

순차규칙 모델링을 활용한 뇌경색증 환자 진단 분석

Analysis for Diagnosis of Patients with Cerebral Infarction by Sequence Modeling

신아미*, 박희준, 이인희, 김윤년

A. M. Shin, H. J. Park, I. H. Lee and Y. N. Kim

요약

본 연구는 데이터 마이닝 기법 중 하나인 순차규칙 모델링을 활용해 뇌경색증 환자의 진단을 분석함으로써 뇌경색증의 선행질환과 합병증을 파악하고 이를 최소화할 수 있는 중재 개발에 기초자료를 제공하고자 시도되었다. 뇌경색증 환자의 진단 관련 데이터는 대구에 위치한 A 의료원에서 2000년에서 2007년 사이 뇌경색증으로 진단받은 환자의 관련 데이터를 추출하여 일괄분석을 위해 데이터마트를 구축하였다. 대상 의료기관에서 총 2,267명이 뇌경색증을 진단 받았고 추출된 진단 데이터는 총 32,692건이었다. 데이터 분석은 Clementine 12.0의 순차규칙 모델링을 이용하였고 8개의 의미 있는 규칙이 발견되었다. 본 연구 결과는 2차 뇌경색 발생 예방 프로그램 개발 및 뇌경색증 환자의 합병증 예방 중재 개발의 기초자료로 활용할 수 있을 것으로 기대된다.

ABSTRACT

This study was tried to analyze the diagnosis of patients with cerebral infarction by sequence modeling that was one of data mining analysis method and find out previous disease or complication of patients with cerebral infarction. Mass data that the diagnosis code of cerebral infarction was I63 in 2000 to 2007 were extracted from A hospital's database and then the data mart was constructed for analysis. Total 2,267 patients illnesses were diagnosed as cerebral infarction and 32,692 cases related diagnosis were extracted. Sequence modeling in Clementine 12.0 program was used to analyze diagnosis of patients with cerebral infarction and 8 meaningful rules were found in this paper. This result could be used as a basic data to make secondary cerebral infarction prevention program and to prevent complication of cerebral infarction.

Keyword : Cerebral infarction, Diagnosis analysis, Sequence modeling, Data mining

1. 서론

뇌혈관 질환은 암 다음으로 사망률이 높은 질환으로 우리나라 3대 사망원인 중 하나이다. 50대 이

접수일 : 2008. 1. 30

심사완료일 : 2007. 2. 6

* 신아미 : 계명대학교 생체정보기술개발사업단

ami535@kmu.ac.kr (주저자)

박희준 : 계명대학교 동산의료원

hjpark@dsmc.or.kr (교신저자)

이인희 : 계명대학교 동산의료원

yiinhee@hanmail.net (공동저자)

김윤년 : 계명대학교 동산의료원

ynkim@dsmc.or.kr (공동저자)

* 본 연구는 지식경제부 지방기술혁신사업(RTI04-01-01) 지원으로 수행되었음.

상에서 사망원인 2위를 차지하고 있으며 사망률은 인구 10만 명당 50대 36.8명, 60대 129.3명, 70대 482.5명, 80세 이상 1405.5명으로 나이가 들수록 사망률이 급격히 늘어나고 있다[6]. 뇌졸중은 뇌의 한 부분에서 갑작스런 혈류의 방해가 일어나 뇌세포가 괴사되고 해당 부위 뇌가 통제하는 신체기능이 파괴되거나 손상되는 것으로 허혈성 뇌졸중과 출혈성 뇌졸중으로 분류할 수 있다. 허혈성 뇌졸중인 뇌경색증은 혈전이나 색전이 뇌혈관을 막을 때 발생하고, 출혈성 뇌졸중은 뇌혈관이 갑자기 터져 발생하는 것으로 혈액이 뇌세포 속으로 스며들게 된다[11]. 대부분의 뇌졸중 자료에서 허혈성 뇌졸중이 출혈성에 비해 3~8배 정도 흔하다고 알려져 있고 뇌졸중은 50% 이상에서 신체적, 기능적 장애를 초래하며, 첫 발병 후 1달 이내 사망률이 22%이고, 10

