

신경망을 이용한 만성질병에 영향을 미치는 식이요인 분석연구*

이심열[†] · 백희영^{**} · 유송민^{***}

동국대학교 가정교육과, 서울대학교 식품영양학과, ** 경희대학교 기계산업시스템공학부^{***}

Analysis of Dietary Factors of Chronic Disease Using a Neural Network

Sim Yeol Lee,[†] Hee Young Paik,^{*} Song Min Yoo^{**}

Department of Home Economics Education, College of Education, Dongguk University, Seoul, Korea

Department of Food and Nutrition,^{**} Seoul National University, Seoul, Korea

College of Mechanical and Industrial System Engineering,^{***} Kyunghee University, Seoul, Korea

ABSTRACT

A neural network system was applied in order to analyze the nutritional and other factors influencing chronic diseases. Five different nutrition evaluation methods including SD Score, %RDA, NAR, INQ and %RDA-SD Score were utilized to facilitate nutrient data for the system. Observing top three chronic disease prediction ratio, WHR using SD Score was the most frequently quoted factor revealing the highest prediction rate as 62.0%. Other high prediction rates using other data processing methods are as follows. Prediction rate with %RDA, NAR, INQ and %RDA-SD Score were 58.5%(diabetes), 53.3%(hyperlipidemia), 51.6%(diabetes), and 58.0%(diabetes) respectively. Higher prediction rate was observed using either NAR or INQ for obesity as 51.7% and 50.9% compared to the previous result using SD Score. After reviewing appearance rate for all chronic disease and for various data processing method used, it was found that iron and vitamin C were the most frequently cited factors resulting in high prediction rate. (*Korean J Community Nutrition* 4(3) : 421~430, 1999)

KEY WORDS : nutritional factors · SD Score · %RDA · NAR · INQ · neural network.

서론

최근 우리나라도 서구화된 식생활 양식이 도입되면서 서구에서 흔히 발생하는 고혈압, 당뇨병, 고지혈증 등의 만성 퇴행성 질환으로 인한 사망이 증가추세에 있다. 특정 영양소와 질병간의 관계에 대한 강한 추정근거에도 불구하고 실제 연구에서 이들간의 상관성이 아주 낮거나 거의 나타나지 않는 경우가 많다(Jacob 등 1979). 이는 사용된 식이섭취 조사 방법이 영양소 섭취량을 제대로 정확하게 평가하지 못했거나, 집단 내에 위험군과 비위험군 간의 충분한 대비가 이루어지지 않아 영양소와 질병간의 실제 상관관계가 측정

되지 못했기 때문일 수 있다. 섭취영양소 분석 시 사용되는 통계기법들은 그 자료가 정규분포를 따른다는 전제하에 실시되나 실제로 그 자료들이 어느 한쪽으로 치우치는 분포를 하고 있는 경우가 많아 분석 전에 자료의 처리가 요구되기도 한다(Millns 등 1995). 특히 일반 통계기법은 표본크기가 작거나 기능적 관계를 알지 못하거나 또는 많은 요인들간의 관계가 복잡하고 상호 관련성이 높을 경우, 분석된 자료의 올바른 해석이 어려우며, 이 안에 내재된 특정 유형이나 영향인자를 찾아내는 것이 쉽지 않다. 이러한 문제점들은 신경망을 이용하여 어느 정도 극복될 수 있다.

신경망은 대규모 자료에 숨겨진 패턴 혹은 지식을 찾아내는 지능적인 자료분석법으로 연관규칙 탐사(심재은 등 1997b)와 더불어 최근 활발한 연구가 진행되고 있다. 신경망기법은 1940년대에 도입된 인공지능적인 기법으로 여러 분야에 응용되고 있으며 새로운 알고리즘과 컴퓨터의 계산속도의 비약적인 발전과 더불어 점차로 그 응용분야가 넓어지고 있다(Cross 등 1995 : Fausette 1994 : Grossberg 1970 : Hirose 등 1991). 초기에는 단순한 논리연산자 정

*이 논문은 1997년도 한국학술진흥재단의 박사후 연구과정 연구비 지원에 의하여 연구되었음.

[†]Corresponding author : Sim Yeol Lee, Department of Home Economics Education, College of Education, Dongguk University, Seoul 100-715, Korea
Tel : (02) 2260-3413, Fax : (02) 2265-1170
E-mail : simlee@cakra.dongguk.ac.kr

도였으나 비약적인 발전 후 여러 분야로 응용되고 있으며, 특히 1980년대 중반에는 의학 및 진료분야에도 응용되기 시작하여 수많은 임상기록에 근거한 자료에 따라 예측할 수 있는 진료시스템이 시도되었다(Anderson 1986). Smith 등(1988)은 Pima 인디언을 대상으로 당뇨병 발병예측 등에 이러한 기법을 이용하였다.

선행 논문에서 섭취 영양소의 SD score를 가지고 만성질환예측에 적용하여 일부 영양소와 질병간에 상관관계를 도출하였다(Lee 등 1998). 그러나 이는 연령이나 성별 등의 요인을 고려하지 않은 것으로, 연령과 성별에 따라 영양소 섭취량 차이가 크며 이들 변수가 질병의 발병과도 관련성을 가지기 때문에 신뢰성있고 객관적인 분석을 위해서는 성별과 연령에 대한 보정이 필요할 것으로 보인다.

영양소 섭취 평가방법으로 널리 사용되고있는 것으로 개인의 섭취량을 성별, 연령에 따른 영양권장량(recommended dietary allowance : RDA)과 비교하기도 한다. 이 외에도 권장량에 대한 섭취비로 나타내나 권장량 이상을 섭취해도 1을 최대치로 하는 영양소 적정섭취비(nutrient adequacy ratio : NAR), 에너지가 충족되는 상태에서 영양소의 균형상태를 파악하는 영양의 질적지수(index of nutritional quality : INQ)가 있다. 성별 및 연령이 보정된 %RDA(영양권장량 백분율), NAR, INQ 등의 영양 평가방법들을 도입하여 이들 가운데 가장 적합한 입력자료를 찾아내고 이를 토대로 조사집단의 특성을 충분히 반영한 발전된 신경망 시스템으로 완성시키는 것이 필요하다.

따라서 본 연구에서는 우리 나라 농촌성인의 만성질환에 영향을 주는 식이요인 연구 및 만성질환 예방을 위한 대책마련을 위해 경기도 연천군에 거주하는 성인들을 대상으로 식이섭취 및 건강조사를 실시한 후 이들 자료를 토대로 하여 당뇨병, 고혈압, 고지혈증 등과 같은 만성 퇴행성 질환 발병 여부를 예측하기에 적합한 신경망 시스템을 구축하여 만성질환에 영향을 주는 식이 및 기타 요인을 추출하고, 또한 이들 요인을 이용하여 질병보유 여부를 예측하고자 하였다.

