

논문 2014-51-3-16

# 개인 식품섭취 선호도에 따른 만성질환 발생 위험도 분석 시스템 설계 및 구현

(Design and Implementation of Chronic Disease Risk Analysis System  
according to Personalized Food Intake Preferences)

전 소 혜\*, 김 남 현\*\*

(So Hye Jeon and Nam Hyun Kim<sup>©</sup>)

## 요 약

IT 기술의 발달로 인터넷에 다양한 콘텐츠가 늘고, 다양한 정보에서 개인의 적합한 정보를 제공받고자 하는 요구가 급증하면서 다양한 개인화 서비스를 제공 방법에 관한 연구가 활발히 진행되고 있다. 본 논문에서는 구매 선호도에 따른 상품 추천 시스템의 분석방법을 개인 식품섭취 선호도에 따른 고혈압 및 당뇨 발생위험도를 분석하는 시스템에 적용하는 방법을 제안하고자 한다. 개인의 식품섭취 선호도 분석을 위해, 피어슨 상관계수를 이용하여 참조데이터와 샘플데이터의 유사도 가중치를 계산하고, 개인과의 유사도에 따른 집단을 구성하여 고혈압과 당뇨의 발생 위험도를 산출한다. 시스템의 유의성 검정을 위해 1,021명의 샘플을 시스템에 적용하였다. 고혈압과 당뇨병의 유병군에서 정상인군에 비해 더 높은 발생 위험도가 산출되는 통계적으로 유의한 경향을 확인할 수 있어 식품섭취 선호도와 고혈압/ 당뇨의 발생 위험도의 관련성이 있음을 확인하였다. 본 연구에서는 개인의 식품섭취 선호도에 따른 고혈압 및 당뇨 발생 위험도 분석 시스템의 유효성을 검증하였다.

## Abstract

While variety of content on the internet has increased with the development of IT and person's needs about suitable information are increasing rapidly, studies for personalized service have been actively performed. In the study, we proposed the Hypertension and Diabetes risk analysis system according to personal food intake preference using the analysis method of buying preferences in product recommendation system. For the analysis of food intake preference, the Pearson correlation coefficient is used to calculate similarity weights between each reference analysis data and sample data and then reference data should be grouping into the similarity weights and calculating risk of hypertension and diabetes each group. To evaluate the significance of this system, 1,021 subjects are applied the system. Hypertension and diabetes groups' risk is significant higher than normal group statistically so, it is confirmed that food intake preference and the diseases were relevant. In this paper, we verify the validity of hypertension and diabetes risk analysis system using a personal food intake preference.

**Keywords** : Hypertension, Diabetes, Food intak Preference, Risk Analysis

\* 정회원, 연세대학교 대학원 생체공학협동과정  
(Graduate Program of Biomedical Engineering,  
Graduate School, Yonsei University)

\*\* 정회원, 연세대학교 의과대학 의공학학교실  
(Department of Medical Enginnering, College of  
Medicine, Yonsei University)

© Corresponding Author(E-mail: knh@yuhs.ac)  
접수일자: 2013년12월3일, 수정완료일: 2014년3월5일

## I. 서 론

IT기술의 발전을 통해 인터넷 서비스는 다양해지고  
사용자의 관점에 초점을 둔 정보를 제공하는 서비스에  
많은 관심이 모아지고 있다. 수많은 콘텐츠와 서비스의  
범람으로 사용자가 원하는 콘텐츠를 원하는 형태로 전

달받는 것이 중요한 기능으로 여겨지고 있다. 또한 많은 양의 데이터 축적으로 데이터로부터 정보를 추출할 수 있는 다양한 알고리즘에 관한 연구가 이루어지면서 웹 마켓이나 포털 등에서 사용자에게 적합한 정보를 제공하기 위한 개인화 관련 연구가 진행되어왔다<sup>[1]</sup>. 건강관리를 위한 서비스에서도 사용자에게 적합한 개인별 추천시스템이나 및 개인별 식단 추천 시스템<sup>[1~4]</sup> 등이 연구되고 있다.

이처럼 여러 분야에서 개인별 맞춤형 정보의 제공에 관한 다양한 접근 방법이 이루어지고 있지만, 국가 통계자료를 이용한 개인별 성향 분석에 대한 연구는 거의 없다. 국가통계자료 중 국민건강 영양조사 자료의 활용 방법은 역학조사를 통해 얻어진 정보를 특정 집단과 성별, 연령에 따른 질병에 관한 일반적인 정보를 제공하는 것에 그치고 있다. 이러한 일반화된 건강정보는 건강에 대한 인식 개선은 가능하지만 각 개인의 성향을 분석한 것이 아니므로 실질적인 영향을 미치지 못한다. 본 연구에서는 한국인을 대상으로 조사되어온 국민건강 영양조사 자료를 기반으로 개인의 데이터를 분석하여 개인의 식품섭취 선호도가 고려된 분석을 통해 고혈압과 당뇨의 위험도 평가 시스템을 구현하고자 하였다.

