 진단을 내리는데 대단히 중요하다. 이러한 관점에서 데이

터 마이닝은 지능적인 분석 방법론을 제시 하여 준다.

결정트리 알고리즘은 분류 또는 예측의 과정이 나무 구조에 의한 추론규칙에 의해서 표현되기 때문에 신경망이나, 판별분석, 회귀분석 등 다른 방법에 비해서 그 과정을 쉽게 이해하고 설명할 수 있다는 점에서 더욱 실질적인 결과를 제공하는 분석 방법이다. 기존에 결정 트리를 이용하여 혈청 단백질(serum protein)로부터 신장암 또는 전립선암을 감별 진단하는 연구가 수행된 바가 있지만<sup>[16~17]</sup> 이를 족부 질환에 적용하여 진단규칙을 생성하는 학습 모형으로의 응용은 본 연구에서 처음으로 시도 하였다.

본 논문은 충남대학교병원 재활의학과 신발클리닉에 다양한 족부 질환으로 내원한 환자들의 임상 데이터를 바탕으로 결정트리 학습 방법을 이용하여 족부 임상 데이터를 분류 하고 식별 하였다. 분석의 결과 전체 15개의 족부 질환에 대한 64개의 규칙을 생성하였고, 이로부터 각 질환에 대한 진단의 결정 과정을 설명하고, 각 질환 진단의 결정에 주요 인자가 무엇인지를 파악 할 수 있다. 특히 특정 질환에 대한 세부적인 설명성은 진단이 모호한 환자에 대한 변별력을 제공한다. 5개의 클래스(영유아, 소아, 청소년, 노인, 전체)로부터 생성된 각 결정트리는 각 클래스의 특징과 클래스간의 상관관계를 설명하였다. 연령대별 일치율의 변화는 연령대의 특성상 소아·청소년은 발달과정이 가장 활발한 시기이며, 정확한 병력을 알기가 힘들고 질환에 따른 증상 또한 성인과 같이 전형적으로 나타나지 않고 그 편차(deviation)또한 크게 나타나기 때문에 나타난 결과라 사료된다. 본 연구에서 탐사된 결정 트리의 결과는 미지의 족부 질환에 대한 객관적인 지표를 제공함에 따라 전문가의 진단과 치료 과정의 결정에 도움을 주고 예측이 가능하게 한다. 또한, 실제 임상 데이터로부터 탐사된 규칙이고 이 결과는 다시 피드백되어 재분석이 이루어짐에 따라 의료 현장에 바로 적용할 수 있는 실질적인 정보를 제공할 것이다.

향후 진단 규칙의 정확도를 높이기 위한 방안으로서 서로 다른 결정 트리 알고리즘의 비교 평가와 선처리 단계에서 데이터의 재가공 등을 통하여 마이닝의 효율성에 대한 검증을 보완하고자 한다. 또한 본 논문에서 조사한 15개 질환 외에 특정 질환 또는 연령대별 질환 특성에 대해 다양한 분석 방법들을 연구하고 진단 항목과 질환 사이의 상관관계를 분석하기 위한 효율적인 알고리즘을 개발 하고자 한다.



## 참 고 문 헌

- [1] Kevin Deschamps, Ivan Birch, Kaat Desloovere, Giovanni A. Matricali, "The impact of hallux valgus on foot kinematics : A cross-sectional comparative study", *Gait & Posture*, Vol. 32, no. 1, pp. 102-106, 2010.
- [2] Yu-Chi Chen, Shu-Zon Lou, Chen-Yu Huang, Fong-Chin Su, "Effects of foot orthoses on gait patterns of flat feet patients", *Clinical Biomechanics*, Vol. 25, no. 3, pp. 502-505, 2010.
- [3] Joshua Burns, Jack Crosbie, "Weight bearing ankle dorsiflexion range of motion in idiopathic pes cavus compared to normal and pes planus feet", *The Foot*, Vol. 15, no. 2, pp. 91-94, 2005.
- [4] 김천식, 홍유식, 정명희, "데이터 마이닝을 이용한 차량 사고자 사망확률 모형", *대한전자공학회논문지*, 제43권, 제9호, 15-31쪽, 2006년.
- [5] Blum RL. "Displaying clinical data from a time-oriented database", *Computers in Biology and Medicine*, Vol. 11, no. 4, pp. 197-210, 1981.
- [6] Tiet Q, Ilgen MA, Byrnes HF, Moos RH. "Suicide attempts among substance use disorder patients: an initial step toward a decision tree for suicide management" *Alcoholism: Clinical and Experimental Research*, Vol. 30, no. 6, pp. 998-1005, 2006.
- [7] 이아름, 박용근, 권대규, 김정자, "가중연관규칙 탐사를 이용한 재활훈련운동과 근육 활성의 연관성 분석", *전자공학회논문지*, 제46권, CI편 제6호, 7-17쪽, 2009년.
- [8] Jiawei Han, Micheline Kamber 공저, 강창완등 옮김, "데이터 마이닝 : 개념과 기법", 사이플러스, 271-290쪽, 2007년.
- [9] 최중후, "데이터마이닝 의사결정나무분석", 고려정보산업, 17-31쪽, 1998년.
- [10] 이중현, 이영신, 이재욱, 박성하. "Orthotics 착용에 따른 정상, 요족, 평발의 생체역학적 보행분석 및 시뮬레이션", *대한기계학회논문집*, 제31권, 제11호, 1115-1123쪽, 2007년.
- [11] 송동호, 이운, 은백린, 이광재, 강상국, 박성진, 유성, 신정빈, 김봉욱. "경골 내염전 환아에서 경비골 역회전 교정장치(TCR)의 치료 효과", *대한소아과학회*, 제50권, 제1호, 79-84쪽, 2007년.
- [12] Hutter CG, Scott W. "Tibial torsion", *J Bone Joint Surg Br*, vol. 21, no. A, pp. 511-518, 1949.
- [13] McDonough MW. "Angular and axial deformities of legs of children", *Clin Podiatry*, vol. 1, pp. 601-620, 1984.
- [14] Quinlan R., "Induction of decision tree", *Machine Learning*, Vol. 1, no. 1, pp. 81-106, 1986.
- [15] Bohannon RW, Smith MB. "Interrater reliability of a modified Ashworth scale of muscle spasticity" *Phys Ther*, Vol. 67, pp. 206-207, 1978.
- [16] Bao-Ling Adam, Yinsheng Qu, John W. Davis, et al., "Serum protein fingerprinting coupled with pattern-matching algorithm distinguishes prostate cancer from benign prostate hyperplasia and healthy men", *Cancer Res*, Vol. 62, pp. 3609-3614, 2002.
- [17] Yonggwon Won, Ho-Jun Song, Taek Won Kang, Jung-Ja Kim, Byoung-Don Han, Seung-won Lee, "Pattern analysis of serum proteome distinguishes renal cell carcinoma from other urologic diseases and healthy persons", *Proteomics*, Vol. 3, no. 12, pp. 2310-2316, 2003.