조사대상 및 방법

1. 조사 지역 및 대상자

본 연구는 서울 북부지역에 위치한 경기도 연천군에 거주하는 성인을 대상으로 시행된 건강 및 식생활 조사에 참여한 대상자 중 결과 자료의 분석이 가능한 2037명에 대하여 분석하였다. 조사지역과 대상자에 대한 정보는 선행논문(Lee 등 1998)에 제시되었다. 건강조사에서 각 대상자의 인구통계학적 정보는 직접 면담을 통해 얻었으며, 공복시

혈액채취, 혈압 측정, 신체 계측과 당내용 검사(glucose tolerance test)를 실시하였다. 이전에 당뇨병이나 고혈압으로 진단 받은 적이 없으면서, 본 연구에서 WHO 기준-공복시 혈당이 140mg/dl 이상이거나 포도당 음복 2시간 후의 혈당이 200mg/dl 이상 : 수축기 혈압이 140mmHg 이상이거나 이완기 혈압이 90mmHg 이상 : 혈청지질농도가 220mg/dl 이상-에 해당하는 대상자들을 각각 당뇨병, 고혈압과 고지혈증으로 선정하였다. BMI기준으로 남자는 25이상, 여자는 27이상인 경우를 비만으로 선정하였다.

2. 식이섭취조사

식이섭취조사는 대상자별로 24시간 회상법을 이용하여 직접면담에 의해 실시되었다. 조사는 하루전날 섭취한 음식명과 각 음식에 사용된 재료명을 분량과 함께 물어보아 기록하였다. 조사된 식이섭취량은 식품영양가표(한국영양학회 1995)를 토대로 하여 영양소 섭취량으로 환산하였다.

3. 신경망시스템

1) 자료전처리

신경망에 적용된 요인으로는 12가지 주요 영양소와 질병 발병에 주 위험요소로 생각되는 건강관련지수인 공복시 혈당(fasting blood glucose : FBG), 포도당 음복 2시간후 혈당(blood glucose concentration at 2-hour after oral glucose tolerance test : GTT), 수축기 혈압(SBP), 이완기 혈압(DBP), 혈중 콜레스테롤 농도, BMI(체질량지수), WHR(허리/엉덩이 둘레비)를 사용하였다. 각 요인들은 다음과 같은 영양평가 방법들로 처리한 후 신경망시스템에 적용시켰다.

(1) 건강관련지수 - SD Score

식이요인 외에 만성질환에 영향을 미치는 인자들을 분석하기 위하여 7가지의 건강관련지수를 SD score에 적용시켰다.

(2) 영양권장량 백분율(%RDA)

산출된 개인별 1일 영양소 섭취량을 각 영양소별 한국인 영양권장량(한국영양학회 1995)과 비교하여 이에 대한 백분율로 계산하였다.

(3) 영양소 적정섭취비(NAR)

영양소 적정섭취비는 각 영양소 섭취량을 영양권장량에 대한 비율로 계산한 후, 1을 최고 상한치로 설정하여, 1이 넘는 경우에는 1로 간주하였다.

(4) 영양의 질적지수(INQ)

영양이 충족된 상태에서 영양소의 균형상태를 파악하는

것으로 식사의 에너지함량과 영양소함량의 비율을 영양권장량에서의 비율과 비교하는 것이다.

(5) %RDA값의 SD Score

%RDA값에 대한 SD Score를 계산한 것으로 영양소를 이용하여 질병을 판별하는 목적에 응용하고자 적용시킨 방법으로 각각의 영양소인자에 대하여 %RDA를 계산한 후 이에 대하여 SD Score값을 계산하였다.

2) 자료재구성

자료를 신경망시스템에 적용시키기에 적합하도록 하기 위하여 각각의 영양소 혹은 건강관련 인자를 앞에서 제시한 5가지 영양섭취평가법에 따라 계산한 후 크기순으로 배열하여 일정범위 이내의 수준을 유지하는 영양소 수치를 나타내는 자료군만 수용하였다. 이 기준으로 %RDA, NAR 및 INQ는 각 300%이내이며 SD Score로는 ±3.0이내이다. 또한 %RDA값으로 SD Score를 계산한 경우는 300%와 ±3.0의 기준을 모두 충족시킨 대상자만 사용하도록 하였다. 이러한 기준을 적용시켰을 때 각 평가법에 따른 남녀 대상자들의 수는 Table 1과 같았다. 총 대상자 2037명 중 자료가 부분적으로 부족하거나 처리된 계산결과가 기준을 벗어난 것을 제외한 것으로서 평가방법에 따른 대상자 수는 1725명(INQ 적용 시)에서 1959명(NAR 적용 시)으로 나타났다. 대상자별로 섭취 영양소 또는 건강관련지수값이 사용되었으며 각 질병별로 자료의 처리 방법(영양평가방법)에 따라 Table 2에서 설명되고 있는 인자들이 사용되었다. 신경망시스템을 각각의 질병에 적용시킬 경우 그 질병을 진단하는 기준이 되는 건강관련지수인자는 제외시켰다.

3) 신경망시스템 적용

본 연구에 적용된 신경망은 입력과 이에 대한 출력값이 미리 알려진 자료를 사용하는 감독학습법(Fausette 1994)을 적용하여 구성하였으며 신경망구조의 학습법은 역전파 알고리즘을 사용하였다. 신경망모델은 기본적으로 일정한 수의 입력뉴런으로 구성된 입력층에서 섭취영양소 등의 입력자료를 수용하며 출력층에서는 단일 질병의 보유유무를 판별할 수 있도록 2개의 뉴런으로 구성하였다. 입력과 출력층은 1개의 은닉층으로 연결되었다.

구축된 신경망모델은 5가지 영양평가방법에 의하여 처리된 자료를 이용하여 학습되었으며 학습완료 후 학습에 사용되지 않은 대상자들의 자료를 시스템에 적용하여 학습된 신경망시스템이 대상자들의 질병보유여부를 판단하게 한 후 기존의 질병보유결과와 비교하여 본 시스템의 판단의 정확성을 평가하도록 하였다(Fig. 1). 학습에 활용되는 자료로

Table 1. Distribution of subjects in the neural network systems with various evaluation methods

Evaluation method	No. of subjects(%)		
	Male	Female	Total
SD Score	708(40.4)	1044(59.6)	1752(100)
%RDA	764(42.0)	1053(58.0)	1817(100)
NAR	833(42.5)	1126(57.5)	1959(100)
INQ	733(42.5)	992(57.5)	1725(100)
%RDA-SD Score	745(41.3)	1057(58.7)	1802(100)