## II. 관련 기술

### 1. 개인화(Personalization) 서비스

개인화는 개인의 요구를 충족시킬 수 있도록 맞춤형된 제품/서비스/정보를 제공하는 것을 의미한다<sup>[6]</sup>. 모든 개인에게 같은 정보를 제공하는 것이 아닌 개인에게 특별한 제품과 서비스를 제공하는 서비스로 마케팅 분야에서 판매증대를 위한 방안이 되고 있다. 개인화 방법으로 사용자가 접근한 로그를 분석하여 개인의 성향을 파악하거나, 구매내역 등을 통한 선호도 기반의 개인 성향을 파악하는 웹 마케팅이 일반화 되고 있다<sup>[7]</sup>. 하지만 개인의 식품 섭취의 선호도 분석을 통한 개인화 서비스에 대한 접근은 미비하다. 국민건강영양조사 지침서<sup>[5]</sup>에서 식생활이 매년 크게 변화하지 않는다는 전제를 고려하여 연구 분석 시 1년간의 섭취양상을 기반으로 조사하고, 건강에 대한 식생활의 영향은 장기간 섭취를 통해 영향을 미치는 것으로 가정한다. 이를 근거로 개인의 식품섭취선호도는 섭취 양상을 내포한 1년간의 식품섭취빈도를 바탕으로 개인의 유사도 가중치를

산정하여 식품섭취 유사성에 따른 집단을 구성하고 집단의 고혈압과 당뇨의 발생 위험도를 평가하여 제시함으로써 통계적 근거를 기반으로 한 정보를 제공하는 시스템을 설계하였다.

### 2. 개인 질병발생 위험도 평가 모델

질병발생 위험도는 특정 조건의 샘플 집단의 유병여부에 대한 분석을 통해 집단별 질병발생 위험도 값을 승산비(Odds Ratio)로 분석하여 산출할 수 있다. 일반적으로 질병발생 위험도에 대한 분석은 역학조사의 일부로 산출된 결과를 보건 통계 분야에서 활용하고 있다. 이러한 분석 방법을 식품섭취 패턴에 기반하여 유사한 집단을 구성하여 해당 집단의 질병발생 위험도 평가를 하는 방법으로 식품섭취 유사도 가중치를 통한 당뇨병발생 위험도를 평가하는 연구가 진행되었다<sup>[8]</sup>.

사전 연구에서 당뇨병 유병 대상자의 식품섭취빈도를 입력하여 유사도를 계산하고, 유사도가 높은 집단의 당뇨병 유병율을 계산한 결과 당뇨병 유병 대상자와 유사한 식품섭취빈도를 갖는 집단에서 더 높은 유병율이 산출되는 결과를 확인할 수 있었다<sup>[9]</sup>. 이러한 분석방법으로 개인의 식품섭취 선호도에 따른 집단을 구성하고 구성된 집단의 당뇨병 발생 위험도를 분석할 수 있는 모델 연구가 진행되었다<sup>[10]</sup>. 본 논문에서는 기존의 연구

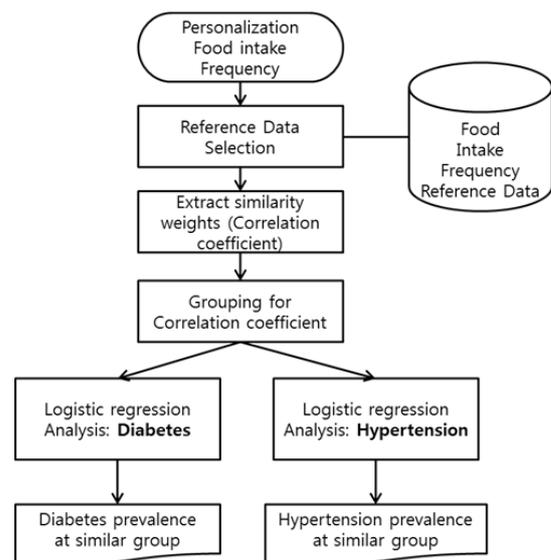


그림 1. 식품섭취 선호도에 따른 당뇨 및 고혈압 발생 위험도 평가 모델 프로세스

Fig. 1. Process of Diabetes and Hypertension Risk Analysis Model with Personalized Food Intake Preference.

를 통해 식품섭취 선호도에 따른 당뇨병 발생 위험도 평가 모델을 응용하여 고혈압과 당뇨병 발생 위험도를 계산할 수 있도록 프로세스를 수정하였다. 수정된 분석 모델의 프로세스는 그림 1과 같다.

입력된 개인의 식품섭취빈도는 분석에 적용할 참조 데이터를 선택하여, 선택된 참조데이터 전체 샘플과 개인의 식품섭취빈도의 유사도를 계산한다. 참조 데이터 전체 샘플과 계산된 유사도에 따라 참조 데이터를 사분 위수로 나누어 식품섭취빈도를 입력한 개인과 가장 유사한 집단 샘플들의 당뇨 및 고혈압 발생 위험률을 산출한다. 산출된 결과는 개인과 비슷한 식품섭취빈도를 가진 대상자들의 당뇨 및 고혈압 유병여부이므로, 개인의 식품섭취와 질병과의 관련성을 파악할 수 있다.