년 이내 사망하거나 지속적인 치료가 요구되는 경우는 79%에 이른다[10, 12, 13]. 통계청 자료에서 보고된 바와 같이 뇌졸중은 일반적으로 고령층에서 발생하지만, 최근 연구에서는 45세 이하의 청년에서 초기 장년층 연령에서 전체 뇌경색 환자에서 비율이 5~10% 정도 차지하고 있는 것으로 보고하고 있다. 이 연령층에서 발생 빈도는 고령층에 비해 낮지만 유병기간과 재활 기간이 길며 가정 및 사회에서의 역할을 고려할 때 가정, 사회, 경제적인 면에서 문제가 심각하다고 할 수 있다[4, 8].

데이터 마이닝이란 패턴인식 기술이나 통계기법, 수학적 알고리즘을 이용하여 의미 있는 새로운 상관관계, 패턴, 추세 등을 발견하는 과정으로 방대한 양의 데이터로부터 합리적이며 잠재적 유용성이 있는 정보를 발견할 수 있다[9]. 데이터 마이닝으로 가능한 작업은 예측, 세분화, 연관성/순차규칙, 인파 모형, Text 분석 등이 있고 이러한 작업들은 회귀분석, 로지스틱 회귀분석, 요인분석, 군집분석, 판별분석, 신경망분석, 의사결정나무 분석, 연관성 분석, 순차 분석, 시계열 분석, 빈도분석 및 교차분석 등의 분석기법을 이용해 이루어진다. 특히 순차 분석은 시간을 두고 순차적으로 발생하는 사건에 대한 순차규칙을 만들어내는 것으로 시간을 두고 순차적으로 발생하는 제품 구매를 분석하거나 시간에 따른 질병의 발생 패턴 등을 분석할 수 있다.

뇌졸중의 발생 빈도가 증가함에 따라 뇌졸중의 발생 시기에 따른 연령 분포, 유형, 장애 형태, 위험 인자, 합병증 등의 역학적 요인에 대한 조사 연구가 이루어지고 있지만 데이터 마이닝과 같은 분석 방법을 이용하여 대용량의 자료를 대상으로 시간의 경과에 따라 진단의 패턴을 분석한 연구는 없었다. 따라서 본 연구에서는 뇌졸중 중 높은 빈도를 차지하는 뇌경색증을 진단 받은 환자에게 시간이 경과함에 따라 추가되는 진단 패턴을 데이터 마이닝의 순차규칙 모델링을 활용해 분석함으로써 뇌경색증의 선행질환과 합병증을 파악하고 이를 최소화할 수 있는 중재 개발에 기초자료를 제공하고자 시도되었다.

2. 본론

2.1 재료 및 방법

뇌경색증 대상자의 진단 분석을 위해 2000년에서 2007년 사이 대구시 A의료원에서 뇌경색증 진단을 받은 환자의 데이터를 이용하였다. 데이터베이스에

서 대뇌 질환 중에서 국제질병사인(International Classification of Disease, ICD) 분류코드 중 주 진단이 뇌경색증(cerebral infarction)이고, Korea classification of disease(KCD)-4 코드 I63인 60세 이하 환자의 진단 정보를 추출 하였다. 뇌경색증 대상자의 진단 패턴을 분석하기 위해 뇌경색증 진단 시점을 중심으로 진단 전·후를 구별하지 않고 뇌경색증 진단을 받은 환자의 모든 진단 정보를 추출 하였다. 진단을 제외한 흡연력, 음주력, 고혈압 유무 등 대상자의 일반적 특성은 의무기록 검토를 통해 자료를 수집하였다.