Table 2. Various factors used in the neural network system

Evaluation method	No. of factors	Factors
SD Score ¹⁾	7	Cholesterol, BMI, WHR, FBG ²⁾ , GTT ³⁾ , SBP, DBP
%RDA	9	Energy, Protein, Calcium, Iron, Vitamin A, Vitamin B ₁ , Vitamin B ₂ , Niacin, Vitamin C
NAR	10	Energy, Protein, Calcium, Iron, Vitamin A, Vitamin B ₁ , Vitamin B ₂ , Niacin, Vitamin C, MAR
INQ	8	Protein, Calcium, Iron, Vitamin A, Vitamin B ₁ , Vitamin B ₂ , Niacin, Vitamin C
%RDA-SD Score	9	Energy, Protein, Calcium, Iron, Vitamin A, Vitamin B ₁ , Vitamin B ₂ , Niacin, Vitamin C

- 1) anthropometric and biochemical factors were used
- 2) fasting blood glucose
- 3) blood glucose concentration at 2 hour in an oral Glucose Tolerance Test

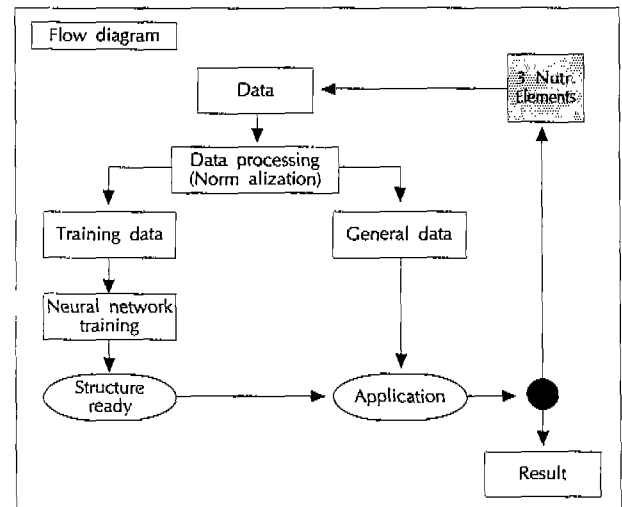


Fig. 1. Data flow structure.

는 WHO 기준에 사용되는 질병판단인자를 수치 크기에 따라 배열한 후 판단기준상 인자수치가 가장 높은 상위 20명과 가장 낮은 하위 20명의 대상자(질병보유자 20명과 미보유자 20명 등 총 40명)가 섭취한 영양소 또는 건강관련지수값이 사용되었다.

대상 영양소 또는 건강관련지수 중에서 3개를 선정하여 신경망의 입력층에 적용시켜 학습시키므로 학습된 신경망시스템을 적용시키면 여러 가지 경우에 대하여 서로 다른 질병예측률을 얻게된다. 12개의 영양소가 사용된 %RAD

의 경우에는 220(₁₂C₃)가지의 다른 경우가 발생한다.

신경망의 학습종료시점은 출력층에서의 출력과 목표값의 차이가 허용오차 이하로 되거나 학습횟수가 10000회에 도달하는 것 중에서 우선하는 것으로 하였다.

결과 및 고찰

본 연구 대상자들의 특성과 식이섭취상태는 이미 선행연구(이심열 등 1998)를 통해 보고되었다. 여자 대상자수가 남자대상자 보다 약간 많은 편이며 연령은 50대가 가장 많았고 이들의 영양소 섭취는 전반적으로 낮은 편이었다.

영양소 또는 건강관련지수가 질병 발생에 미치는 영향을 보기 위하여 적합한 신경망 시스템을 구축, 활용하고자 하였으며 질병예측이 뛰어난 인자를 판단하기 위하여 3개씩의 인자들을 선정하여 적용시킨 후 가장 높은 질병 인식률을 보인 인자들의 인용빈도를 비교함으로써 이들 인자의 상대적인 중요도를 살펴보았다.

1. 질병예측률

1) 건강관련지수-SD Score

Table 3은 건강관련지수의 SD Score를 신경망시스템에 적용시킨 후 예측률 3위까지를 나타낸 것이다. 당뇨병에서는 공복 시 혈당과 포도당 음복 2시간 후 혈당(GTT)이 질병보유여부를 결정하는 데 사용되므로 이들을 제외한 5개의 인자만이 사용되어 3개의 입력뉴런을 지닌 신경망에 사용되었다. 가장 높은 예측률은 콜레스테롤, BMI와 WHR가 조합된 경우로서 51.9%를 보였고 BMI는 3회 모두 인

Table 3. Anthropometric and biochemical factor combination of top third classification ratios for predicting each disease by the neural network

Disease	Factors combination			Prediction rate(%)
DM ¹⁾	Cholesterol	BMI	WHR	51.9
	BMI ⁵⁾	WHR	SBP	51.0
	Cholesterol	BMI	SBP	48.6
Hypertension ²⁾	Cholesterol	BMI	GTT	49.4
	WHR	FBG	GTT	47.7
	BMI ⁶⁾	WHR	FBG	47.4
Hypelipidemia ³⁾	WHR	FBG	SBP	54.4
	WHR	GTT	SBP	52.7
	WHR	FBG	GTT	51.6
Obesity ⁴⁾	WHR	FBG	DBP	62.0
	Cholesterol	WHR	DBP	58.9
	Cholesterol	WHR	GTT	57.8

1) FBG, GTT were not used

2) SBP(systolic blood glucose), DBP(diastolic blood glucose)were not used, 3) Cholesterol was not used, 4) BMI was not used

5) Body Mass Index, 6) Waist Hip Ratio

용되었다. SBP와 DBP가 제외된 고혈압의 경우에는 콜레스테롤, BMI와 GTT가 사용된 경우가 가장 높은 예측률을 보이고 있으나 49.4%로 다른 질병에 비하여 다소 낮은 결과를 보였다. 예측률 3위까지 중에서 BMI, 공복 시 혈당, GTT와 WHR이 각각 2회 인용되고 있어 특별히 3회 모두 인용되는 인자는 없었다. 고지혈증에서는 WHR, 공복시 혈당과 SBP를 사용한 경우가 54.4%의 예측률을 보여 가장 높게 나타났으며 WHR이 3회 모두 인용되었다. 비만은 BMI를 제외한 6개의 인자가 사용되었으며 WHR, 공복 시 혈당과 GTT를 사용한 경우가 62.0%로서 질병 중 가장 높은 예측률을 보였다. 각 질병에 대하여 일부 지수가 제외된 상태에서 비교되므로 절대적으로 비교하기는 어렵지만 4가지의 질병 예측 시 WHR가 총 인용 가능한 12회 중 10회로 가장 빈번하게 사용되었다.