### III. 식품섭취빈도에 따른 개인 질병발생 위험도 평가 시스템

#### 1. 시스템 아키텍처

제안된 시스템은 분석에 사용되는 국가 연구조사인 국민건강영양조사를 통해 수집된 최신의 데이터를 업데이트 할 수 있는 모듈, 개인의 식품섭취빈도를 입력할 수 있는 자료입력 모듈, 입력된 자료를 유사도 분석 및 승산비 산출 알고리즘을 이용한 개인 질환 발생 위험도를 예측하는 모듈, 데이터를 저장하는 데이터베이스로 구성되며 산출된 결과를 사용자에게 보여주는 인터페이스 모듈을 포함한다. 본 시스템은 클라이언트/서버 환경의 웹서비스로 JSP(Java Server Page)를 사용하여 구현하였다. 데이터베이스는 MySQL 5.5.32를 사용하였고 웹서버 Tomcat 6.0을 사용하였다. 본 연구에서 제안하는 개인 질병발생 위험도 평가 시스템의 구성도는 다음과 같다.

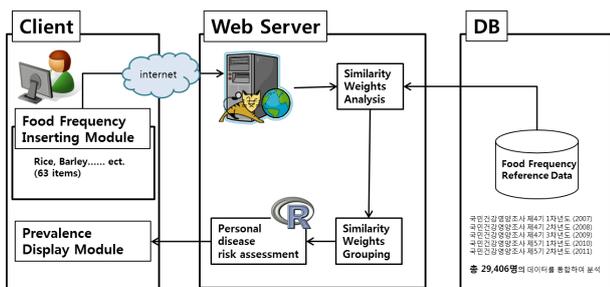


그림 2. 시스템 아키텍처  
Fig. 2. System Architecture.

#### 2. 식품섭취빈도 참조데이터 데이터베이스

본 시스템의 참조 데이터는 국민건강 영양조사 원시 자료로부터 수집된 자료를 기반으로 식품섭취빈도와 고혈압, 당뇨병의 유병여부 값을 추출하여 데이터베이스에 업로드 한다. 분석을 위해 업로드 한 참조데이터는 국민건강영양조사 2007~2011년도 자료를 통합하여 사전 테스트를 위한 1,021명의 샘플을 제외한 29,406명을 참조데이터로 구성하였다. 다음의 화면은 참조데이터를 업로드하는 페이지로 분석에 활용될 참조데이터를 추가 업로드 할 수 있는 유연한 구조로 설계하였다. 국민건강영양조사 데이터 추가 발생 시 추가하여 분석에 이용할 수 있다.

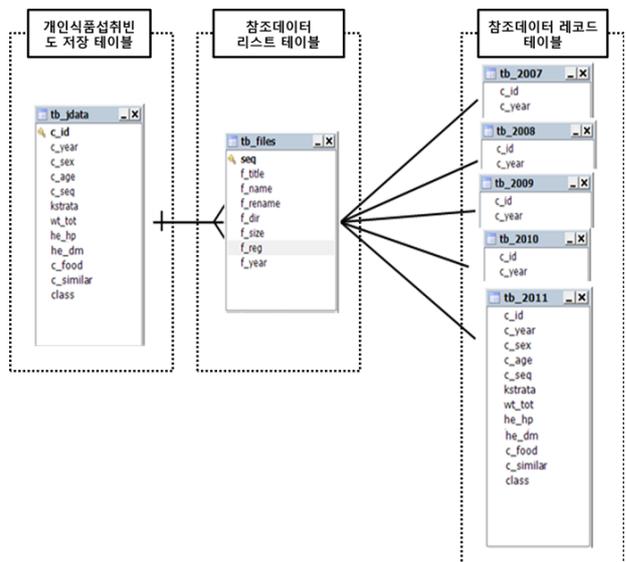


그림 3. 시스템 데이터베이스 ERD (Entity Relationship Diagrams)

Fig. 3. ERD(Entity Relationship Diagrams) of database in the System.

#### 3. 개인별 식품섭취빈도 분석

분석 모듈에서 평가하고자 하는 개인의 식품섭취빈도를 입력받아 업로드 되어 있는 참조데이터와 유사도가중치 계산을 통해 상관관계를 추출하고, 상관관계수에 따른 유사도 그룹을 산정한다. 그룹별 유병여부를 분석하여 평가하고자 하는 개인과 유사도가 높은 그룹의 승산비와 95% 신뢰구간 결과를 제공한다.