대상자의 일반적 특성은 SPSS WIN 12.0 프로그램을 사용해 빈도와 백분율을 구하였고, 시간이 경과함에 따라 뇌경색증 환자에게 내려진 진단 패턴의 규칙을 발견하기 위해 데이터 마이닝 프로그램 중 하나인 Clementine 12.0 프로그램의 순차규칙 모델링을 이용하여 분석을 실시하였다. 순차규칙 모델링을 이용한 뇌경색증 환자 진단 분석의 구체적인 과정은 그림 1과 같다.

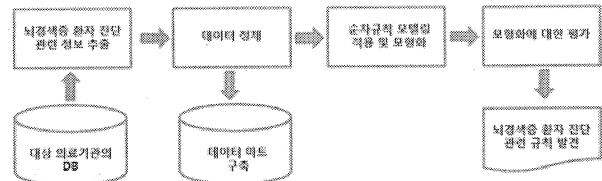


그림 1. 순차규칙 모델링을 활용한 뇌경색증 진단 분석 과정

2.2 뇌졸중 환자 진단 데이터 마트 구축

Clementine 12.0 프로그램을 대상의료원의 데이터 베이스에 연결하여 2000년에서 2007년 사이 대상의료원에서 뇌경색증(I63) 진단을 받은 대상자의 자료를 추출한 결과 총 2,267명이 뇌경색증을 진단 받았고, D/W에서 추출된 진단 데이터는 총 32,692건이었다. 이 중 등록번호와 진단 날짜, 진단명이 같은 중복되는 데이터는 제외하고 전체 진단의 0.5% 이상을 차지하는 다빈도 진단 12가지(E11, E78, G45, G46, G81, I10, I20, I63, I69, K25, K29, N18)를 선정하여 총 10,488건에 대한 데이터 마트를 구축하였다.

2.3 대상자의 일반적 특성

2000년 1월에서 2007년 10월까지 뇌경색증(I63) 진단을 받은 대상자는 총 2,267명이었고 이중 남자

가 1,434명(63.3%), 여자가 833명(36.7%)이었으며, 평균 나이는 50.67 ± 7.38 세였다. 흡연력이 있는 대상자는 784명(34.6%), 음주력이 있는 대상자는 732(32.3%), 고혈압 진단을 받은 대상자는 1,009명(44.5%)이 있었다.

2.4 뇌경색증 환자 진단 분석

2,267명의 뇌경색증 환자에게 내려진 진단코드는 A02에서 Z96까지 매우 다양하였고, 빈도 또한 1건에서 최고 5,366건까지 다양하게 나타났다. 뇌경색증 환자에게 내려진 다빈도 진단코드는 그림 2와 같이 E11(인슐린-비의존성 당뇨병), E78(지단백질 대사 장애 및 기타 지혈증), G45(일과성 대뇌 허혈 성 발작 및 관련 증후군), G46(대뇌혈관 질환에서의 뇌혈관 증후군), G81(편마비), I10(본태성 고혈압), I20(협심증), I63(뇌경색증), I69(뇌혈관 질환의 후유증), K25(위궤양), K29(위염 및 십이지장염), N18(만성 콩팥기능상실)이 있었고, 모두 전체 진단건수의 0.5% 이상을 차지하였다. 그 중 I63은 재발을 고려하여 세분화된 진단 중 중복되는 항목을 제거하였지만 I63의 세분화된 코드로 재진단 받는 경우가 있어 진단 건수가 5,366건으로 가장 많이 진단되었고 그 다음으로 I10, E11, E78 순으로 높은 빈도를 나타내었다.