2) %RDA

각 영양소별 %RDA를 신경망 인자로 적용시킨 후 예측률 3위까지의 결과가 Table 4에 제시되었다. 당뇨병과 고지혈증에 대한 예측률은 50%를 넘었으나 고혈압과 비만에 대해서는 50%를 약간 밑돌았다. 당뇨병에서 에너지, 철분과 비타민 A 조합의 예측률이 58.5%로서 %RDA를 사용하여 적용시킨 결과 중에서 가장 높은 예측률을 보였다. 3위까지의 예측률은 55.1%로서 비교적 높게 나타나고 있으나 에너지, 비타민 A와 철분이 2회씩 인용되었을 뿐 3회 모두 사용된 영양소는 없는 것으로 나타났다. 고혈압에서는 에너지, 철분과 비타민 B₁의 조합 시 예측률이 가장 높았으나 4가지 질병 가운데에서는 가장 낮았다. 고지혈증의 경우에는 칼슘, 나이아신 그리고 비타민 C를 조합한 경우 예측률이 가장 높았다. 비만에서는 비타민 B₂, 나이아신, 비타민 C의 조합이 47.1%의 예측률로서 비교적 낮았다. 4가지 질병에서 철분과 비타민 C가 6회, 칼슘과 에너지가 5회씩 인용되었다.

Table 4. Nutrient combination of top three ratios for predicting chronic disease by the neural network using % RDA

Disease	Nutrients combination			Prediction rate(%)
DM	Energy	Iron	Vitamin A	58.5
	Vitamin A	Niacin	Vitamin C	56.0
	Energy	Calcium	Iron	55.1
Hypertension	Energy	Iron	Vitamin B ₁	47.0
	Calcium	Iron	Vitamin C	46.4
	Energy	Niacin	Vitamin C	45.0
Hyperlipidemia	Calcium	Niacin	Vitamin C	54.5
	Protein	Iron	Vitamin B ₁	53.1
	Energy	Vitamin B ₁	Vitamin C	52.8
Obesity	Vitamin B ₂	Niacin	Vitamin C	47.1
	Calcium	Vitamin A	Vitamin B ₁	45.0
	Protein	Calcium	Iron	44.9

SD Score만을 사용하여 자료를 처리한 선행연구결과 (Lee 등 1998)에 비해 성별, 연령을 보정한 %RDA를 입력 인자로 사용한 본 연구결과 비만을 제외한 모든 만성질환에서 낮은 예측률을 보였다. 질병 보유군과 비보유군간의 영양소섭취량을 비교한 연구에서도 유의한 차이를 보이던 여러 영양소들이 연령과 성별을 보정한 후에는 그 유의성이 소실되었다(심재은 등 1997a). 이는 이들을 고려하지 않은 분석에서 보인 영양소 섭취량 차이가 열량섭취, 연령, 성별 등 영양소 이외의 요인에 의한 것이었음을 나타내고 있다.

3) NAR(영양소 적정섭취비)

Table 5는 섭취영양소의 NAR값을 적용시킨 결과이다. %RDA값을 적용시킨 경우에 비해 비만을 제외한 다른 질병에서는 예측률이 저하되었으며 특히 당뇨병의 예측률이 크게 저하되었다. 당뇨의 경우 가장 높은 예측률은 칼슘, 비타민 A, 비타민 B₂의 조합으로 48.1%였으며 3위까지 에너지, 칼슘, 철분, 비타민 A가 2회 인용되었다. 고혈압에서는 비타민 A, 비타민 B₁, 비타민 C의 조합이 46.7%로 가장 높은 예측률을 보였고 3위까지는 비타민 A와 비타민 B₁이 모두 3회씩 인용되었다. 고지혈증의 경우에는 칼슘, 철분과 비타민 B₂의 조합 시 인식률이 모든 질병에 대한 결과 중 가장 높았으며 칼슘과 비타민 B₂가 2회 인용되었다. 비만에서는 칼슘, 철분과 비타민 C의 조합이 예측률 51.7%로 1위를 보이고 있으며 비타민 C가 3위까지 모두 인용되었다. 4가지 질병에 대하여 최대 12번까지 인용될 수 있는 것에 비하여 칼슘, 비타민 C, 비타민 A와 철분이 각각 5회 인용되었다.

4) INQ

Table 6은 INQ를 입력인자로 사용한 경우의 결과를 나타낸 것이다. NAR값을 적용시킨 경우에 비해 예측률이 약간 상승하였으나 고지혈증에서만 저하되었다. INQ값은 섭취에너지를 보정한 것으로 고지혈증 예측 시 에너지가 어느 정도 영향을 미친 것으로 보인다. 당뇨에서 비타민 A, 나이아신과 비타민 C 조합의 예측률이 51.6%로서 가장 높았으며 나이아신은 매번 인용되었다. 고혈압에서는 칼슘, 철분 그리고 비타민 B₂의 조합이, 고지혈증에서는 철분, 비타민 A와 비타민 B₂의 조합이, 비만에서는 비타민 A, 나이아신과 비타민 C의 조합이 가장 높은 예측률을 보였다. 4가지 질병 모두에 대하여 종합적으로 철분이 8회, 비타민 C와 나이아신이 각각 6회 인용되었다.

5) %RDA-SD Score

%RDA 값을 이용하여 SD Score를 계산한 후 이 값을 입력인자로 처리한 경우에 대한 결과를 분석하여 보았다

(Table 7). 이 경우 %RDA 값만을 입력인자로 이용한 경우와 비교할 때 별 차이를 나타내지 않았다. 전체적으로는 비타민 C가 6회로 가장 많이 인용되었고 철분과 비타민 B₁이 각각 5회 인용되었다. %RDA 값의 SD Score를 입력

Table 5. Nutrient combination of top three ratios for predicting chronic disease by the neural network using NAR

Disease	Nutrients combination			Prediction rate(%)
DM	Calcium	Vitamin A	Vitamin B ²	48.1
	Energy	Iron	Vitamin A	47.4
	Energy	Calcium	Iron	45.5
Hypertension	Vitamin A	Vitamin B ₁	Vitamin C	46.7
	Vitamin A	Vitamin B ₁	Niacin	46.0
	Energy	Vitamin A	Vitamin B ₁	46.0
Hyperlipidemia	Calcium	Iron	Vitamin B ₂	53.3
	Energy	Calcium	Vitamin B ²	48.9
	Vitamin A	Vitamin B ₁	Vitamin C	47.8
Obesity	Calcium	Iron	Vitamin C	51.7
	Vitamin B ₂	Niacin	Vitamin C	48.2
	Iron	Niacin	Vitamin C	47.5

Table 6. Nutrient combination of top three ratios for predicting chronic disease by the neural network using INQ

