

<https://doi.org/10.7236/JIIBC.2023.23.6.199>  
JIIBC 2023-6-29

# 당뇨병 환자의 라이프로그 데이터를 이용한 식단 추천 시스템

## Diet Recommendation System using Life Log Data of Diabetic Patients

김선아\*, 황만수\*\*, 김능회\*\*\*

Seonah Kim\*, Mansoo Hwang\*\*, Neunghoe Kim\*\*\*

**요약** 국민건강보험공단에 따르면 불규칙한 식습관, 스트레스, 운동 부족 등의 요인으로 젊은 당뇨 환자가 2017년 대비 2021년에 3,564,059명으로 24.3% 증가 추세를 보였다. 모든 연령층에서 증가 중인 당뇨병은 약물치료, 규칙적인 운동, 식사요법 관리가 필요하다. 이 중 식사요법은 적절한 열량 및 균형 잡힌 3대 영양소 섭취를 필요로 하므로 매우 체계적인 관리가 필요하다. 현재 당뇨 식단 추천은 개인 정보, 건강 정보, 사회적, 문화적 측면을 고려하여 식단을 제안 하지만, 건강 정보에 대한 다중 변수를 고려하지 않은 부분이 미흡하다. 이에 본 논문은 당뇨병 환자의 라이프로그 데이터를 이용한 식단 추천 시스템을 제안하여 데이터 다중 변수를 고려해 개인의 건강 상태에 따른 맞춤형 식단을 추천하고자 한다.

**Abstract** The National Health Insurance Corporation reported a 24.3% increase in young diabetes patients, rising to 3,564,059 in 2021 from 2017, which is attributed to factors like irregular eating patterns, heightened stress, and insufficient physical activity. Diabetes, which is increasing in all age groups, requires medication, regular exercise, and dietary management. Of these aspects, dietary therapy demands systematic management as it involves ensuring sufficient calorie intake and a balanced consumption of the three major nutrients. The current diabetes diet recommendations consider personal, health, social, and cultural factors, yet they fall short of addressing various health variables comprehensively. Therefore, this paper proposes a diet recommendation system using life log data from diabetic patients, which recommends customized dietary suggestions according to the individual's health status by considering multiple variables in the data.

**Key Words** : Diabetes, Diet Recommendation System, Life Log Data

\*학생회원 군산대학교 소프트웨어학부

\*\*정회원 신한대학교 IT융합공학부(교신저자)

\*\*\*정회원 군산대학교 소프트웨어학부(교신저자)

접수일자 2023년 10월 13일, 수정완료 2023년 11월 15일

계재확정일자 2023년 12월 8일

Received: 13 October, 2023 / Revised: 15 November, 2023 /

Accepted: 8 December, 2023

Corresponding Author: mshwang@shinhan.ac.kr,

nunghoi@kunsan.ac.kr

School of IT Convergence Engineering, Shinhan University, Korea,  
School of Software, Kunsan National University, Korea

## I. 서 론

최근 고열량, 고지방, 고단백 식단 섭취로 인한 식생활 문제 뿐만 아니라 불규칙한 식습관, 스트레스, 운동 부족 등으로 젊은 당뇨병 환자가 증가하고 있다. 국민 건강보험공단에 따르면, 기존의 당뇨병 환자의 나이 연령대는 전체 환자 중 중장년, 노년의 비율이 높은 숫자로 대부분을 차지했지만, 현재 국내 추세에 따르면 연령을 가리지 않고 당뇨병 환자 발생이 증가하고 있다<sup>[1]</sup>.

당뇨병은 혈액 속의 포도당이 세포 속으로 들어가 에너지원으로 이용되지 못해 혈당이 비정상적으로 올라가는 질환이다. 단순히 혈당만 상승하는 병이 아닌 당뇨병은 잘 조절하지 않으면 시간이 지날수록 많은 합병증을 유발한다. 당뇨병은 제1형 당뇨병, 제2형 당뇨병, 기타 당뇨병, 임신성 당뇨병으로 나뉘는데 제1형 당뇨병은 췌장의 베타세포가 파괴되어 인슐린이 전혀 분비되지 않는 병이다. 제2형 당뇨병은 몸의 인슐린 저항성이 커지면서 인슐린의 작용이 원활하지 않고 상대적으로 인슐린 분비의 장애가 생겨 혈당이 올라가는 병이다. 기타 당뇨병은 특정한 원인에 의해 발생하고 임신성 당뇨병은 임신 중에 발견된다. 한국인 당뇨병의 대부분이 제2형 당뇨병이다.

당뇨병 치료의 목적은 합병증의 예방으로 모든 환자는 진단 초기부터 적극적인 생활 습관 개선과 적절한 약물 치료를 필요로 한다. 적극적인 생활 습관에는 규칙적인 운동과 식사요법이 포함되는데 이중 식사요법은 좋은 영양 상태를 유지하기 위한 당뇨병 환자의 중요한 치료 방법의 하나로서 환자는 이를 위해 자신이 섭취하는 식단과 그 식단에 대한 영양 정보를 파악해야 한다<sup>[2]</sup>. 따라서 당뇨를 관리하는데 체계적인 식단 섭취는 당뇨병 환자 뿐만 아니라 현대 사회의 모든 개인에게 중요한 역할을 한다는 것을 알 수 있다.

당뇨병 환자는 자신의 건강 상태에 따라 매일 규칙적인 식사를 통해 적절한 열량 및 3대 영양소를 균형 있게 섭취해야 한다. 이처럼 개인별 맞춤 관리가 필요한 당뇨병 환자는 자신의 건강 상태에 따라 어떠한 영양소를 균형 있게 섭취해야 하는지 알 수 있는 식단 추천이 필요하다.

본 논문은 환자의 현재 건강 상태 여부를 파악하여 이에 따라 맞춤 관리를 할 수 있는 식단 추천 시스템을 제안한다. 당뇨병 환자의 건강을 위해 관리해야 하는 지표를 설정해 건강점수를 도출하고 나이, 섭취 열량, 영양소를 고려 식단을 구성한다. 따라서 일반적으로 제공되는 식단이 아닌 본인의 상태에 맞는 식단을 제공받는다.

## II. 배경

### 1. 당뇨인의 식단 구성

당뇨병 환자는 일정한 시간에 알맞은 양의 음식을 규칙적으로 섭취해야 한다. 이는 단순한 식사를 의미하는 것이 아닌 치료를 목적으로 하는 방법 중 하나이다. 식사요법의 목표로는 올바른 식습관과 생활 습관으로의 변화를 통해 적절한 혈당, 혈압, 지질 농도를 유지하고, 급성 및 만성 합병증을 예방하는 것이다.