식품섭취빈도의 유사성은 상관관계 기반의 알고리즘으로 산출하여 개인과 사용자들의 식품섭취빈도에 대한 공통 선호도 값을 통해 유사도 평가하여 유사 집단을

구성한다. 데이터간의 유사도를 판별하기 위해 사용자와 유사한 식품섭취빈도를 갖는 이웃을 유사도 가중치로 분석한다. 본 논문에서는 유사도 가중치를 계산하기 위하여 피어슨 상관계수를 이용하였다. 피어슨 상관계수의 공식은 수식 (1)과 같다<sup>[10~11]</sup>.

$$sim(U_k, J_k) = \frac{\sum_{k=1}^n (U_k - \bar{U})(J_k - \bar{J})}{\sqrt{\sum_{k=1}^n (U_k - \bar{U})^2 \sum_{k=1}^n (J_k - \bar{J})^2}} \quad (1)$$

유사도 집단 그룹화는 순위에 따른 사분위수로 그룹화하여 4개의 그룹간의 질병발생위험도를 평가할 수 있다. 피어슨 상관계수의 유사도 가중치 계산은 오차와 표준편차를 이용하게 되어 표본 데이터의 누락이 있는 경우 예측값이 부정확해지는 단점이 있으나<sup>[11]</sup>, 식품섭취빈도 조사자료 특성상 충분한 데이터가 확보되고 희박성 문제에 노출될 가능성이 적어 식품섭취빈도 유사도 가중치 연구에 유리한 이론으로 본 연구에 개인의 식품섭취빈도와 유사도 가중치를 분석하는 방법으로 적용하였다.

개인의 식품섭취빈도 입력하여 개인별 분석이 가능하도록 하는 데이터 입력 화면이다. 개인 식품섭취빈도 입력 및 함께 분석하고자 하는 참조 데이터를 확인하여 사용자가 선택하여 분석할 수 있다. 과거 모든 데이터와 비교도 가능하고, 최근 통계자료만을 기반으로도 분석할 수 있다. 개인의 식품섭취빈도는 국민건강영양조사에서 사용한 63가지 식품의 1주일 단위의 섭취 횟수로 입력할 수 있도록 하였다.

그림 4. 개인 식품섭취빈도 입력  
Fig. 4. Insertion of personal food intake frequency.

4. 당뇨 및 고혈압 발생 위험도 결과 화면

결과 데이터는 로지스틱 회귀분석을 통해 산출된 승산비(Odds Ratio)와 신뢰구간(95% CI: 95% Confidence interval)을 최종 변수로 출력한다. 출력된 데이터 결과를 화면에서 확인할 수 있다. 첫 번째 테이블에서 유사도 그룹별 고혈압 발생 위험도를 나타내고, 하단 두 개의 테이블에서 고혈압 발생 위험도 수치에 대한 신뢰구

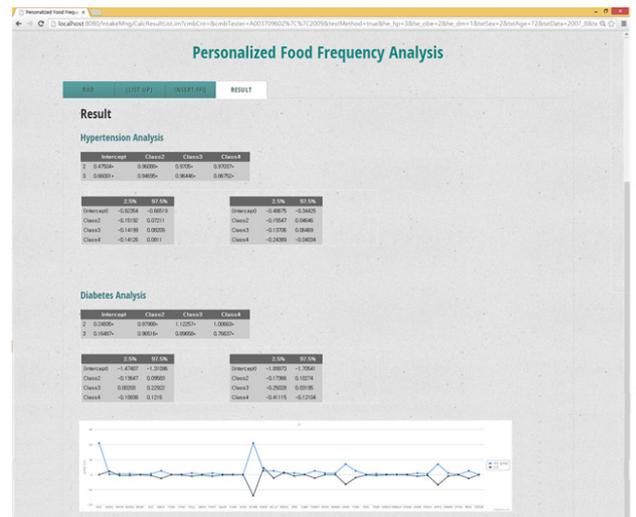


그림 5. 당뇨병 및 고혈압 발생 위험도 결과  
Fig. 5. Result of risk analysis for diabetes and hypertension.

표 1. 시스템 분석 결과 출력 파라미터  
Table 1. Output parameter of analyzing result.

출력 파라미터	설명
Class1 OR	REF.
Class2 OR	Class2 승산비
Class3 OR	Class3 승산비
Class4 OR	Class4 승산비
Class2 95% CI Low	Class2 95% CI 최소값
Class2 95% CI High	Class2 95% CI 최고값
Class3 95% CI Low	Class3 95% CI 최소값
Class3 95% CI High	Class3 95% CI 최고값
Class4 95% CI Low	Class4 95% CI 최소값
Class4 95% CI High	Class4 95% CI 최고값

* 참고	Class1 유사도 사분위수 0~25% (유사도가 가장 낮은 집단)
	Class2 유사도 사분위수 26~50%
	Class3 유사도 사분위수 51~75%
	Class4 유사도 사분위수 75~100% (유사도가 가장 높은 집단)

간 값을 출력한다. 네 번째 테이블은 유사도 그룹별 당뇨병 발생 위험도이고, 당뇨병 발생 위험도에 대한 신뢰구간 값을 출력한다. 질병별 출력되는 변수는 표 1과 같은 함수들로 각 그룹별 위험도를 평가한다.