Disease	Nutrients combination			Prediction rate(%)
DM	Vitamin A	Niacin	Vitamin C	51.6
	Iron	Vitamin B ₂	Niacin	50.3
	Protein	Calcium	Niacin	49.6
Hypertension	Calcium	Iron	Vitamin B ₂	51.2
	Iron	Niacin	Vitamin C	47.2
	Calcium	Iron	Vitamin C	46.6
Hypelipidemia	Iron	Vitamin A	Vitamin B ₁	48.3
	Iron	Vitamin A	Vitamin B ₂	47.4
	Calcium	Vitamin B ₂	Vitamin C	47.2
Obesity	Vitamin A	Niacin	Vitamin C	50.9
	Iron	Niacin	Vitamin C	48.1
	Protein	Calcium	Iron	47.8

Table 7. Nutrient combination of top three ratios for predicting chronic disease by the neural network using %RDA-SD Score

Disease	Nutrients combination			Prediction rate(%)
DM	Iron	Vitamin B ₂	Vitamin C	58.0
	Protein	Vitamin B ₁	Vitamin C	56.0
	Energy	Protein	Niacin	54.3
Hypertension	Energy	Protein	Iron	48.3
	Energy	Vitamin A	Vitamin C	47.3
	Vitamin A	Vitamin B ₁	Vitamin B ₂	47.3
Hypelipidemia	Iron	Vitamin B ₁	Niacin	56.7
	Energy	Niacin	Vitamin C	51.8
	Protein	Iron	Vitamin B ₁	51.4
Obesity	Vitamin A	Vitamin B ₁	Vitamin B ₂	47.7
	Vitamin B ₂	Niacin	Vitamin C	45.9
	Calcium	Iron	Vitamin C	44.0

인자로 사용한 경우의 가장 높은 예측률을 SD Score 한가지만을 사용한 선행 연구결과(Lee 등 1998)와 비교할 때 당뇨병의 경우 73.1%에서 58.0%로, 고혈압에서는 53.7%에서 48.3%로, 고지혈증에서는 78.3%에서 56.7%로 모두 크게 낮아졌으며 비만에서만 48.8%에서 47.7%로 약간의 차이를 보였다.

2. 인용빈도율

신경망 분석을 통한 높은 질병인식률에서 특정 영양소의 인용빈도가 높아 기여도가 높게 나온 경우 그 영양소의 섭취량이 상대적으로 많거나 혹은 섭취량이 적을 때 발병률이 높은 것을 의미한다. 각 질병별 상위 20위내의 예측 시 이용된 영양소들의 인용횟수를 가지고 인용빈도를 계산하였다. 인

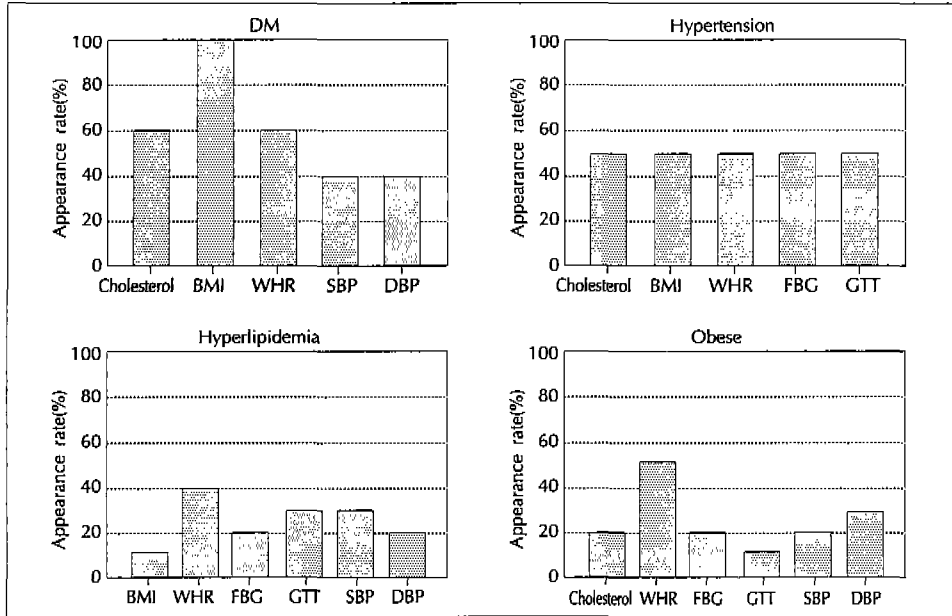


Fig. 2. Appearance rate of anthropometric and biochemical factors for top 5th or 10th recognition rate of the neural network system for predicting chronic disease(top 5 : DM, hypertension, top 10 : hyperlipidemia, obese).

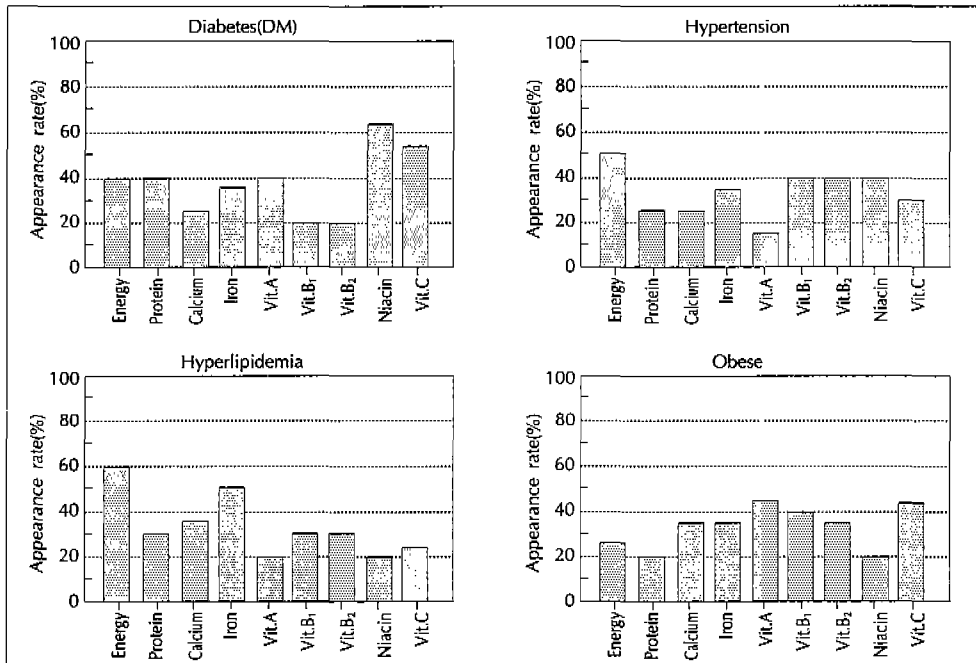


Fig. 3. Appearance rate of each nutrients with %RDA for top 20th recognition rate of the neural network system for predicting chronic disease.