혈당 상승을 유발하는 성분이 들어가는 음식은 피해야 하며 예로는 설탕이나 꿀 등의 단순당, 동물성 단백질에 함유된 포화지방산과 콜레스테롤, 염분, 음주 등이 있다. 혈당 조절과 지방의 농도를 조절할 수 있는 식이 섬유소 등의 당뇨 조절에 도움이 되는 성분이 들어간 음식을 섭취해야 한다<sup>[3]</sup>.

또한 대한 당뇨병 학회에 따르면 당뇨인은 식품 교환표를 이용하여 식단을 구성한다. 식품 교환표란 식품들을 영양소 구성이 비슷한 것끼리 6가지 식품군으로 나누어 묶은 표이다. 6가지 식품군은 곡류군, 어육류군, 채소군, 지방군, 우유군, 과일군을 말하며, 같은 군 내에서는 자유롭게 바꿔 먹을 수 있도록 설정되어 있다. 당뇨인의 하루 섭취 열량에 맞춰 식품군별 교환 단위를 섭취해야 다양한 영양소의 필요량을 충족시킬 수 있다<sup>[4]</sup>.

### 2. 당뇨인의 하루 섭취 열량

대한 당뇨병 학회에 따르면 당뇨병 환자가 섭취하는 음식의 열량은 당뇨인이 유지해야 할 표준체중과 일상생활에서 활동하는 정도에 따라 달라진다. 따라서 당뇨병 환자는 자신의 적절한 표준체중을 알고 이를 바탕으로 필요 열량을 계산해 섭취해야 하는 음식의 양을 결정해야 한다. 표준체중이란 일상생활에서 건강을 유지하기 위한 적절한 체중을 의미하며 이는 성별과 신장에 따라 달라진다. 표 1은 성별에 따른 표준체중을 계산하는 공식이다. 표준체중을 구하는 방법에는 여러 가지 방법이 있으나 본 논문에서는 대한 당뇨병 학회의 내용과 같이 비교적 쉽고 정확한 방법을 사용한다.

표 1. 표준체중 계산

Table 1. Standard Weight Calculation

	남자	여자
표준체중(kg)	키(m)의 제곱 × 22	키(m)의 제곱 × 21
표준체중 : 일상생활에서 건강을 유지하기 위한 적절한 체중		

환자의 하루 필요 열량을 계산하는 방법은 다음과 같다. 표 2는 육체 활동 정도에 따른 하루 필요 열량 계산 공식으로 활동 정도와 표준체중을 사용한다. 해당 공식은 활동 정도가 커지면 표준체중에 곱해지는 값이 커진다. 많이 움직일수록 열량 소모가 많아 더 많은 열량을 섭취해야 하는 것은 당연하다.

표 2. 하루 필요 열량 계산  
 Table 2. Calculation of Daily Caloric Requirements

활동 정도	하루 필요 열량(Kcal)
육체 활동이 거의 없는 경우	표준체중 × 25-30 (Kcal/일)
보통의 활동을 하는 경우	표준체중 × 30-35 (Kcal/일)
심한 육체 활동을 하는 경우	표준체중 × 35-40 (Kcal/일)

표 1과 표 2의 공식을 사용하면 당뇨병 환자가 하루 섭취해야 하는 열량이 계산된다. 이때 임신부나 수유부는 총열량 요구량이 증가하므로 위에서 계산한 필요 열량에 300~500 열량을 추가해야 한다. 또한 임신 전 체중이 표준체중의 범위에 속한 경우 임신기간 동안 약 10kg의 체중 증가가 있도록 해야 하며 어린이나 청소년의 경우 연령과 성장 정도를 고려해 결정해야 한다. 이처럼 여러 가지 변수가 존재해 고려해야 할 사항들이 있으므로 정확한 필요 열량은 임상 영양사와 상담하는 것을 권장한다<sup>[5]</sup>.

### III. 관련 연구

[6, 7]의 연구는 머신러닝 알고리즘을 사용해 당뇨병을 예측하고 감지하며, 이에 따른 개인 맞춤형 단계적 건강 관리를 위한 식단 및 운동을 사용자에게 제공한다. [8]의 연구의 경우 이미지 처리 모델을 활용해 음식을 인식하고 해당 음식의 영양 정보를 기반으로 사용자에게 맞춤형 식단을 추천한다.

[9]의 연구는 당뇨병 환자들의 식습관을 자세히 파악하고 특징화하기 위해 개인 건강 온톨로지를 개발한다. 영양 질문과 관심사를 알 수 있는 질문에 대한 답변 데이터와 섭취한 식사 데이터에 인공지능 기술을 활용하여 시간적 패턴을 분석한다. 이를 통해 개인 건강 지식 그래프를 생성하여 개인화된 건강 관리를 지원한다.

[10, 11]의 연구는 당뇨병 환자의 건강 데이터보다는

개인의 생활 방식, 음식의 사회적, 문화적, 종교적 중요성과 맛, 영양, 예산 등의 다양한 측면을 고려하여 식단을 제안한다.

위의 논문들은 개인 정보, 건강 정보, 행동 정보, 문화적, 사회적 등의 다양한 데이터를 고려하여 식단을 제안한다. 하지만 당뇨병 환자의 건강 관리에 관련 있는 중요한 데이터 즉, 심혈관질환, 당뇨 상태, 지방대사 상태에 대한 다중 데이터를 고려하는 부분이 미흡하다. 결과적으로 추천에 관한 접근 방식과 사용하는 데이터 변수 종류, 주요 지표에 차이가 존재한다. 따라서 본 논문에서는 당뇨병 관리와 밀접한 연관이 있는 심혈관, 당뇨, 지방대사에 대해 건강 상태를 보다 정확하게 파악할 수 있는 요인들을 추가하여 당뇨병 환자에게 더욱 도움이 되는 식단을 추천하고자 한다.

[12]의 연구는 머신러닝 기법을 사용하여 당뇨병을 진단하고 진단한 결과와 유형에 따라 DSR(Diet Recommendation System)을 통해 적절한 식이를 추천한다. [13]의 연구는 나이, 신장, 혈당 수준 등의 건강 데이터를 사용하여 머신러닝 알고리즘을 통해 군집화하고 기존 사용자 간의 유사도를 계산하여 적절한 영양 정보를 추천한다. [12, 13]의 연구는 식단을 추천하지만 동작 프로세스에 대한 세부적인 내용을 언급하고 있지는 않아 응용과 활용에 어려운 점이 있다.