#### IV. 식품섭취빈도에 따른 개인 질병발생 위험도 평가 시스템 유의성 분석

##### 1. 실험데이터 및 실험 방법

유의성 분석을 위한 실험데이터는 국민건강영양조사 제 4기와 제5기(2007-2011) 자료에서 참조데이터 구성 전에 미리 샘플링 한 값으로, 연도별 약 200 샘플을 추출하여 1,021명의 샘플데이터를 집합으로 사용하였다. 1,021명의 샘플데이터를 개인 식품섭취빈도 분석 프로세스에 적용하여 29,406명의 데이터와 유사도 가중치를 산출하여 29,406명과 계산된 유사도가중치 크기에 따라 사분위수로 나누어 4개 집단의 승산비를 계산하였다. 계산된 1,021명의 실제 질병 진단 내역과 비교하여 유병 집단에서의 승산비 평균과 실제 정상인 집단의 승산비 평균을 확인하고자 하였다.

표 2. 기술통계  
Table 2. General characteristics of samples.

	N	AGE		
		Mean	SD	
2007	201	48	15.3	
2008	204	50	16.8	
2009	210	48.4	15.7	$\chi^2 = 0.530$
2010	204	50	16.2	$df = 4$
2011	202	50	16.6	$p = 0.971$
Total	1021	49	16.1	

##### 2. 당뇨병 발생 위험도 유의성 분석 결과

표 3에서 유사도가 가장 높은 Q4의 경우, Q1을 레퍼런스로 정상군에서 승산비가 4.79 당뇨병 군에서 13.06로 산출되었다. Q4의 당뇨병 군은 Q4의 정상인군에 비해 Q1에 비해 2.7배 높은 당뇨병 발생 위험도를 보이는 집단으로 분류되었다. 사후검정을 통해 확인한 결과 정상인군과 공복혈당장애군은 구분할 수 있을 만큼의 차

이가 발생하지 않았지만, 정상인군과 당뇨병군은 통계적인 분류가 가능할 만큼의 차이를 보이는 것으로 확인되었다. Q2와 Q3 집단에서도 Q4와 유사한 성향을 보이고 있다. Q2에서 정상군과 공복혈당장애군을 구분할 수 있는 그룹으로 분류하지 않았으나, 정상군과 당뇨병군은 구분할 수 있는 그룹으로 분류하고 당뇨병군에서 정상인군에 비해 1.6배 높은 당뇨병 발생위험도가 산출되었다. Q3 집단에서도 사후검정에 의한 그룹분류는 유사하게 되었고 정상인군과 당뇨병군의 당뇨병발생위험도의 차이가 1.3배로 차이가 있음 확인할 수 있었다. Q2, Q3에서 모두 정상인군과 당뇨병군을 구분할 수 있었으나 Q4 그룹에 비해 차이가 작게 나오므로서 Q4를 이용한 분석이 가장 유의성이 높음을 확인할 수 있었다. 그림6은 전체 샘플의 유사도 집단별 질병 현황에 따른 당뇨병 발생 위험도를 표시한 것으로 Q4에서 유의한 결과가 나타나는 것을 확인할 수 있다. Q1을 기준으로 Q2, Q3, Q4에서 도출된 당뇨병 발생 위험도의 수치의 평균은 당뇨병 유병 여부에 대해 유의수준 .000 수준에서 유의함을 검증할 수 있었다.

남녀 집단을 분류하여 분석한 결과도 전체 분석결과와 비슷하게 도출되었다. 남성의 경우, Q4에서 정상군에서 3.69의 당뇨병 발생 위험도가 산출된 반면, 당뇨병군에서 11.42으로 3.1배 높은 당뇨병 발생 위험도를 확인할 수 있었다. 사후검정결과 정상군과 공복혈당장애군은 집단을 구분할 수 있을 정도의 차이가 발생하지는 않았으나 당뇨병군은 확연히 구분할 수 있는 정도의 차이를 보

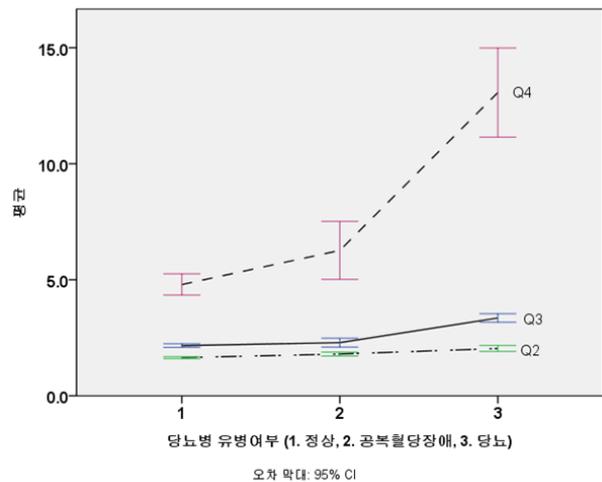


그림 6. 유병여부별 당뇨병 발생 위험도 평균  
Fig. 6. Average of Diabetes risk whether the prevalence.

표 3. 당뇨병 유무에 따른 당뇨병 발생 위험도 평균  
Table 3. Diabetes risk whether the Diabetes Morbidity.