용빈도율(appearance rate)은 인용가능한 횟수(최대 20회) 가운데 실제로 인용된 횟수를 백분율로 표시한 것이다. 건강 관련지수로는 5개 또는 6개만의 인자를 사용하므로 3개씩 추출하면 가능한 조합수가 각각 10개, 20개이며 이때 이의 절반인 5위와 10위까지 각각 분석하였다(Fig. 2).

5개의 인자만이 사용된 당뇨의 경우 BMI가 5회 모두 인

용되어 100%를 기록하고 있다. 고혈압에서는 5개의 인자가 모두 균등하게 인용되었다. 6개의 건강관련지수가 사용된 고지혈증과 비만에서는 모두 WHR이 가장 자주 인용된 것으로 나타났다. Colditz 등(1992)의 연구에서도 BMI가 높을수록 인슐린 비의존성 당뇨병의 위험률이 높았으며 연령, BMI, 알콜섭취가 고혈압발병에 큰 영향인자로 보고되

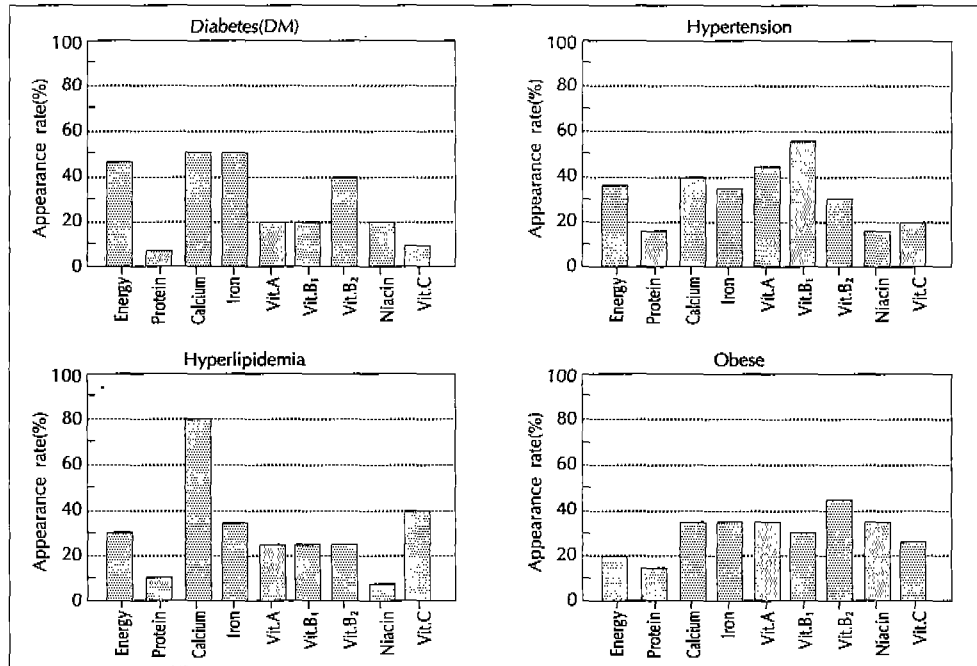


Fig. 4. Appearance rate of each nutrients with NAR for top 20th recognition rate of the neural network system for predicting chronic disease.

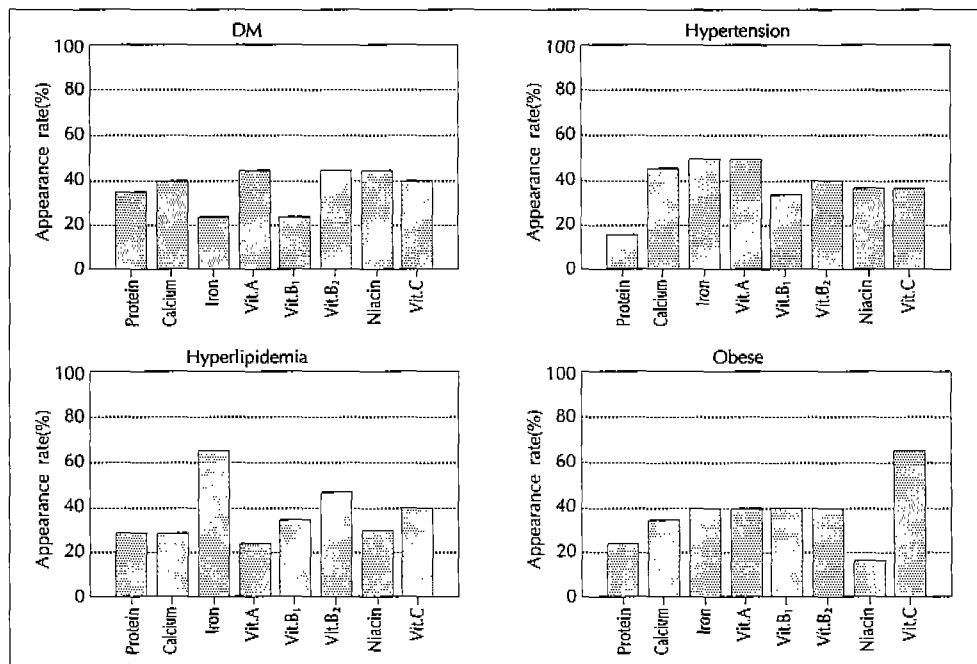


Fig. 5. Appearance rate of each nutrients with INQ for top 20th recognition rate of the neural network system for predicting chronic disease.

었다(심재은 등 1997a).

Fig. 3은 %RDA를 이용한 경우 20위 내에서의 인용빈도를 나타내고 있으며 당뇨와 비만에서는 두드러지게 많이 인용된 인자는 없었으며 고지혈증과 고혈압에서는 에너지가 각각 50%와 60%로 가장 높았다. 선행 연구에서도 고혈압의 경우에는 에너지가 많이 인용되었다. Fig. 4에서와 같

이 NAR을 사용하여 처리한 자료를 사용한 결과 고지혈증에서 칼슘이 80%로 현저히 높게 나타났으며 고혈압에서는 비타민 B₁의 인용빈도율이 55%로 약간 높게 나타났고 당뇨에서는 칼슘, 철분, 에너지가 비만에서는 비타민 B₂의 인용빈도율이 다른 영양소에 비해 상대적으로 약간 높게 나타났다. INQ를 사용한 결과에서는 고지혈증에서의 철분과 비