## IV. 라이프로그 데이터를 사용한 식단 추천 시스템

그림 1은 본 논문에서 제안하는 시스템의 전체 아키텍처이다. 먼저, 제안하는 시스템의 개요를 전체적으로 간략하게 소개한 후, 추후에 더 자세한 내용을 세부적으로 소개하려 한다. 본 시스템은 당뇨병 환자의 건강 상태를 파악할 수 있는 점수를 도출하여 그에 따른 조치를 취하기 위한 식단을 추천한다. 건강점수는 심혈관질환, 당뇨 관리, 지방대사의 세 가지 지표마다 각각 계산된다. 이 세 가지 건강점수 중 가장 높게 나온 지표를 기준으로 맞춤형 식단을 추천한다.

제안하는 시스템에서 세 가지 지표별 건강점수를 계산하는 방법과 각 지표별 식단 추천을 위한 리스트 생성에 대한 상세한 설명은 이어지는 내용에서 세부적으로 다루고자 한다.

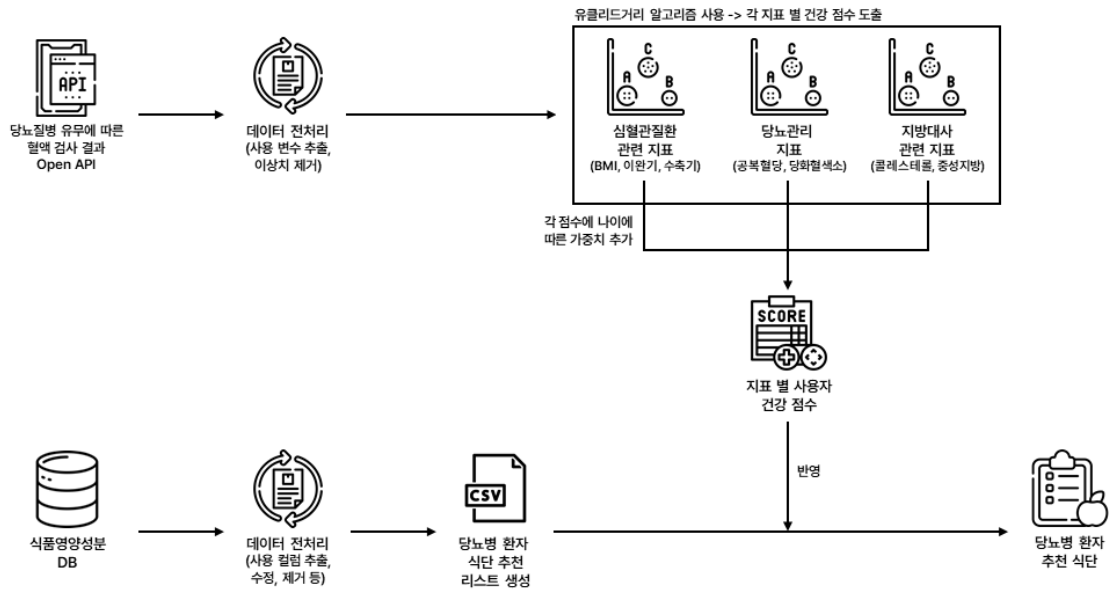


그림 1. 전체 시스템 아키텍처  
Fig. 1. Whole System Architectur

### 1. 지표별 사용자 건강점수 계산

지표별로 사용자 건강점수를 계산한다. 이 점수는 후에 당뇨병 환자의 식단을 추천할 때 반영되는 값으로 환자가 현재 어떤 식단 관리가 필요한지 알 수 있다.

#### 가. 사용 데이터 소개 및 수집

혈액은 우리 몸의 조직이 생활하는 데 필요한 영양분과 산소를 공급하는 등 여러 가지 중요한 역할을 한다. 혈액 내 존재하는 세포의 수, 모양의 변화 및 대사 물질 수치를 통해 여러 질환을 판단할 수 있어 혈액 검사의 결과로 환자의 건강 상태를 파악할 수 있다.

빅데이터 플랫폼인 LIFELOG<sup>[14]</sup>에서 제공한 혈액 검사 결과 OpenAPI를 사용한다. 이 데이터셋은 당뇨 질환 유무에 따른 성별/연령대별/비만 여부에 따른 혈액 검사 결과 제공 API로 사용자 로그 데이터를 의미한다. 사용하는 Open API는 4개의 요청 변수와 13개의 출력 변수를 가진다. 요청변수로는 gender, age, diabetes, obesity를 가지고 출력변수로는 공복혈당, 총 콜레스테롤, 중성지방, 헤모글로빈 등의 혈액 검사에 따른 결과인 사용자 로그 데이터를 가진다. 표 3은 요청 변수에 따른 혈액검사 및 혈압측정 결과인 모든 출력변수를 나타낸다.

표 3. 사용자 로그 데이터 출력변수  
Table 3. User Log Data Output Variables

출력변수	명칭	타입
FBS	공복혈당	String
TTL_CHOL	총 콜레스테롤	String
LDL	저밀도 지방단백질	String
HDL	고밀도 지단백질	String
Neutral Fat	중성지방	String
Hb	헤모글로빈	String
HbA1c	당화혈색소	String
Blood Sugar	혈당	String
Height	키	String
Weight	몸무게	String
BMI	체질량지수	String
Systolic Blood Pressure	수축기혈압	String
Diastolic Blood Pressure	이완기혈압	String

#### 나. 데이터 전처리

사용하려는 변수의 컬럼만을 추출하여 데이터를 CSV 파일로 저장한다. 사용하려는 변수는 키, 나이, BMI, 이완기혈압, 수축기혈압, 당화혈색소, 콜레스테롤, 중성지방으로 해당 컬럼의 데이터만을 추출한다. 이후, 사용자의 데이터 중 데이터 값이 -1인 이상치 데이터와 값이 튀는 값을 제거한다. 이상치 데이터란 입력 시 오류 또는 측정 시 제대로 측정되지 않은 값으로 이를 사용하여 분석했을 때 결과적으로 의미가 없는 데이터이다.

#### 다. 지표 생성 및 지표별 사용자 건강점수 계산

사용자의 건강 상태에 대한 점수를 도출하기 위한 지표를 생성한다. 최종적으로 생성된 지표는 총 세 가지이며 이는 사용한 데이터 변수들의 상관관계를 통해 도출되었다. 심혈관질환 지표, 당뇨관리 지표, 지방대사 관리 지표 이렇게 세 가지의 지표를 생성한다.

