

# 당뇨병 예측을 위한 신경망 모델 개발에 관한연구

## Development of Diabetes Mellitus prediction model using artificial neural network

서혜숙, 최진욱, 김희식\*

서울대학교 의과대학 의공학교실, 서울시립대학교 공과대학 제어계측공학과\*

Hyesook Suh, Jinwook Choi, Hiesik Kim\*

Department of Biomedical Engineering, College of Medicine, Seoul National  
University, Department of Control & Instrumentation Engineering Seoul City  
University\*

### ABSTRACT

There were many cases to apply artificial intelligence to medicine. In this paper, we present the prediction model of the development of the NIDDM(noninsulin-dependent diabetes mellitus). It is not difficult that doctor diagnose patient as DM(diabetes mellitus). However NIDDM is usually developed later on 40 years old and symptom appears gradually. So screening test or prediction model is needed absolutely. Our model predicts development of NIDDM with still normal data 2 year ago. Prediction models developed are both MLP(multilayer perceptron) with backpropagation training and RBFN(radial basis function network). Performance of both models were evaluated with likelihood ratio. MLP was about two and RBFN was about three. We expect that models developed can prevent development of DM and utilize normal data.

## I. 서론

당뇨병은 가장 흔한 내분비질환으로 대사장애와 합병증(눈, 신장, 신경, 혈관)이 특징적이다. [1] 1993년 우리나라의 연천 지역 주민을 대상으로 한 조사결과에 의하면 30세 이상의 일반 주민에서 7.2 %의 유병률을 보이고 있다. [2] 진단은 어렵지 않다. 금식시 정맥내 포도당 농도가 7.8 mmol/L (140 mg/dL) 이상이거나, 포도당을 75 g 섭취하고 2시간 후에 측정 한 정맥내 포도당 농도가 11.1 mmol/L (200 mg/dL) 이상이면 당뇨병으로 진단된다. 그러나 비인슐린의존성당뇨병(NIDDM)은 보통 40세 이후에 발병하고, 증상이 점진적으로 나타나기 때문에 정기 검진시 증가된 혈장내 포도당 농도를 보이는 무증상 환자가 자주 당뇨병 진단을 받게 된다. [1] 이를 위해 본 연구에서는 비선형적 모델의 패턴인식에 사용되는 인공신경망 기법을 도입하여 건강검진등의 임상검사결과로부터 향후 2년 후에 당뇨병이 발병할 가능성에 대해 예측해주는 신경망 모델을 개발하였다.

## II. 본론

### 1. 데이터 획득

#### 1) 데이터의 구성

1993년과 1995년 연천 보건소 검진 데이터 1193명 중 완전한 데이터 120명(환자 60명, 정상 60명)을 선택하여 17개 공통변수를 사용하였다.

#### 2) 데이터의 변형

각 변수의 범위를 0과 1사이로 transformation하는 선형사상을 찾고, 그 양쪽 경계를 넘는 값은 각 경계값이 되도록 프로그래밍 하였다.

#### 3) 학습데이터(train data) & 시험데이터(test data)

가. 환자와 정상 60명씩을 짝수와 홀수로 교대로 섞어 120명의 혼합데이터(total mixed data)를 만들었다.

나. 위의 혼합데이터를 편의로 10쌍 만들어 학습데이터와 시험데이터로의 사용이 다양하게 변화도록 하였다.

다. 학습데이터와 시험데이터의 비율을 1:1, 2:1, 3:1, 4:1, 5:1로 변형시켜 실험하였다.

변수명	의미
AGE	나이
SEX	성별
GLU2	공복시혈당
GLU3	식후2시간후 혈당
HEIGHT	키
WEIGHT	몸무게
AC	복위
HC	둔위
SBP	수축기혈압
DBP	확장기혈압
GOT	효소의 일종
GPT	효소의 일종
CHOLE	콜레스테롤
TG	triglyceride
HDL	고밀도지단백질
BMI	wt/ht <sup>2</sup> (kg/m <sup>2</sup> )
WHR	waist-hip ratio

### 2. 예측 모델

MATLAB 5.1을 사용하여 인공신경망을 디자인하고 프로그래밍하였다.

#### A. MLP(multilayer perceptron)

##### 1) 입력층(input layer)

변수 17개 모두를 입력층노드로 사용하여 앞

에서 0에서 1로 mapping된 값을 입력값으로 취하였다.

2) 은닉층(hidden layer) :

single hidden layer MLP : 입력층노드수를 17개로 고정하고 은닉층노드수를 다양하게 변화시켜 보았다.

전이함수(transfer function)는 tansigmoid를 사용하였다.

3) 출력층(output layer)

출력층의 노드수는 한 개로 고정하고, 당뇨병의 가능성을 예측하는 값을 출력하도록 하였다.

전이함수는 선형(purelinear)을 사용하였다. 이 때 그 값을 3개 영역(고위험군, 중간군, 저위험군)으로 나누어 보도록 프로그래밍 하였다.

기준 :            고위험군 >= 0.7  
                   -0.7 <= 중간군 < 0.7  
                   저위험군 < -0.7

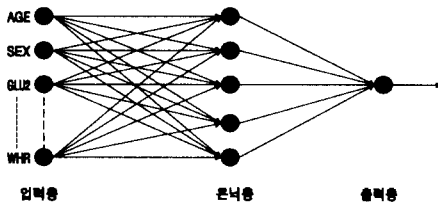


그림 1. MLP model : 17x5x1

4) 학습방법

- maximum epoch = 50000
- error goal = 0.01
- learning rate = 0.001
- momentum = 0
- target : 정상은 -1, 환자는 +1
- backpropagation

B. RBFN(radial basis function network)

1) 입력층(input layer)

MLP와 같다.