표 1. 대상자의 일반적 특성

항목	빈도 (%)
성별	남 1,434 (63.3) 여 833 (36.7)
나이	평균 \pm 표준편차 50.67 \pm 7.38
흡연력	유 784 (34.6) 무 791 (34.9) 과거 흡연자 44 (1.9) 기록 없음 648 (28.6)
음주력	유 732 (32.3) 무 823 (36.3) 기록 없음 712 (31.4)
가족력	유 323 (14.2) 무 920 (40.6) 기록 없음 1,024 (45.2)
고혈압	유 1,009 (44.5) 무 857 (37.8) 기록 없음 401 (17.7)

(N=2,267)

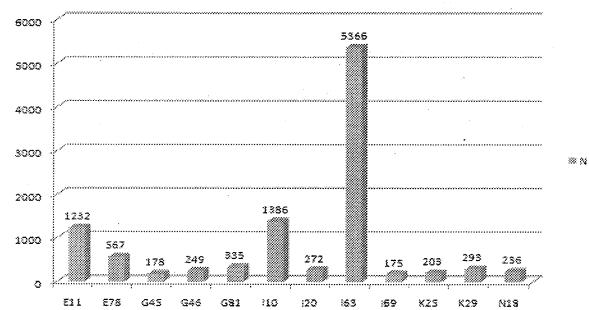


그림 2. 뇌경색증 환자의 진단 분석 결과

I63을 세분화하여 분석한 결과는 그림 3과 같이 I63, I63.0(뇌전동맥의 혈전증에 의한 뇌경색증), I63.1(뇌전동맥의 색전증에 의한 뇌경색증), I63.2(뇌전동맥의 상세불명 폐색 또는 협착에 의한 뇌경색증), I63.3(대뇌동맥의 혈전증에 의한 뇌경색증), I63.4(대뇌동맥의 색전증에 의한 뇌경색증), I63.5(대뇌동맥의 상세불명 폐색 또는 협착에 의한 뇌경색증), I63.6(대뇌정맥 혈전증에 의한 비화농성 뇌경색증), I63.8(상세불명의 뇌경색증)로 다양하게 나타났다. 그 중에서 I63이 3,169건으로 가장 많았고, 그 다음으로 I63.9가 781건 진단되었다.

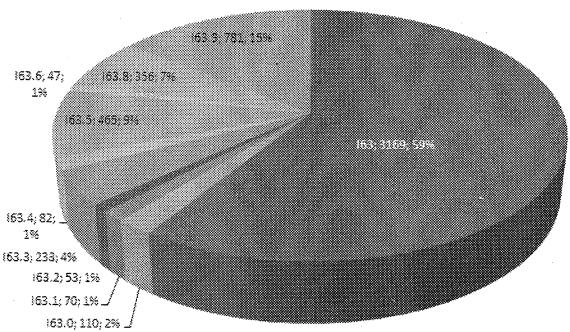


그림 3 진단 코드 I63 분석 결과

2.5 순차규칙 모델링을 활용한 뇌경색증 환자 진단 분석

그림 2에 나타난 뇌경색증 환자에게 내려진 다빈도 진단 12가지를 대상으로 순차규칙 모델링을 활용해 분석하였다. 분석 시 뇌경색증은 재발의 경우를 고려하여 I63을 세분화된 진단명을 이용하였다. 그 결과 표 2와 같이 총 8개의 의미 있는 순차규칙

을 발견할 수 있었다. 그 중에서 I63을 진단 받은 후 I10을 진단받은 경우가 705건으로 가장 많았고, 지지도와 신뢰도 또한 각각 96%, 32%로 전체 규칙 중 가장 높았다. I10에서 I63.9, I10에서 I63을 제외한 모든 규칙은 90% 이상의 높은 지지도를 나타냈지만 신뢰도는 I63에서 I10이 32%, I63에서 I63.9가 25%이고 다른 규칙들은 20% 미만이었다.