심혈관질환은 심장 자체의 원인으로 발생하기보다는 고혈압, 당뇨, 고지혈증, 비만 등의 합병증으로 발생하는 질환으로 당뇨 관련 중요 지표로 설정하였다<sup>[15]</sup>. 사용한 당뇨환자의 라이프로그 데이터 변수 중 BMI, 수축기혈압, 이완기 혈압의 변수는 심혈관질환과 관련성이 높아 해당 지표의 기준으로 설정한다. 수축기 혈압과 이완기 혈압은 값이 높은 경우 심혈관질환의 위험을 높인다. 이에 대해 이완기 혈압의 경우 13.7%, 수축기 혈압의 경우 11.98%로 심혈관질환 발병에 영향을 미친다<sup>[16]</sup>. 또한 BMI가 증가할수록 관동맥질환 발생이 높아진다. Nurse's Helath Study에 따르면 BMI가 21kg/m<sup>2</sup>를 기준으로 할 때 관동맥질환 발생 위험은 BMI가 25-28.9kg/m<sup>2</sup>일 때 2.08배, BMI가 29kg/m<sup>2</sup> 이상일 때 5.56배로 늘어난다<sup>[17]</sup>. 이를 이유로 심혈관질환 지표의 변수로 BMI, 수축기혈압, 이완기혈압을 설정한다.

당뇨관리 지표는 공복혈당, 당화혈색소 변수로 구성한다. 공복혈당은 식사 후 최소 8시간이 지난 후의 혈당 수치를 의미하며, 당화혈색소는 혈중의 글루코오스와 헤모글로빈 사이의 화학적 반응을 나타내는 지표이다. 이들은 당뇨 관리와 혈당 조절 상태를 평가하는 데 사용되며 당뇨병 진단 기준으로도 사용되어 이를 이유로 두 변수를 당뇨관리 지표 기준으로 설정한다<sup>[18]</sup>.

지방대사는 당뇨 환자의 건강 상태를 악화하는 요인 중 하나이다. 지방대사 관리 지표로는 콜레스테롤(LDL), 중성지방 변수로 설정한다. 사용할 LDL은 저밀도 지단백으로 우리 몸에 나쁜 콜레스테롤을 의미하고 중성지방은 식사 후에 혈당에서 만들어지며 지방 조직에 저장된다. 둘 다 지방 성분으로 수치가 높을수록 심혈관질환 발병 위험 가능성이 증가할 뿐만 아니라 혈당이 빠르게 올라가 위험하다. 이를 이유로 LDL과 중성지방 변수들을 지방대사 관리 지표의 기준으로 설정한다.

위에서 생성한 세 가지 지표를 사용하여 지표 별 사용자 건강점수를 도출한다. 이 점수는 당뇨인의 건강 상태를 확인할 수 있는 점수로 후에 당뇨병 환자 추천 식단에 반영하여 식단을 추천할 때 사용하는 값이다. 수치가 정상인 사람과의 비교를 통해 값은 높게 나올수록 해당 지표에 대한 건강 상태가 좋지 않음을 의미한다.

각 지표별 사용되는 변수들의 정상수치를 기준으로 하여 거리 기반 알고리즘인 유클리드 거리를 사용해 오차 값을 계산한다.

표 4. 지표별 사용 변수 기준  
 Table 4. Use Variables Criteria by Indicator

	정상	위험
BMI	18.5 ~ 22.9 (kg/m <sup>2</sup> )	
수축기혈압	140 미만 (mmHg)	140 이상 (mmHg)
이완기혈압	85 미만 (mmHg)	85 이상 (mmHg)
공복혈당	126 미만(mg/dL)	126이상(당뇨판정)
당화혈색소	4 ~ 5.7%	5.8~6.4% (당뇨 전 단계) 6.5% 이상 (당뇨 판정)
콜레스테롤 (LDL)	100 미만 (mg/dL)	100 이상 (mg/dL)
중성지방	150 미만 (mg/dL)	150 이상 (mg/dL)

변수들의 정상수치는 표 4와 같이 정의하였다. 이 중 수축기, 이완기 혈압은 심혈관질환이 동반되지 않은 고혈압을 기준으로 정의하였다<sup>[2]</sup>. 각 지표 변수의 정상 수치 값과 당뇨인의 수치 값을 유클리드 거리를 사용해 차이를 계산한다. 이후, 각 지표별 점수에 나이에 따른 가중치를 곱해 최종적으로 세 가지 지표별 건강 점수를 도출한다. 당뇨병 환자들은 일반인들에 비해 경험하는 피로 정도가 2배 이상 높고 나이가 많을수록 피로에 더 많은 영향을 받는다는 연구 결과가 있다<sup>[19,20]</sup>. 이에 따라 계산된 건강 점수가 동일한 사람이라도 나이가 많을수록 당뇨를 관리해야 하는 중요도는 높아져 나이대별로 가중치를 부과하였다. 총합을 1로 잡아 40대의 사용자인 경우 0.1을, 50대는 0.2, 60대는 0.3, 70대 이상 사용자는 0.4로 가중치를 부과하였다. 계산된 값은 환자의 건강 상태를 알려주는 점수로 심혈관질환의 상태, 당뇨 관리의 상태, 지방대사의 상태를 의미한다. 이는 현재 본인의 몸의 건강 상태와 모든 값의 수치가 정상인인 사람과의 차이를 파악할 수 있다.

#### 2. 당뇨병 환자 추천 식단 리스트 생성

당뇨병 환자에게 추천할 식단 리스트를 생성한다. 이는 위에서 추출한 세 가지 지표에 대한 관리를 위해 생성된 리스트로 당뇨인의 라이프로그 데이터를 바탕으로 사용할 수 있는 추천 식단 리스트이다.

##### 가. 사용 데이터 소개 및 수집

당뇨환자의 식단 추천을 위한 리스트를 생성하기 위해

식품의약품안전처<sup>[21]</sup>의 식품 영양성분 데이터베이스에서 제공하는 전국 통합식품 영양성분 데이터셋을 사용한다. 이 데이터셋은 식품의약품안전처에서 제공하는 대한민국 식품영양성분 통합 DB로 다양한 기관에서 생산된 영양 성분 데이터를 체계적으로 조직화하여 하나의 통합 데이터베이스화되어 있다. 농축산물, 수산물, 가공식품, 음식 항목이 수록되어 있으며 전체 식품 수는 90,105개의 데이터로 그에 따른 성분수는 농축산물, 수산물, 가공식품, 음식으로 총 348개의 데이터를 가진다.