2) 은닉층(hidden layer) :

single hidden layer : 입력층노드수를 17개로 고정하고 은닉층노드수는 사용한 matlab5.1에서 지원하는 학습함수에 의해 결정된다. 전이함수(transfer function)는 방사형기저함수(RBF)를 사용하였다.

3) 출력층(output layer)

MLP와 같다.

4) 학습방법

- maximum epoch = 5000
- error goal = 0.01
- target : 정상은 -1, 환자는 +1

3. 결과의 판정방법

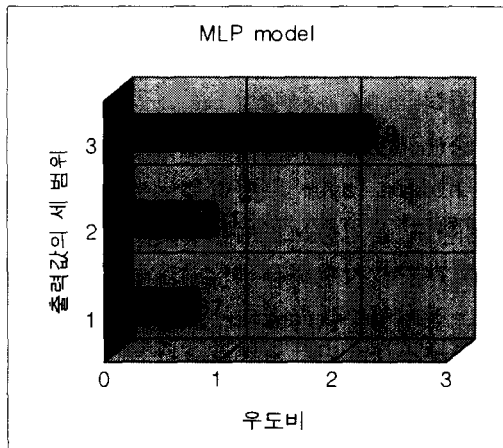
우도비(LR:likelihood ratio) : 10쌍의 data set을 사용하여 얻은 결과를 평균하여 해석하였다. 출력값을 세 가지 범위로 나누어 환자군의 분율과 정상군의 분율의 비로 산출한다.

질 병 예측값	환자군		우도비(LR)
	환자군	정상군	
고위험군	a	b	$a/(a+c+e)/b/(b+d+f)$
중 간 군	c	d	$c/(a+c+e)/d/(b+d+f)$
저위험군	e	f	$e/(a+c+e)/f/(b+d+f)$

4. 결과

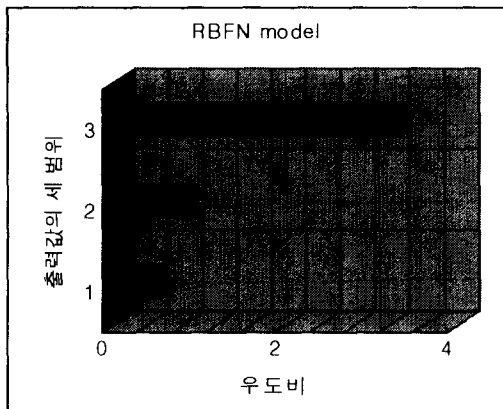
A. MLP model : 10쌍의 평균(TR:TE=2:1) 시험데이터출력분포(Test=40)

예측값	환자군	정상군	우도비(LR)
고위험군	7.9(0.395)	3.6(0.18)	2.19
중 간 군	6.3(0.315)	7.8(0.39)	0.81
저위험군	5.8(0.29)	8.6(0.43)	0.67



B. RBFN model: 10쌍의 평균(TR:TE=2:1)  
시험데이터 출력분포(Test=30)

예측값	환자군	정상군	우도비(LR)
고위험군	6.5(0.433)	2(0.133)	3.26
중간군	5.0(0.333)	6(0.40)	0.83
저위험군	3.5(0.233)	7(0.467)	0.50



### III. 결론

MLP와 RBFN 모두 학습데이터를 적용하면 100% 환자와 정상을 분류한다. 시험데이터를 적용하면 MLP는 약 2배, RBFN은 약 3배 정도의 우도비(LR)를 보인다. 이는 두 예측모델의 결과값이 고위험군에 속하는 경우 MLP는 향후 2년뒤에 정상일 가능성보다 당뇨병 환자로 발병할

가능성이 약 2배 정도 높고, RBFN은 약 3배 정도 높다고 해석할 수 있다. 따라서 검사결과가 현재는 정상범위라 하더라도 앞으로 발병 위험군에 속하는 피검자에게 적극적인 예방과 주의 깊은 관찰 및 추적검사가 필요하다고 조언할 수 있다. 향후 연구방향은 더 많은 데이터를 사용한 예측모델의 개발이 이루어지고, 이런 정상데이터의 분류를 위한 특수한 기법 개발 및 도입이 필요하며, 타당도 향상을 위한 노력이 필요하다. 본 연구에서 개발된 예측모델은 검진 결과 대부분의 정상데이터를 사용해 향후 발병을 예측하는데 활용할 수 있고, 예측모델에서 고위험군으로 분류된 피검자는 예방과 조기진단을 위해 노력하는 효과를 거둘 수 있을 것으로 기대된다.

### IV. 참고 문헌

1. Fauci, et al.; Harrison's Principles of Internal Medicine; 14th edition, vol. 2: 2060-2080
2. Chan Soo Shin, et al.; Risk factors for the development of NIDDM in Yonchon County, Korea; Diabetes Care, vol 20, no 12, 1987 : 1842 - 1846
3. Curt A. Levey, et al.; A neural network knowledge base for medical diagnosis; J Neural Network Computing, 1990: 6-18
4. Simon Haykin; Neural Networks: Macmillan, 1994: 1-2
5. Kyong-Sik Om, et al.; Statistical RBF network with applications to an expert system for characterizing diabetes mellitus; EELS transactions, 1997