표 2. 순차규칙 모델링을 활용한 뇌경색증 환자 진단 분석 결과

전항	후항	사례빈도	지지도 %	신뢰도 %
I63	G81	248	96.03	11.392
I63	I63.5	266	96.03	12.219
I63	E78	343	96.03	15.756
I63	E11	381	96.03	17.501
I10	I63.9	234	45.96	22.457
I63	I63.9	547	96.03	25.126
I63	I10	705	96.03	32.384
I10	I63	343	45.96	32.917

2.6 고찰

국내외 여러 연구들을 통해 뇌졸중의 위험요인은 나이, 성별, 고혈압, 심장질환, 당뇨병, 유전인자, 혈중지질, 일과성 허혈성 발작, 흡연, 식이, 비만, 알코올 섭취, 인종, 만성감염 등이 있는 것으로 나타났고[1], 신동혁 등[4]은 청장년층 허혈성 뇌졸중의 위험요인으로 흡연, 음주, 고혈압, 심장질환, 고지혈증, 당뇨병이 있다고 보고하였다. 본 연구의 일반적 특성에서도 전체 대상자의 60% 이상이 남자였고, 흡연력이 있는 경우 34%, 음주력이 있는 경우 32%, 고혈압이 있는 경우가 44%인 것으로 나타나 선행연구에서 나타난 위험요인들이 본 연구에서도 동일하게 적용됨을 확인할 수 있었다. 하지만 본 연구는 대상의료기관이 전자의무기록을 사용하기 전에 뇌경색을 진단받은 환자의 데이터가 포함되어 기록을 찾을 수 없는 경우가 많아 이를 감안하면 흡연력, 음주력, 고혈압이 있는 대상자의 수는 더 많을 것이라 사료된다.

나이는 뇌졸중의 위험 요인 중 하나로 나이가 증가함에 따라 발생 빈도 또한 증가한다. 우리나라로 고령화 사회로 진입함에 따라 1970년대에는 50대에서 뇌졸중이 가장 많이 발생했지만 1980년대, 1990

년대로 들어서면서 60-70대의 뇌졸중 발병이 유의하게 증가하는 양상을 나타내고 있다[7]. 본 연구에서는 뇌졸중의 발병이 급격하게 증가하는 60-70대의 경우는 뇌경색증과 관련된 질환 이외에 노화와 관련된 여러 질환을 가지고 있을 가능성을 고려하여 60세 이하의 대상자의 진단 데이터를 대상으로 분석을 실시하였다. 연구결과 대상자의 평균 연령이 50.67세로 45세 이하의 청장년층보다는 50대에서 뇌경색의 발생 빈도가 높음을 짐작할 수 있다.

고혈압은 뇌졸중의 위험인자로 알려져 있고 실제 뇌졸중 환자를 대상으로 한 연구에서도 고혈압이 뇌졸중의 가장 큰 위험요인으로 나타났다[3]. 그러나 본 연구에서는 본태성 고혈압을 진단받은 후 뇌경색증을 진단받은 경우 보다는 뇌경색증을 진단받은 후 본태성 고혈압을 진단 받은 경우가 사례 빈도와 신뢰도가 더 높은 것으로 나타났다. 이는 고혈압이 있었지만 특별한 증상이 없어 자신이 고혈압이 있다는 사실을 모르고 지내던 대상자가 뇌경색증의 증상이 나타나 병원을 내원하여 치료를 받는 과정에서 고혈압이 있음을 알게 되고 진단을 받았기 때문이라 사료된다. 당뇨병과 고지혈증 또한 뇌졸중의 위험요인지만 본 연구에서는 뇌경색을 진단 받은 후 당뇨병과 지단백질 대사 장애 및 기타 지혈증을 진단받은 것으로 나타나 특별한 증상이 나타나지 않는 질환에 대한 조기발견에 대한 교육 및 정기검진이 필요하다고 사료된다. 또한 유경호 등[5]이 허혈성 뇌졸중 환자의 고혈압과 당뇨병의 유병기간과 치료의 순응도를 조사한 결과 89.9%의 환자가 기준에 고혈압 진단을 받았지만 이를 중 정기적으로 치료를 받은 경우는 45.5%에 불과하고 당뇨병 역시 77.7%의 환자가 기준에 진단을 받았지만 32.5%만이 혈당을 정기적으로 조절한 것으로 나타나 고혈압이나 당뇨병의 경우 지속적인 치료와 합병증 예방을 위한 교육이 중요하다고 할 수 있다.