#### 나. 데이터 전처리

사용하려는 목적에 맞게끔 연구에 필요하지 않은 컬럼을 삭제하고 필요한 데이터를 추가한다. 섭취하는 사용자가 당뇨임을 고려하여 빵류, 과자류, 아이스크림류 등의 섭취 불가한 식품을 삭제한다. 마지막으로, 식품 리스트를 당뇨인이 섭취해야 하는 식품 교환표의 6가지 식품군에 맞게 전처리한다. 6가지 식품군으로는 곡류군, 어육류군, 채소군, 지방군, 우유군, 과일군이 있다. 이중 식단으로 섭취하는 곡류군, 어육류군, 채소군, 지방군을 제외하고 우유군과 과일군은 간식으로 섭취한다. 간식도 하루 섭취 식단의 구성 중 하나이지만 본 논문에서는 고려하지 않는다. 또한 지방군에 속하는 모든 식품은 적은 양의 섭취로도 높은 열량을 내므로 따로 고려하지 않는다.

#### 다. 지표별 식단 추천 리스트 생성

위의 전처리된 데이터를 사용해 당뇨병 환자에게 추천할 식품 리스트를 생성한다. 최종적으로 생성되는 리스트는 세 가지이다. 각 지표별 추천 리스트로 심혈관질환 추천 식품 리스트, 당뇨관리 추천 식품 리스트, 지방대사 관리 추천 식품 리스트가 있다. 생성하는 과정은 다음과 같다. 식품 데이터의 영양소 변수들을 섭취했을 경우 도움이 되는 영양소와 도움이 되지 않는 영양소로 정리한 후 도움이 되는 영양소가 세 가지 중 어느 지표 관리에 도움 되는지 표기한다. 숫자로 표기하며 심혈관 질환에 도움되는 영양소가 있다면 1, 당뇨관리에 도움되는 영양소가 있다면 2, 지방대사 관리에 도움되는 영양소가 있다면 3으로 표기해 준다. 이후, 어느 지표에 도움이 되는 식품인가를 알 수 있는 숫자를 기준으로 세 가지 리스트들을 CSV 파일로 각각 생성한다. 이러한 프로세스를 통해 각 지표마다 추천 가능한 식품 리스트를 생성한다.

### 3. 당뇨병 환자 식단 추천

위의 사용자 건강 점수를 각 지표별 추천 식단 리스트에 반영하여 각자의 건강 상태에 맞게 식단을 추천한다. 사용자의 건강 점수 중 제일 높게 나온 점수를 기반으로 해당 지표에 도움이 되는 식단 리스트에서 추천하며 골고루 섭취해야 하는 6가지 식품군에 따라 랜덤으로 식품군 별 음식을 추천하여 식단을 구성한다. 이미 섭취한 식품의 경우 리스트에서 제외되며 이후 모든 식품이 리스트에서 제외되었을 경우, 다시 처음과 같이 식품들이 모두 리셋이 되는 시스템으로 구성한다. 이에 따라, 최종적인 추천은 사용자의 하루 섭취 열량과 식품 교환표에 따른 1일 치 식단이 도출된다.

## V. 실험

본 논문에서 제안하는 시스템의 이해를 돕기 위해 실제 환경 데이터를 활용하여 시스템의 각 단계별로 설명한다.

### 1. 지표별 사용자 건강점수 계산

#### 가. 라이프로그 데이터 수집 및 전처리

먼저, 당뇨인의 라이프로그 OpenAPI 데이터를 받아와 CSV 파일로 저장한다. 그림 2는 저장한 데이터를 Python의 pandas를 사용해 데이터프레임 형식으로 불러온 모든 사용자 로그 데이터이다.

성별	키	나이	BMI	이완기혈압	수축기혈압	혈당수치	공복혈당	당화혈색소	콜레스테롤	중성지방	
0	1	170.1	47	21.50	90	130	-1	112	-1.0	81	113
1	1	182.2	46	27.23	90	130	268	140	7.7	39	654
2	1	176.2	52	25.03	100	150	258	108	6.2	142	198
3	1	178.7	68	19.29	110	160	401	116	6.0	52	124
4	1	164.2	62	26.44	98	175	-1	196	8.4	120	205
5	1	157.8	63	26.38	90	130	353	154	7.9	149	164
6	1	166.1	79	-1.00	80	130	-1	-1	-1.0	-1	-1
7	1	163.0	76	25.50	70	140	-1	133	-1.0	76	341
8	2	157.1	45	25.60	-1	-1	-1	103	6.0	100	176
9	2	165.7	48	34.30	90	140	-1	117	5.9	133	208
10	2	146.2	52	23.63	70	110	342	94	6.0	125	130

그림 2. 전처리 전 사용자 로그 데이터

Fig. 2. Unprocessed User Log Data

전처리를 통해 사용할 컬럼을 제외하고 사용하지 않을 컬럼과 -1 또는 값이 튀는 이상치 데이터를 삭제한다. 그림 2는 전처리 전, 그림 3은 전처리 후 사용자 로그 데이터를 Python의 pandas 데이터프레임을 사용해 불러온 결과이다. 성별 컬럼의 1 값은 남성을 의미하며 2 값은 여성을 의미한다.

성별	키	나이	BMI	이완기혈압	수축기혈압	공복혈당	당화혈색소	콜레스테롤	중성지방	
1	1	182.2	46	27.23	90	130	140	7.7	139	154
2	1	176.2	52	25.03	100	150	108	6.2	142	198
3	1	178.7	68	19.29	110	160	116	6.0	52	124
4	1	164.2	62	26.44	98	175	196	8.4	120	205
5	1	157.8	63	26.38	90	130	154	7.9	149	164
9	2	165.7	48	34.30	90	140	117	5.9	133	208
10	2	146.2	52	23.63	70	110	94	6.0	125	130
14	2	152.5	64	26.53	92	185	123	9.2	136	129

그림 3. 전처리 된 사용자 로그 데이터  
 Fig. 3. Preprocessed User Log Data

나. 지표별 사용자 건강점수 계산

위에서 생성된 세 가지 지표인 심혈관질환, 당뇨관리, 지방대사에 대한 각 건강점수를 계산한다. 먼저, 사용자의 로그 데이터 변수 수치 값과 각 변수의 정상 수치 값을 유클리드 거리 기반 알고리즘을 사용해 세 가지 지표 별로 오차를 계산한다. 이는 Scipy 라이브러리의 distance.euclidena 메서드를 사용하여 유클리드 거리를 계산하였다. 사용자 별로 계산된 세 가지 건강 점수는 가시성을 위해 데이터프레임의 오른쪽에 새로운 컬럼의 변수로 생성해 데이터를 저장하였다. 추가되는 열은 심혈관질환 점수, 당뇨관리 점수, 지방대사 관리 점수로 cd\_socre, dm\_score, fmm\_score로 표기해 값을 저장한다. 그림 4는 유클리드 거리를 사용해 계산된 건강 점수 값이다.