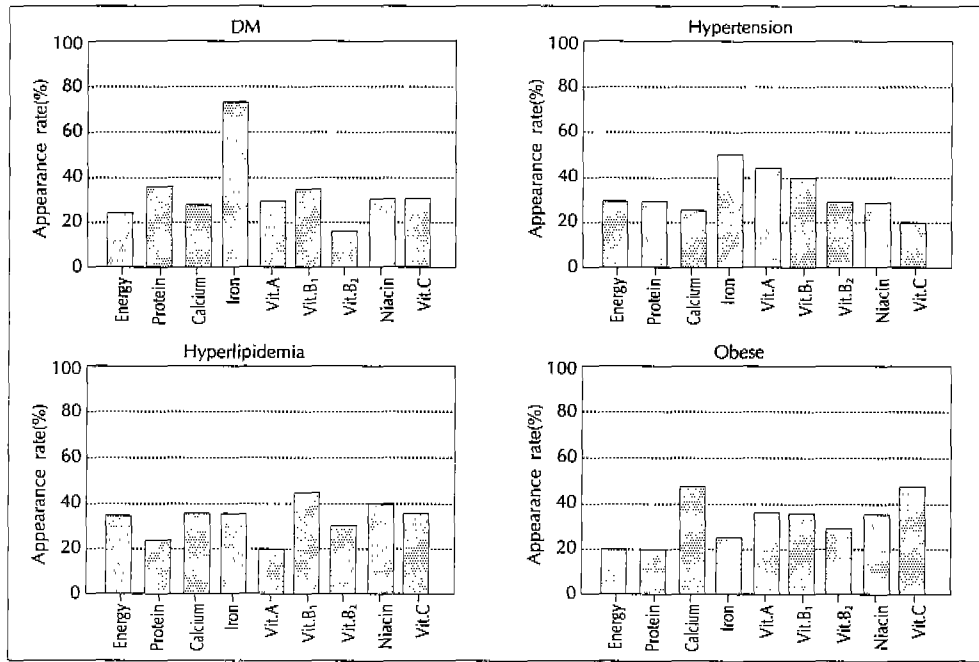


Fig. 6. Appearance rate of each nutrients with RDA-SD Score for top 20th recognition rate of the neural network system for predicting chronic disease.

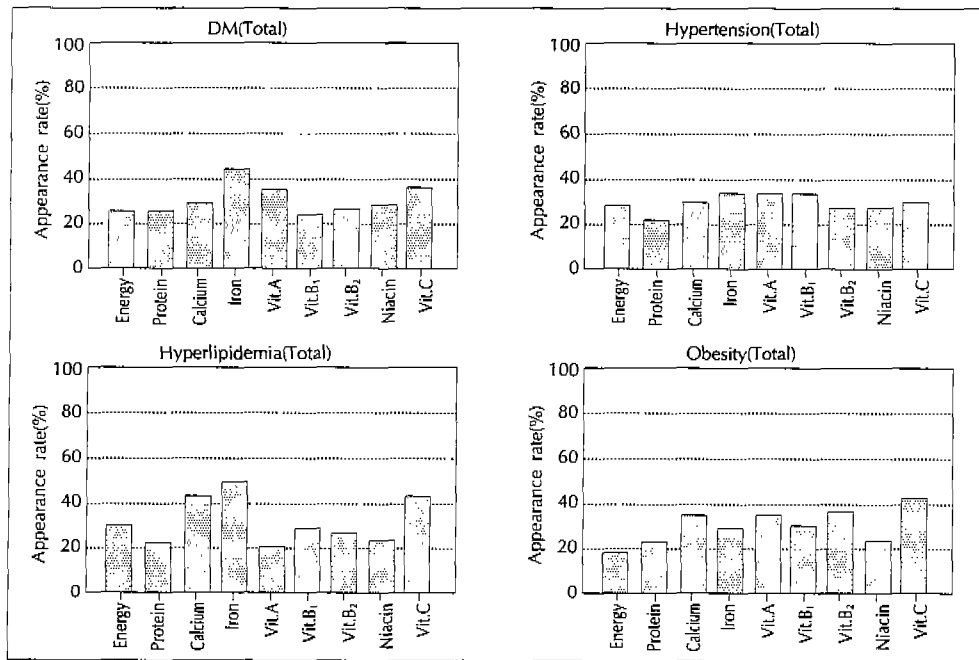


Fig. 7. Average Appearance rate of each nutrients for top 20th recognition rate of the neural network system for predicting chronic disease.

만에서의 비타민 C가 각각 65%로 20위중 13회가 인용되어 가장 높게 나타나고 있으며(Fig. 5) 특히 고지혈증에서 철분의 인용빈도율이 높게 나타난 것은 Lee 등(1998)의 결과와 일치한다. Fig. 6에서 분석된 것과 같이 %RDA와 SD Score를 모두 사용한 경우에는 당뇨와 고혈압에서 철분이 각각 70%와 50%를 나타내고 있고 비만에서는 칼슘과 비타민 C가 50%로 나타났다. 비타민 B₁, 비타민 B₂, 나이아신 등의 경우는 에너지 섭취량과 이들 영양소의 섭취량이 비례하기 때문에 에너지와 더불어 자주 인용된 측면이 있을 것으로 생각된다.

Fig. 7은 %RDA, NAR, INQ, %RDA-SDS를 입력인자로 하여 질병예측 시 상위 20위내의 영양소 인용빈도율을 분석한 결과의 평균을 그림으로 나타낸 것이다. Fig. 7에 나타난 바와 같이 평균적으로 당뇨에서는 철분 41%, 비타민 C 37%의 평균인용빈도율을 보였고 고혈압에서는 철분, 비타민 A, 비타민 B₁ 그리고 비타민 C가 모두 34%를 보였다. 고지혈증에서는 철분이 48%로 가장 높은 평균인용빈도율을 보였고 칼슘과 비타민 C가 41%로 그 다음으로 나타났으며 비만에서는 비타민 C가 41%로 가장 높았다. 따라서 인용빈도율의 평균을 분석한 결과 대체로 영양소 중 철분과 비타민 C가 많이 인용되어 질병예측 시 기여도가 높은 영양소로 나타났다. 이는 항산화비타민이 심혈관계 질환예방에 기여한다는 보고(Jha 등 1995; Kenkt 등 1994)와 비타민 C 섭취량과 당내용성 사이의 음의 상관관계가 있다는 보고자료(Feskings 등 1995)를 뒷받침해주고 있다. 특히 당뇨병과 고지혈증 예측 시 철분의 높은 기여도는 주목할 만하며 이의 관련 기작에 관한 연구가 필요할 것으로 보인다.

요약 및 결론

본 연구는 우리나라 농촌성인의 만성질병에 영향을 주는 식이요인 및 만성질병 예방을 위한 대책마련을 위해 경기도 연천군에 거주하는 성인 2037명을 대상으로 식이섭취 및 건강조사를 실시하여 이들 자료를 토대로 당뇨병, 고혈압, 고지혈증 등과 같은 만성 퇴행성질환 발병여부를 예측할 수 있는 적합한 인자를 찾아내어 신경망 시스템을 구축하며 만성질병에 영향을 주는 식이요인을 추출하고 또한 질병보유 여부를 예측하고자 실시되었다. 그 결과를 요약하면 다음과 같다.