재원기간 중 뇌졸중 환자에게 나타난 합병증으로 김경태, 안재우, 김법영 등[2]은 폐렴, 상부 위장관 합병증, 우울증, 간염, 요로감염을 보고하였고, 김수일과 이미영[3]은 신경인성 장, 신경인성 방광, 우울증, 골다공증을 보고하였다. 본 연구에서 뇌경색증 진단을 받은 환자는 편마비(G81), 뇌경색증(I63.5, I63.9), 지단백질 대사 장애 및 기타 지혈증(E78), 인슐린 비의존성 당뇨병(E11), 본태성 고혈압(I10)과 같은 진단을 주로 받는 것으로 나타나 선행연구와는 다른 결과를 보였다. 이는 선행연구의 경우 급성기 입원한 시점을 중심으로 합병증을 조사한데 반해 본 연구는 대상자가 뇌경색증 진단을 받은 시점을 중심으로 뇌경색증이 발병하기 전·후 외래진료

혹은 입원과 관련된 진단이 모두 포함되어 있기 때문에 문인 것으로 사료된다.

3. 결론

본 연구는 데이터 마이닝 기법 중 하나인 순차규칙 모델링을 활용하여 뇌경색증 환자의 진단을 분석하고 뇌경색증에 선행하는 진단과 합병증으로 동반되는 진단의 패턴을 발견하고자 시도되었다. 본 연구 결과를 바탕으로 본태성 고혈압(I10)을 진단 받은 환자는 뇌경색증(I63)을 진단받고, 뇌경색증을 진단받은 환자는 편마비(G81), 뇌경색증(I63.5, I63.9), 지단백질 대사 장애 및 기타 지혈증(E78), 인슐린 비의존성 당뇨병(E11), 본태성 고혈압(I10)과 같은 합병증을 진단받게 될 가능성이 높다고 할 수 있다. 따라서 고혈압을 예방할 수 있는 중재뿐 아니라 고혈압의 효과적인 관리를 위한 교육을 통한 뇌경색증 예방이 필요하다. 또한 뇌경색증이 발병한 환자는 재발을 예방하고 편마비, 지단백질 대사 장애 및 기타 지혈증, 인슐린 비의존성 당뇨병과 같은 합병증을 예방할 수 있는 중재가 요구된다.

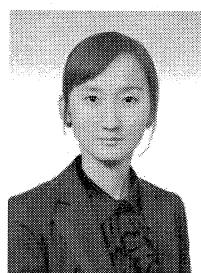
본 연구는 단일 의료기관에서 뇌경색증을 진단받은 60세 이하의 대상자만을 대상으로 분석을 실시하였고 I63의 진단코드 중 세분화되어 있지 않은 데이터가 많이 포함되어 있는 한계점이 있다. 따라서 향후 연구에서는 좀 더 많은 의료기관의 60세 이상의 노인 혹은 전체 뇌경색증 환자의 진단 자료를 대상으로 분석하고 좀 더 보편화할 수 있는 규칙을 도출할 것을 제언한다.

참 고 문 헌

- [1] 국정란, “응급의료센터에 내원한 허혈성 뇌졸중과 출혈성 뇌졸중 환자의 특성 비교”, 울산대학교 간호학과 석사학위논문, 2005
- [2] 김경태, 안재우, 김법영 등, “뇌졸중의 최근 역학적 동향”, 대한재활의학회지 Vol. 27 pp.178-185, 2003
- [3] 김수일, 이미영, “노인 뇌졸중 입원환자의 역학적 연구”, 한국전문물리치료학회지 Vol.12 No.2 pp.98-104, 2005
- [4] 신동혁, 최필조, 송형곤, 정연곤, “응급실로 내원한 청장년층 허혈성 뇌졸중 환자의 임상적 특성에 관한 연구”, 『대한응급의학회지』 Vol.16 No.1 pp.114-121, 2005