BMI	이완기혈압	수축기혈압	공복혈당	당화혈색소	콜레스테롤	중성지방	cd_score	dm_score	fmm_score
27.23	90	130	140	7.7	139	154	11.989533	14.142136	39.204592
25.03	100	150	108	6.2	142	198	18.153151	18.006943	63.780875
19.29	110	160	116	6.0	52	124	32.218506	10.004499	54.589376
26.44	98	175	196	8.4	120	205	37.503754	70.052052	58.523500
26.38	90	130	154	7.9	149	164	11.709415	28.086296	50.960769
34.30	90	140	117	5.9	133	208	12.448293	9.002222	66.730802
23.63	70	110	94	6.0	125	130	33.548963	32.001406	32.015621
26.53	92	185	123	9.2	136	129	45.685631	4.609772	41.677332
29.43	110	160	196	8.9	168	177	32.674775	70.073105	73.164199

그림 4. 유클리드 계산 후의 건강 점수  
 Fig. 4. Health Score after Euclidean Calculation

이후, 계산된 각 지표별 점수에 나이에 따른 가중치를 추가한다. 나이에 따른 가중치는 나이대별로 부과한다. 40대는 0.1, 50대는 0.2, 60대는 0.3, 70대 이상은 0.4로 가중치 총합을 1로 잡아 곱해주었다. 계산한 점수는 전과 같이 가시성 있게 데이터프레임의 오른쪽에 새로운 컬럼을 추가하고 데이터를 저장하였다. 나이 가중치를 추가한 심혈관질환 점수, 당뇨관리 점수, 지방대사 관리 점수로 a\_cd\_socre, a\_dm\_score, a\_fmm\_score의 각 컬럼명을 가진다.

cd_score	dm_score	fmm_score	a_cd_score	a_dm_score	a_fmm_score
11.989533	14.142136	39.204592	1.198953	1.414214	3.920459
18.153151	18.006943	63.780875	3.630630	3.601389	12.756175
32.218506	10.004499	54.589376	9.665552	3.001350	16.376813
37.503754	70.052052	58.523500	11.251126	21.015616	17.557050
11.709415	28.086296	50.960769	3.512825	8.425889	15.288231
12.448293	9.002222	66.730802	1.244829	0.900222	6.673080
33.548963	32.001406	32.015621	6.709793	6.400281	6.403124
45.685631	4.609772	41.677332	13.705689	1.382932	12.503200

그림 5. 나이 가중치를 반영한 건강 점수  
 Fig. 5. Age-Weighted Health Score

이와 같이, 사용자별 건강점수를 계산하고 나이에 따른 가중치를 반영하여 최종 건강 점수를 도출했다. 이를 당뇨병 환자의 식단에 반영하여 식단을 추천한다.

2. 당뇨병 환자 추천 식단 리스트 생성

가. 식품 데이터셋 수집 및 전처리

먼저, 식품 의약품 안전처의 식품 영양성분 데이터베이스를 받아온다. 그림 6은 저장한 데이터를 Python의 pandas를 사용해 데이터프레임 형식으로 불러온 식품 데이터이다.

NO	SAMPLE ID	식품코드	DB명	상품종류	식품명	연도	지역/제조사	채취시기	식품대분류	도요사염타당노신(22.5(n-3))(g)	도코사핵사노신(22.6(n-3))(g)	트랜스지방산(g)	
1	D000006-94-AVG	D000006	음식	종류	병고	2019	중주	평균	구이류	-	0	0.02	0.1
2	D000007-ZZ-AVG	D000007	음식	종류	닭고기	2019	전국(대표)	평균	구이류	-	0.02	0.01	0.2
3	D000008-66-AVG	D000008	음식	종류	닭고기	2019	호남	평균	구이류	-	0.02	0.02	0.3
4	D000009-ZZ-AVG	D000009	음식	종류	닭고기	2019	전국(대표)	평균	구이류	-	0	0	0.04
5	D000010-ZZ-AVG	D000010	음식	종류	떡볶이	2019	전국(대표)	평균	구이류	-	0	0	0
7675	D018915-ZZ-AVG	D018915	음식	외식	삼육항태권국	2020	충남	평균	국밥	-	-	-	0

그림 6. 전처리 전 식품 데이터셋  
 Fig. 6. Unprocessed Food Dataset

데이터를 사용하려는 목적에 맞게끔 전처리하는 과정에서 필요하지 않은 NO, 식품코드, 상품종류, 채취시기, 연도 등의 컬럼을 삭제하고 당뇨인의 식단이 될 수 없는 식품들을 제거한다. 밀가루가 포함되어 있거나 당이 많이 들어 있는 음식, 6가지 식품군에 포함되지 않는 음식을 제거하였다. 또한, 6가지 식품군에 따라 식품을 분류하는 작업을 진행하였다. 어묵볶음의 경우 어육류군에 포함하고 가지나물의 경우 채소군에 포함하는 방식으로 전처리를 진행하였다.

식품명	식품군	식품대분류	식품상세분류	1회제공량	에너지 (kcal)	수분 (g)	단백질 (g)	지방 (g)	탄수화물 (g)	...
평불고기	어육류군	구이류	육류구이	500	368.80	412.6	33.50	8.5	39.7	...
닭갈비	어육류군	구이류	육류구이	400	595.61	276.4	45.90	25.8	44.9	...
닭꼬치	어육류군	구이류	육류구이	70	176.72	35.31	11.56	8.57	13.35	...
더덕구이	채소군	구이류	채소류구이	100	184.00	58.1	3.10	5.2	31.1	...
돼지갈비	어육류군	구이류	육류구이	100	240.32	56.3	19.50	14.4	8.1	...
...	...	...	...	...	...	...	...	...	...	...
설악향태진국	NaN	국 및 탕류	어패류국탕	100	90.00	-	8.00	6	1g 미만	...