1) 건강관련지수를 SD Score로 평가하여 신경망에 적용시킨 결과, 최대 62.0%(비만)의 질병예측률을 보였으며 이때 WHR의 인용빈도가 높게 나타났다.

2) SD Score를 입력자료로 사용한 선행연구에 비해 성

별, 연령을 보정한 %RDA, NAR, INQ 값을 신경망의 입력자료로 활용한 결과 비만의 경우를 제외한 다른 질병들의 전반적인 예측률이 더 낮게 나타났다.

3) 사용된 여러 영양평가방법들 중에서 권장량에 대한 백분율(%RDA)을 입력인자로 하였을 경우 질병예측률이 가장 높게 나타나 영양소인자를 신경망에 적용 시 %RDA를 입력인자로 사용함이 적합함을 보였다.

4) 각 영양평가 방법별 최대예측률이 질병에 따라 다르게 나타났는데 %RDA에서는 58.5%(당뇨병), NAR에서는 53.3%(고지혈증), INQ에서는 51.6%(당뇨)이었으며 %RDA 값을 SD Score에 적용한 경우에는 58.0%(당뇨)이었다.

5) 질병예측 시 상위 20위 내의 영양소 인용빈도율을 4가지 영양평가방법에 대하여 분석한 결과 영양소 중 철분(당뇨: 41%, 고혈압: 34%, 고지혈증: 48%)과 비타민 C(당뇨: 37%, 고혈압: 34%, 고지혈증: 41%, 비만: 41%)가 많이 인용되어 질병예측 시 가장 기여도가 높은 영양소로 나타났다. 질병예측에 사용된 건강관련인자 중에서는 허리-엉덩이둘레비(WHR)의 기여도가 가장 높았으며, 당뇨병에서 BMI의 기여도도 두드러지게 나타났다.

본 연구에서는 조사대상 지역이 농촌이므로 식이가 비교적 단순하고 별 변화가 없을 것이라는 가정 하에 24시간법을 사용하였으나 이는 최근의 식이섭취를 반영하므로 이에 따른 영향을 배제할 수는 없다. 만성질환은 오랜 기간에 걸쳐 발병하므로 개인의 장기간의 영양소 섭취상태를 평가할 수 있는 식품섭취빈도 조사자료를 신경망에 적용시켜 식이와 질병과의 관계를 밝히는 지속적인 연구가 필요할 것이다. 질병예측률이 평균 50% 내외로 높지는 않게 나타났으며 이는 신경망의 입력인자로 섭취 영양소요인에 주안점을 두었기 때문인 것으로 보인다. 질병발병에는 기타 다른 많은 요인들(건강관련요인, 흡연, 음주 등)이 동시에 작용할 수 있으므로 앞으로 이들 모든 요인들을 함께 고려하여 적용시키는 입력모델을 개발하는 것이 필요하다.

■ 감사의 글

본 연구의 수행을 위해 협조해 주신 서울대학교 의과대학 의료관리학교실의 연천지역 의료관리 시범 사업단과 연천군 보건의료원에 깊은 사의를 표합니다.

참고문헌

- 심재은 · 류지영 · 백희영 · 신찬수 · 이홍규 · 박용수(1997a) : 반정량적 식품섭취 빈도법을 이용한 농촌성인의 만성퇴행성질환에 영향을 미치는 식이요인 연구. *한국역학회지* 19(1) : 42-57
 심재은 · 백희영 · 신원선 · 김교정(1997b) : 연관규칙 탐사를 이용한 우

- 리나라 농촌 성인의 식이섭취평가. *한국역학회지* 19(2) : 220-230
- 이십열 · 주달래 · 백희영 · 신찬수 · 이용규(1998) : 24시간 회상법으로 조사한 연천지역 성인의 식생활 평가(1) : 영양소 섭취 평가. *한국영양학회지* 31(3) : 333-342
- 한국 영양학회(1995) : 식품 영양가표. 한국인 영양 권장량. 제 6 차 개정
- Anderson JA(1986) : Cognitive capabilities of a parallel system. *Disordered Systems and Biological Organization*, Springer-Verlag
- Colditz GA, Manson JE, Stampfer MJ, Rosener B, Willet WC, Day NE(1994) : Comparison of dietary assessment methods in nutritional epidemiology : weighed records vs 24h recalls, food-frequency questionnaire and estimated-diet records. *British J of Nutrition* 72 : 619-43
- Cross SS, Harrison RF, Kennedy RL(1995) : Introduction to neural networks, pp.1075-1079, *Lancet*
- Fausett L(1994) : *Fundamentals of Neural Network*, Englewood Cliffs, NJ, Prentice Hall
- Feskens EJM, Stengard J, Virtanen SM, Pekkanen J, Rosenen L, Nissinen A, Tuomilehto J, Kromhout D(1995) : Dietary factors determining diabetes and impaired glucose tolerance. *Diabetes Care* 18(1) : 1104-12
- Grossberg S(1970) : Nonlinear neural network : principles, mechanisms and architectures. *Neural Networks* 10(1) : 17-61
- Hirose Y, Yamashita K, Hijjiya S(1991) : Back-propagation algorithm which varies the number of hidden units. *Neural Networks* 4 : 61-66
- Jacobs DR, Anderson JT, Blackburn H(1979) : Diet and serum cholesterol : Do zero correlations negate the relationship?. *Am J Epidemiol* 110 : 77-87
- Jha P, Flather M, Lonn E, Farkouh M, Yusuf S(1995) : The antioxidant vitamins and cardiovascular disease : A critical review of epidemiologic and clinical trial data. *Ann Intern Med* 123 : 860-72
- Kenkt P, Reunanen A, Jarcinen R, Sepponen R, Heliavaara M, Aromaa A(1994) : Antioxidant vitamin intake and coronary mortality in a longitudinal population study. *Am J Epidemiol* 139 : 1180-1189
- Lee SY, Paik HY, Yoo SM, Lee HK(1998) : Identification of dietary factors related to hypertension, diabetes, hyperlipidemia, and obesity using neural network. *Korean J of Epidemiol* 20(2) : 226-233
- Mills H, Woodward M, Bolton-Smith C(1995) : Is it necessary to transform nutrient variables prior to statistical analyses? *Am J Epidemiol* 141 : 251-262
- Smith JW, Everhart BS, Dickson WC, Knowler WC, Johannes RS (1988) : Using the ADAP learning algorithm to forecast the onset of diabetes mellitus, *Proceedings of the Symposium on Computer Applications in Medical Care*, pp.261-265
- World health organization(1988) : Report of a study group : Diabetes Mellitus. World health organization technical report series, Geneva, 727