저 자 소 개



김진호(정회원)  
 2008년 전북대학교 생체정보  
 공학과 학사 졸업  
 2011년 전북대학교 헬스케어  
 공학과 석사 졸업  
 2011년~현재 전북대학교 헬스케어  
 공학과 박사 과정

<주관심분야 : 바이오인포매틱스, 데이터마이닝,  
 족부생체역학>



박인식(정회원)  
 1994년 호주 족부의사 자격증 및  
 면허증 취득  
 1997년 영국 웨스트민스터대학 족  
 부의학과 졸업  
 2003년 미국 공인 페도티스트  
 2003년~2009년 경기대학교 대학  
 원 겸임교수

2010년 대한 페도틱협회 회장  
 2010년~현재 바이오메카닉스(주) 대표이사  
 <주관심분야 : 족부생체역학, 의료용 신발, 발의  
 교정, 보행, 보조기>



김봉옥(정회원)  
 1978년 연세대학교 의학사  
 1985년 연세대학교 대학원  
 의학박사  
 1988년~현재 충남대학교 의과  
 대학 재활의학교실 교수  
 2010년~현재 충남대학교병원  
 재활의학과장

2010년~현재 대전 장애인보조기구 사례관리  
 센터장  
 2010년~현재 한국 의지보조기 학회장  
 <주관심분야 : 의료용 신발, 발의교정, 보행, 보조  
 기, 재활공학, 소아재활>



양윤석(정회원)  
 1996년 서울대학교 제어계측  
 공학과 학사 졸업.  
 1998년 연세대학교 협동과정 생체  
 공학과 석사 졸업.  
 2002년 서울대학교 공과대학 협동  
 과정 의용생체공학전공  
 박사 졸업

2002년~2005년 한국전자통신연구원(ETRI).  
 2005년~현재 전북대학교 바이오메디컬공학부  
 교수  
 <주관심분야 : 바이오센서, 임베디드 시스템, 신  
 경재활공학, 에너지 수확>



원용관(정회원)  
 1987년 한양대학교 전자공학과  
 학사 졸업.  
 1991년 미국 미주리주립대학교  
 컴퓨터공학 석사 졸업.  
 1995년 미국 미주리주립대학교  
 컴퓨터공학 박사 졸업.

1995년~1996년 한국전자통신연구원(ETRI).  
 1997년~1999년 한국통신(KT).  
 1999년~현재 전남대학교 전자컴퓨터공학부  
 교수.  
 <주관심분야 : 패턴인식, 신호 및 영상처리, 컴퓨  
 터 의료진단 및 기기>



김정자(정회원)-교신저자  
 1985년 전남대학교 계산통계학과  
 학사 졸업  
 1988년 전남대학교 전산학과 석사  
 2002년 전남대학교 전산학과 박사  
 1988년~2002년 전남대학교  
 전산학과 외래교수

2002년~2004년 전남대학교 전자통신 연구소  
 post-doc  
 2004년~2006년 한국 Bio-IT 파운드리 사업단  
 광주센터 연구교수  
 2006년~현재 전북대학교 바이오메디컬공학부  
 교수  
 <주관심분야 : 바이오인포매틱스, 데이터 마이닝,  
 생체역학, 족부 의공학>