- [5] 유경호, 배희준, 권순역 등, “한국 뇌졸중 자료은행에 등록된 10,811명의 급성기 허혈성 뇌졸중 환자 분석: 병원기반 다의료기관 전향적 자료등록 연구”, 대한신경과학회지 Vol.24 pp.535-543, 2006
- [6] 통계청, 2007년 사망 및 사망원인통계결과. 출처 <http://www.nso.go.kr>, 2008.
- [7] 황석재, “우리나라 뇌졸중의 최근 20년간 변화양상에 대한 연구”, 경희대학교 의학과 석사학위논문, 2000
- [8] Bogousslavsky J, Van Melle G, Regli F, "The Lausanne Stroke Registry: analysis of 1,000 consecutive patients with first stroke". Stroke Vol.19 pp.1083-1092, 1998.
- [9] Fayyad U, Piatetsky-Shapiro G, & Smyth P, "Advances in Knowledge Discovery and Data Mining". Menlo Park, Calif, AAAI Press, 1996
- [10] Hardie, K., Hankey, G. J., Jamrozik, K., Broadhurst, R. J., & Anderson, C, "Ten-year risk of first recurrent stroke and disability after first-ever stroke in the perth community stroke study", Stroke Vol.35 pp.731-735, 2004
- [11] Mower, D, "Brain attack : Treating acute ischemic CVA", Nursing Vol.27 pp.34-39, 1997
- [12] Kristensen B, Malm J, Carlberg B, Stegmayr B, Backman C, Fagerlund M, et al., "Epidemiology and etiology of ischemic stroke in young adults aged 18 to 44 years in northern Sweden", Stroke Vol.28 pp.1702-1709, 1997
- [13] Siqueira JI, Santos AC, Fabio SR, Sakamoto AC, "Cerebral infarction in patients aged 15 to 40 years", Stroke Vol.27 pp.2016-2019, 1996

신 아 미

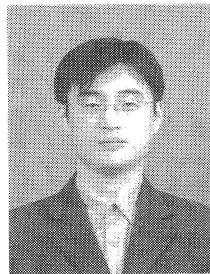


2002년 2월 계명대학교 간호학과 졸업 (학사)
 2007년 8월 계명대학교 대학원 간호학과 졸업 (석사)
 2002년 4월 - 2005년 8월 계명대학교 동산의료원 간호사
 2005년 9월 - 2007년 7월 계명대학교 간호대학 실습조교
 2007년 8월 - 현재 계명대학교 생체정보기술개발

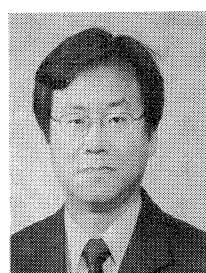
사업단 연구원
관심분야 : 재활복지공학, 의료정보공학



박 희 준
1999년 2월 경북대학교 전자공
학과 졸업 (학사)
2001년 2월 금오공과대학교 전
자공학과 졸업 (석사)
2006년 2월 경북대학교 전자공
학과 졸업 (박사)
2005년 3월 - 현재 계명대학교 동산의료원 의료정
보학교실 연구강사
관심분야 : 의용계측, 생체신호처리, u-healthcare



이 인 희
2000년 2월 대구대학교 재활과
학대학 물리치료학과 졸
업 (학사)
2005년 2월 계명대학교 대학원
의학과 졸업 (석사)
2000년 4월 - 현재 계명대학교
동산의료원 물리치료사
2006년 3월 - 현재 계명대학교 생체정보기술개발사
업단 연구원
관심분야 : 재활복지공학, 신호처리, 정보공학



김 윤 난
1980년 2월 경북대학교 의학과
졸업 (학사)
1983년 2월 경북대학교 의학과
졸업 (석사)
1987년 2월 경북대학교 의학과
졸업 (박사)
1984년 3월 - 현재 계명대학교
의과대학 내과학교실 교수
2004년 10월 - 현재 계명대학교 생체정보기술개발
사업단장
관심분야 : 심장질환, 의료정보학