그림 7. 전처리 후 식품 데이터셋  
Fig. 7. Preprocessed Food Dataset

나. 지표별 식단 추천 리스트 생성

이후, 각 심혈관질환 지표, 당뇨관리 지표, 지방대사 관리에 도움이 되는 영양성분을 도출하고 이에 따라 식품 추천 리스트를 생성한다. 그림 8은 실험 후 생성된 식품 추천 리스트 CSV 파일이다. cd\_list.csv 파일은 cardiovascular disease의 줄임말로 심혈관질환 지표에 도움이 되는 식품 리스트이고 dm\_list.csv 파일은 diabetes management로 당뇨관리 지표에 도움이 되는 식품 리스트이며 fmm\_list는 fat metabolism로 지방대사 관리 지표에 도움이 되는 식품 리스트이다.

cd_list.csv	2023-06-06 오전 12:43	Microsoft Excel ...
dm_list.csv	2023-06-06 오전 12:43	Microsoft Excel ...
fmm_list.csv	2023-06-06 오전 12:43	Microsoft Excel ...

그림 8. 생성된 식품 추천 리스트  
Fig. 8. Generated Food Recommendation List

3. 당뇨병 환자 식단 추천

위에서 도출한 사용자 건강 점수를 각 지표별 추천 식단 리스트에 반영하여 제일 높은 점수를 기반으로 해당 식단 리스트에서 추천한다. 점수의 값이 높을수록 수치가 정상인 사람과의 차이를 의미하므로 상태가 좋지 않은 지표에 대한 식단 리스트가 선택된다. 추천 방법은 식품군별로 랜덤 추출을 통해 식단을 구성하여 최종적으로 사용자가 섭취해야 하는 하루 적정 섭취 열량과 1일 치 식단이 추천된다. 사용자 로그 데이터를 사용해 하루 섭취 열량을 계산하고 kcal 컬럼을 데이터프레임에 추가해 값을 저장했다. 그림 9는 사용자별로 계산된 kcal 값이 새로운 컬럼으로 추가된 그림이다.

cd_score	dm_score	fmm_score	a_cd_score	a_dm_score	a_fmm_score	kcal
11.989533	14.142136	39.204592	1.198953	1.414214	3.920459	2068.781715
18.153151	18.006943	63.780875	3.630630	3.601389	12.756175	2008.408820
32.218506	10.004499	54.589376	9.665552	3.001350	16.376813	2219.820460
37.503754	70.052052	58.523500	11.251126	21.015616	17.557050	2634.529425
11.709415	28.086296	50.960769	3.512825	8.425889	15.288231	2224.335300
12.448293	9.002222	66.730802	1.244829	0.900222	6.673080	2054.319300
33.548963	32.001406	32.015621	6.709793	6.400281	6.403124	1669.147205
45.685631	4.609772	41.677332	13.705689	1.382932	12.503200	2191.942500
32.674775	70.073105	73.164199	9.802432	21.021931	21.949260	1943.582287
48.364950	50.022495	47.381431	19.345980	20.008998	18.952572	1585.612297

그림 9. 사용자 하루 섭취 열량  
Fig. 9. Daily Calorie Intake for Users

마지막으로, 제안하는 시스템을 통해 결과를 도출한다. 상단의 첫 번째 사용자의 데이터를 사용해 식단 추천을 진행한 결과 하루 섭취 열량, 제일 높은 값의 지표, 1일 치 식단이 도출된다. 총 세 가지 식품군이 추천되며 어육류군은 육류와 어패류 각 1가지씩, 채소군은 3가지씩으로 총 18가지의 식품이 추천된다. 그림 10은 실험에 따른 최종 결과이다.

```

user_health_score=user_health_score(bmi, dbp, sbp, fbs, hba1c, ldl, tg, age)
make_kcal=make_kcal(weight, height, sex)
recommend_user(user_health_score, make_kcal)
✓ 0.0s
섭취 kcal: 2068.781715
값이 제일 높게 나온 지표 : a_fmm_score(지방대사)
    
```

1일치 식단  
곡류군: 잡곡밥 / 어육류군: 두부구이, 소고기 장조림 / 채소군: 김치국/더덕구이/노각무침  
곡류군: 흑미밥 / 어육류군: 갈치구이,만둣국 / 채소군: 배추김치/브로콜리볶음/깍두기  
곡류군: 보리밥 / 어육류군: 오리불고기,가지미꾸이 / 채소군: 무된장국/쪽김나물무침/배추김치

그림 10. 최종 결과  
Fig. 10. Final Result

VI. 결론 및 향후 과제

본 연구에서는 당뇨 환자의 건강 상태를 알 수 있는 라이프로그 OpenAPI 데이터를 활용해 데이터 간의 상관관계를 분석하여 세 가지 지표를 생성한다. 이후, 세 가지 지표를 사용하여 사용자의 건강 상태를 판단할 수 있는 점수를 도출하고, 이 점수를 바탕으로 영양소에 따른 개별화된 맞춤형 식단 추천 시스템을 제안하였다.

제안하는 시스템을 통해 당뇨병 환자는 본인의 건강 상태를 점수라는 정량적인 값으로 파악할 수 있다. 이는 본인의 건강 상태가 어느 부분이 좋지 않은지 파악하여, 해당 결과에 따라 필요한 식단을 추천받음으로써 관리를 할 수 있어 환자의 건강을 개선하는 데 도움이 된다. 또



한, 개별화된 식단을 통해 올바른 식습관을 계획하는 데 도움이 되어 건강한 생활을 지속적으로 유지하여 후에 발생 가능한 합병증 위험을 줄일 수 있다. 이로 인해 당뇨 관리의 효율성이 향상되어 당뇨병 환자들의 삶에 긍정적인 영향을 미칠 것으로 기대된다. 또한, 현재 당뇨병을 앓고 있는 환자 뿐만 아닌 당뇨 발생 가능성이 농후한 일반 사용자도 제안하는 시스템을 통해 상태가 좋지 않은 지표를 파악하고 이에 대한 관리를 받을 수 있다.

추후 과제로는 본 연구의 결과를 더욱 발전시켜 당뇨병을 가진 환자들의 실시간 혈액 데이터를 받아와 식단을 추천해 주는 자동화 시스템을 개발하고자 한다. 또한, 추가적으로 환자 개인의 식단 취향 데이터를 고려해 사용자의 만족도를 높일 수 있는 더욱 정교한 식단 추천 알고리즘을 개발하고자 한다.