DM class		N	Mean	±	SD	F	p	scheffe	
TOTAL	Q1		1(ref)			-	-		
	Q2	정상군 (1)	784	2.16	± 1.13	38.50**	.000	1,2<3	
		공복혈당장애(2) 당뇨병 (3)	161 76	2.29 3.35	± 1.22 ± .50				
	Q3	정상군 (1)	784	1.65	± .58	18.80**	.000	1,2<3	
		공복혈당장애(2) 당뇨병 (3)	161 76	1.80 2.03	± .55 ± .56				
	Q4	정상군 (1)	784	4.79	± 6.44	49.78**	.000	1,2<3	
		공복혈당장애(2) 당뇨병 (3)	161 76	6.26 13.06	± 8.02 ± 8.41				
			1021	5.64	± 7.27				
	MALE	Q1		1(ref)			-	-	
		Q2	정상군 (1)	294	1.97	± 1.07	21.27**	.000	1,2<3
			공복혈당장애(2) 당뇨병 (3)	76 33	2.26 3.24	± 1.19 ± .75			
		Q3	정상군 (1)	294	1.74	± .64	9.56**	.000	1,2<3
공복혈당장애(2) 당뇨병 (3)			76 33	1.93 2.21	± .64 ± .66				
Q4		정상군 (1)	294	3.68	± 5.83	22.17**	.000	1,2<3	
		공복혈당장애(2) 당뇨병 (3)	76 33	5.95 11.42	± 8.09 ± 8.52				
			403	4.74	± 6.90				
FEMALE		Q1		1(ref)			-	-	
		Q2	정상군 (1)	490	2.27	± 1.15	20.18**	.000	1,2<3
			공복혈당장애(2) 당뇨병 (3)	85 43	2.31 3.43	± 1.25 ± .83			
		Q3	정상군 (1)	490	1.58	± .54	7.79**	.000	1,2<3
	공복혈당장애(2) 당뇨병 (3)		85 43	1.67 1.90	± .43 ± .42				
	Q4	정상군 (1)	490	5.46	± 6.86	30.70**	.000	1,2<3	
		공복혈당장애(2) 당뇨병 (3)	85 43	6.54 14.33	± 8.00 ± 8.20				
			618	6.22	± 7.46				

였다. Q2와 Q3에서도 유사한 구분이 가능한 분석이 이루어 졌으나, Q4에 의한 분석이 가장 유의함을 확인할 수 있었다. 여성의 경우, Q4에서 정상군은 5.46, 당뇨병군은 14.33으로 도출되어 당뇨병군에서 확연하게 큰 당뇨병발생위험도가 산출되었다. 사후검정 결과에서 정상인군과 공복혈당장애군 그룹은 구분할 수 없으나, 정상인군과 당뇨병군을 구분할 수 있었다.

3. 고혈압 발생 위험도 유의성 분석 결과

표 4에서 Q4는 유사도가 가장 높은 집단으로 유사도

가 가장 낮은 Q1을 레퍼런스로 분석한 결과 정상군에서의 승산비가 2.719, 고혈압 전단계군에서 4.074, 고혈압군에서 5.406으로 산출되었다. Q4의 고혈압 군은 Q4의 정상인군에 비해 Q1에 비해 2배 높은 고혈압 발생 위험도를 보이는 집단으로 분류되었다. 사후검정을 통해 확인한 결과 정상인군과 고혈압전단계군, 고혈압군을 독립 집단으로 구분할 수 있다. Q2와 Q3에서 역시 정상인군, 고혈압전단계군, 고혈압군을 구분할 수 있을 만큼의 차이가 보였다. Q2에서는 고혈압군이 정상인군에 비해 1.6 배 높은 고혈압 발생 위험도가 산출되었고, Q3에서 고혈

표 4. 고혈압 유무에 따른 고혈압 발생 위험도  
Table 4. Hypertension risk whether the Hypertension Morbidity.

HP class		N	Mean	±	SD	F	p	scheffe
TOTAL	Q1		1(ref)		-	-		
	Q2	정상군 (1)	521	.86	± .66	38.16**	.000	1<2<3
		고혈압전단계(2)	262	1.07	± .76			
		고혈압 (3)	238	1.33	± .71			
	Q3	정상군 (1)	521	1.39	± .57	50.91**	.000	1<2<3
		고혈압전단계(2)	262	1.64	± .61			
		고혈압 (3)	238	1.83	± .51			
	Q4	정상군 (1)	521	2.71	± 3.05	43.68**	.000	1<2<3
		고혈압전단계(2)	262	4.07	± 4.24			
		고혈압 (3)	238	5.40	± 4.49			
			1021	3.69	± 3.90			
	MALE	Q1		1(ref)		-	-	
Q2		정상군 (1)	160	.76	± .63	23.04**	.000	1,2<3
		고혈압전단계(2)	135	.90	± .74			
		고혈압 (3)	108	1.34	± .72			
Q3		정상군 (1)	160	1.48	± .61	19.10**	.000	1<2<3
		고혈압전단계(2)	135	1.66	± .66			
		고혈압 (3)	108	1.96	± .59			
Q4		정상군 (1)	160	2.29	± 2.70	25.00**	.000	1,2<3
		고혈압전단계(2)	135	3.18	± 3.90			
		고혈압 (3)	108	5.61	± 4.95			
			403	3.48	± 4.04			
FEMALE		Q1		1(ref)		-	-	
	Q2	정상군 (1)	361	.90	± .67	23.94**	.000	1<2,3
		고혈압전단계(2)	127	1.25	± .75			
		고혈압 (3)	130	1.33	± .71			
	Q3	정상군 (1)	361	1.36	± .56	28.01**	.000	1<2,3
		고혈압전단계(2)	127	1.63	± .56			
		고혈압 (3)	130	1.73	± .41			
	Q4	정상군 (1)	361	2.90	± 3.18	27.65**	.000	1<2,3
		고혈압전단계(2)	127	5.01	± 4.40			
		고혈압 (3)	130	5.23	± 4.08			
			618	3.83	± 3.81			

압군이 정상인군에 비해 1.3배 높은 고혈압 발생 위험도가 평가되었다. 전체 샘플 분석 결과 Q3 그룹에서 집단별 독립성이 가장 컸다.

그림 7은 전체 샘플의 유사도 집단별 질병 현황에 따른 고혈압 발생 위험도를 표시한 것으로 Q4에서 고혈압 유병 대상자의 고혈압 발생 위험도 평균이 높은 것을 확인할 수 있다. Q1을 기준으로 Q2, Q3, Q4에서 도출된 고혈압 발생 위험도의 수치의 평균은 고혈압 유병 여부에 대해 유의수준 .000 수준에서 유의함을 검정할 수 있었다.

남녀 집단을 분류하여 분석한 결과 남성의 경우 Q4에서 정상군에서 2.291의 고혈압 발생 위험도가 산출된 반면, 고혈압군에서 5.613으로 2.5배 높은 고혈압 발생 위험도를 확인할 수 있었다. 사후검정결과 Q4에서 정상군과 고혈압 전단계군은 집단을 구분할 수 있을 정도의 차이가 발생하지는 않았으나 고혈압군은 확연히 구분할 수 있는 정도의 차이를 보였다. Q2에서의 사후검정 결과 역시 Q4와 유사하게 도출되었으나 Q3에서는 정상인군, 고혈압전단계군, 고혈압군의 집단 구별이 가능한 차이를 보임을 확인하였다. 여성의 경우, Q4에서 정상군

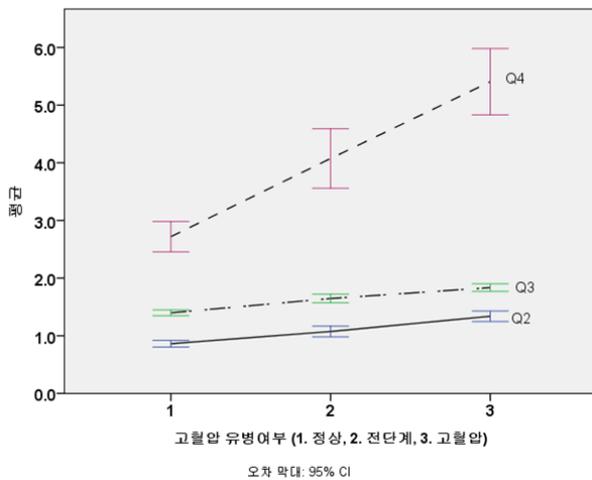


그림 7. 유병여부별 고혈압 발생 위험도 평균  
Fig. 7. Average of Hypertension risk whether the prevalence.

은 2.908 고혈압군은 5.234으로 도출되어 고혈압군에서 고혈압 발생 위험도가 1.8배 높게 산출되었다. Q2, Q3에서도 역시 정상인군에서보다 고혈압군에서 더 높은 고혈압발생위험도를 산출하여 각 그룹별 산정된 고혈압 발생 위험도 값이 실제 고혈압 유병 대상자에서 더 높은 값이 나옴을 확인할 수 있다. 사후검정 결과에서 고혈압전단계군과 고혈압군을 통계적으로 유의하게 구분할 수는 없었으나, 정상인군과 고혈압군을 구분할 수 있었다.