## References

- [1] National Health Insurance Corporation : <https://www.nhis.or.kr/nhis/together/wbhaea01600m01.do?mode=view&articleNo=10829528>
- [2] Korea Disease Control and Prevention Agency National Health Information Portal : [https://health.kdca.go.kr/healthinfo/biz/health/gnrlzHealthInfo/gnrlzHealthInfo/gnrlzHealthInfoView.do?cntnts\\_sn=5305](https://health.kdca.go.kr/healthinfo/biz/health/gnrlzHealthInfo/gnrlzHealthInfo/gnrlzHealthInfoView.do?cntnts_sn=5305)
- [3] Korea Diabetes Association : [https://www.diabetes.or.kr/general/dietary/dietary\\_01.php](https://www.diabetes.or.kr/general/dietary/dietary_01.php)
- [4] Korea Diabetes Association : [https://www.diabetes.or.kr/general/dietary/dietary\\_03.php](https://www.diabetes.or.kr/general/dietary/dietary_03.php)
- [5] Korea Diabetes Association : [https://www.diabetes.or.kr/general/dietary/dietary\\_02.php](https://www.diabetes.or.kr/general/dietary/dietary_02.php)
- [6] Nagaraj P, Muneeswaran V, Dharanidharan A, Balanathanan K, Arunkumar M, Rajkumar C, "A Prediction and Recommendation System for Diabetes Mellitus using XAI-based Lime Explainer", 2022 International Conference on Sustainable Computing and Data Communication Systems (ICSCDS), Vol., No., pp. 1472-1478, Apr. 2022. DOI: <https://doi.org/10.1109/ICSCDS53736.2022.9760847>
- [7] Israt Jahan Eva, Sayra Jannat Anika, Mariya Akter Mitu, Abdus Sattar, "Diabetes Health Care Routine Using Machine Learning", 2022 First International Conference on Electrical, Electronics, Information and Communication Technologies (ICEEICT), Vol., No., pp. 1-5, Feb. 2022. DOI: <https://doi.org/10.1109/ICEEICT53079.2022.9768659>
- [8] Robert A Sowah, Adelaide A Bampoe-Addo, Stephen K Armoo, Firibu K Saalia, Francis Gatsi, Baffour Sarkodie-Mensah, "Design and development of diabetes management system using machine learning." International journal of telemedicine and applications, Vol. 2020, No., pp. 17, Jul. 2020. DOI: <https://doi.org/10.1155/2020/8870141>
- [9] Oshani Seneviratne, Jonathan Harris, Ching-Hua Chen, Deborah L. McGuinness, "Personal Health Knowledge Graph for Clinically Relevant Diet Recommendations", eprint arXiv:2110.10131, Vol., No., pp. - , Oct. 2021. DOI: <https://doi.org/10.48550/arXiv.2110.10131>
- [10] Salliah Shafi Bhat, Gufran Ahmad Ansari, "Predictions of Diabetes and Diet Recommendation System for Diabetic Patients using Machine Learning Techniques," 2021 2nd International Conference for Emerging Technology (INCET), Vol., No., pp. 1-5, May. 2021. DOI: <https://doi.org/10.1109/INCET51464.2021.9456365>
- [11] K. Renuka Devi, J. Bhavithra, A. Saradha, "Diet recommendation for glycemic patients using improved K-means and Krill-Herd optimization.", ICTACT Journal on Soft Computing, Vol. 10, No. 03, pp. 2096-2101, Apr. 2020. DOI: <https://doi.org/10.21917/ijsc.2020.0298>
- [12] Nadia Tabassum, Abdul Rehman, Muhammad Hamid, Muhammad Saleem, Saadia Malik, Tahir Alyas, "Intelligent nutrition diet recommender system for diabetic's patients", Intelligent Automation and Soft Computing, Vol. 30, No. 17, pp. 319-335, Aug. 2021. DOI: <https://doi.org/10.32604/iase.2021.018870>
- [13] Maryam Sadat Amiri Tehrani Zadeh, Juan Li, Shadi Alian, "Personalized Meal Planning for Diabetic Patients Using a Multi-Criteria Decision-Making Approach", 2019 IEEE International Conference on E-health Networking, Application & Services (HealthCom), Vol., No., pp. 1-6, Oct. 2019. DOI: <https://doi.org/10.1109/HealthCom46333.2019.9009593>
- [14] LIFELOG : <https://www.bigdata-lifelog.kr/portal>
- [15] Young-Ho Lee, Jee-Song Park, Jong-Hun Kim, Byung-Mun Lee, "Personalized Diet Recommendation Service for management of coronary heart disease", The Journal of KIIT, Vol. 8, No. 5, pp. 189-197, May. 2010.
- [16] Gwi-ok Jeong, Jin-ho Jeon, Hye-suk Son, Jeong-hak Gang, Hwi-dong Kim, Gyu-il Jo, Chae-eon Lee, "A Study on the Risk Factors Affecting Blood Pressure", Korean Journal of Epidemiology, Vol. 17, No. 2, pp. 201-213, Dec. 1995.
- [17] Chang-Gyu Park, M.D., "Obesity and Cardiovascular Disease", Korea Circulation Journal, Vol. 27, No. 1, pp. 130-134, 1997.
- [18] American Diabetes Association and others, "Standards of medical care in diabetes", Diabetes care, Vol. 28, No. 1, pp. S11-S61, Jan. 2010.

DOI: <https://doi.org/10.2337/dc10-S011>

- [19] Rupali Singh, Patricia M. Kluding, "Fatigue and related factors in people with type 2 diabetes", *The Diabetes Educator*, Vol. 39, No. 3, pp. 320-326, Mar. 2013. DOI: <https://doi.org/10.1177/0145721713479144>
- [20] Yeong-Mi Seo, Won-Hee Choi, Jong-Ryeal Hahm, "Influential Factors on Fatigue in Patients with Type 2 Diabetes Mellitus", *Journal of the Korea Academia-Industrial cooperation Society*, Vol. 15, No. 9, pp. 5700-5707, Sep. 2014. DOI: <https://doi.org/10.5762/KAIS.2014.15.9.5700>
- [21] Ministry of Food and Drug Safety : <https://various.foodsafetykorea.go.kr/nutrient/>

### 저 자 소 개

#### 김 선 아(학생회원)



- 2019년 3월 ~ 현재 : 군산대학교 소프트웨어학부 학사과정
- 주관심분야 : 추천 시스템, 의료데이터, 인공지능, 소프트웨어공학

#### 황 만 수(정회원)



- 1986년 8월 : 중앙대학교 전자계산학과 (이학석사)
- 2001년 8월 : 숭실대학교 컴퓨터학과 (공학박사)
- 1993년 3월 ~ 현재 : 신한대학교 IT융합공학부 정교수
- 주관심분야 : 요구공학, 소프트웨어공학

#### 김 능 회(정회원)



- 2008년 8월 : 고려대학교 컴퓨터학과 (이학석사)
- 2017년 8월 : 고려대학교 컴퓨터·전파통신공학과 (공학박사)
- 2021년 3월 ~ 현재 : 군산대학교 소프트웨어학부 조교수
- 주관심분야 : 요구공학, 소프트웨어공학