## V. 결론 및 고찰

본 연구에서는 개인의 식품섭취 선호도에 따른 당뇨병과 고혈압 발생 위험도를 분석하는 시스템을 제안하였다. 국민건강영양조사 자료로부터 사전 샘플링한 1,021명의 데이터를 이용하여 유의성 검정을 할 샘플을 구성하였다. 검정 샘플링에서 제외된 29,406명의 데이터의 식품섭취빈도를 기반으로 분석하고자 하는 개인과의 상관관계를 계산하여 유사도 그룹을 사분위수로 구분하고, 그룹별 고혈압과 당뇨병 발생 위험도를 산출하게 된다. 유의성 분석 결과 고혈압과 당뇨병의 실제 유병군에서 정상인군에 비해 더 높은 발생 위험도가 산출되는 통계적으로 유의한 경향을 확인할 수 있어 식품섭취 선호도와 고혈압/ 당뇨병의 발생 위험도의 관련성이 있음을 확인하였다. 제안한 시스템은 기존의 역학조사 방법에서 특정 연령군이나 특정 지역에 따른 질병 발생 위

험도를 산출하는 방식을 착안하여 개인의 식품섭취 성향에 따른 고혈압과 당뇨병 발생 위험도를 산출함으로써 개인화된 건강 서비스를 제공하기에 새로운 방법으로 사용될 수 있다. 그러나 분석 사용된 참조데이터의 경우, 조사 시점이 한정되어있는 단층연구로 추적에 의한 자료가 아니라는 한계가 있다. 다년간의 조사 자료를 기반으로 참조분석 데이터를 구성하고, 유의성 분석에 활용한 데이터 역시 5개 연도에서 각각 샘플링하여 분석함으로써 시간적 변화의 영향을 감쇄시키고자 하였다. 향후 본 연구의 결과를 토대로 인구통계학적 정보를 추가 분석하고, 특정 질병에 영향을 미치는 식품의 연구 결과를 추가 하여 가중치를 적용하는 연구가 필요하다. 이러한 추가 연구를 통해 개인의 섭취 식품과 만성질환의 관련성 분석을 통해 개인화된 질환 관리 정보로 활용을 할 수 있을 것으로 기대된다.

## REFERENCES

- [1] C. H. Lee, S. Y. Lee and T. C. Chung, "Application recommender system based on personalized collaborative-filtering using user's emotion information from smartphone," KOREA INFORMATION SCIENCE SOCIETY, Vol.39 No.1A pp. 224-226, 2012.
- [2] B. K. Lim and B. T. Zhang, "Personalized Menu Recommendation Algorithm using Hypernetwork," KOREA INFORMATION SCIENCE SOCIETY, Vol.39 No.1B pp. 393-395, 2012.
- [3] D. H. Kim, S. H. Kim, K. Y. Cho and J. S. Cho, "u-DailyCare : Design and Implementation of prevention and Management Service for chronic and Adult disease," KOREA INFORMATION SCIENCE SOCIETY, Vol.39 No.1A pp. 197-199, 2012.
- [4] J. H. Park, W. I. Park, Y. K. Kim and J. H. Kang, "A Customized Device Recommender System based on Context-Aware in Ubiquitous Environments," The Institute of Electronics Engineers of Korea - Computer and Information, Vol.46 No.3, pp. 15-23, 2009.
- [5] Korea Centers For Disease Control & Prevention, "Guideline: The Fifth Korea National Health and Nutrition Examination Survey (KNHANES V-2)", 2012.
- [6] R. R. Dholakia, interactivity and revisits to

- website, 2000, <http://ritim.cbn.uri.edu/wp/>
- [7] J. K. Kim, J. H. Suh, D. H. Ann and Y. H. Cho, "A Personalized Recommendation Methodology based on Collaborative Filtering," INTELLIGENT SOCIETY, vol. 8, pp. 139-157, 2002.
- [8] R. M. van Dam, E. B. Rimm, W. C. Willett, M. J. Stampfer, and F. B. Hu, "Dietary Patterns and Risk for Type 2 Diabetes Mellitus in U.S. Men," Annals of Internal Medicine, vol. 136, no. 3, pp. 201-209, 2002.
- [9] S. H. Jeon and N. H. Kim, "Prevalence of Diabetes and Impaired fasting glucose according to Food frequency Similarity in Korea," Journal of the Korea Academia-Industrial cooperation Society, vol. 14, pp. 751-758, 2013.
- [10] S. H. Jeon and N. H. Kim, "Diabetes Risk Analysis Model with Personalized Food Intake Preference," Journal of the Korea Academia-Industrial cooperation Society, Vol. 14, No. 11, pp. 5771-5777, 2013.
- [11] K. Y. Jung and J. H. Lee, "Comparative Evaluation of User Similarity Weight for Improving Prediction Accuracy in Personalized Recommender System," The Institute of Electronics Engineers of Korea - Computer and Information, Vol.42 No.6, pp. 63-74, 2005.

저 자 소 개



전 소 혜(정회원)  
2008년 연세대학교 의용전자  
공학과 학사  
2014년 연세대학교 대학원  
생체공학협동과정 박사  
<주관심분야: 의료정보시스템,  
PHR, 유헬스, 데이터마이닝>



김 남 현(정회원)-교신저자  
1977년 연세대학교 전기공학과  
학사  
1982년 연세대학교 대학원  
전기공학과 석사  
1987년 연세대학교 대학원  
전기공학과 박사  
1988년~현재 연세대학교 의과대학 의학공학교실  
교수  
1995년~1999 연세대학교 의과대학 의학공학교실  
주임교수  
2008년~2012년 연세의료원 의료정보실장  
<주관심분야: 의료정보시스템 개발 및 표준화, 유  
헬스, PHR 등>