

# 근적외선 분광법 기반 비침습식 혈당 수치 추정 알고리즘 연구

강영만\* · 한순희\*\*

Research on a Non-invasive Blood Glucose level Estimation Algorithm based on Near- infrared Spectroscopy

Young-Man Kang\* · Soon-Hee Han\*\*

## 요 약

혈당 수치를 확인하기 위해 사용하는 채혈식 혈당 측정기의 불편함을 해소하기 위한 다양한 방법이 시도되고 있다. 본 논문에서는 근적외선 센서를 이용하여 획득한 스펙트럼 데이터로부터 머신러닝 기술을 사용하여 비침습적으로 혈당 수치를 추정하고자 하였다. 연구에 사용한 비침습식 혈당측정기는 가시광선을 포함하여 총 6개로 구성된 근적외선 발광부와 이를 수신하는 수광부를 가지는데, 손가락과 같은 인체의 특정 부위에 대한 스펙트럼 데이터를 수집하기 위해 만든 기기이다. 혈당 수치에 따라 유의미한 차이가 있는지를 검증하기 위해 머신 러닝 알고리즘들을 통해 혈당 수치 추정을 시도하였다. 수집한 데이터에 5가지 머신 러닝 알고리즘 기법을 적용하면서 다양한 하이퍼 파라미터를 조정한 결과, 서포트 벡터 회귀 알고리즘이 가장 좋은 성능을 나타냄을 확인하였다.

## ABSTRACT

Various methods are being attempted to resolve the inconvenience of blood glucose meters used to check blood sugar levels. In this paper, we attempted to estimate blood sugar levels non-invasively using machine learning technology from spectral data acquired using a near-infrared sensor. The non-invasive blood glucose meter used in the study has a total of six near-infrared ray emitters, including visible rays, and a light receiver that receives them. It is a device created to collect spectral data on specific parts of the human body, such as the fingers. To verify whether there was a significant difference depending on blood sugar level, we attempted to estimate blood sugar level through machine learning algorithms. As a result of applying five machine learning algorithm techniques to the collected data and adjusting various hyper parameters, it was confirmed that the support vector regression algorithm showed the best performance.

## 키워드

Near-Infrared Spectroscopy, Blood glucose, Non-invasive, Machine Learning  
근적외선 분광법, 혈당, 비침습식, 기계 학습

\* 전남대학교 문화콘텐츠학부 교수(ymkang@jnu.ac.kr)

† 교신저자 : 전남대학교 문화콘텐츠학부 교수

• 접수일 : 2023. 10. 11

• 수정완료일 : 2023. 11. 11

• 게재확정일 : 2023. 12. 27

• Received : Oct. 11, 2023, Revised : Nov. 11, 2023, Accepted : Dec. 27, 2023

• Corresponding Author : Soon-Hee Han

Division of Culture Contents, Chonnam National University

Email : shhan@jnu.ac.kr

## I. 개요

당뇨병 환자들이 일반적으로 사용하는 자가혈당 측정기는 손가락 끝에서 채혈되는 소량의 혈액으로 혈당을 측정한다. 이러한 채혈식 혈당 측정기의 사용에 따른 고통과 불편함을 해소하고자 전 세계적으로 오랜 시간 동안 비침습적으로 혈당 수치를 추정하기 위한 많은 연구들이 진행되어 왔다. 이들 연구에는 다양한 측정 센서와 다양한 추정 알고리즘에 대한 시도가 있었다[1-2]. 비침습식 혈당 추정은 어려운 과제로 이를 극복하기 위해 최근 몇 년 동안 여러 머신 러닝 모델이 사용되었다. 또 다른 연구에서는 제1형 당뇨병 (Type 1 Diabetes) 환자의 혈당 수치를 추정하는 개인화 모델을 구현하기 위해 랜덤 포레스트 회귀 (Random Forest Regression) 모델을 사용하였다[3].

PhotoPlethysmoGraphy(PPG) 데이터로부터 다양한 머신 러닝 모델을 사용하여 혈당 수치, 수축기 및 이완기 혈압을 동시에 추정하는 시스템을 제시하는 연구도 있다[4]. Ju-Hyun Baek 등은 근적외선 분광법을 이용하여 생쥐의 꼬리 부분에서 얻어진 스펙트럼 데이터를 주성분 분석과 부분 최소 제곱법 회귀 모델의 적용으로 개선된 재현성 및 정확성을 가진 검량 모델을 얻고자 하였다[5].

본 연구에서는 근적외선 센서로 획득한 스펙트럼 데이터로부터 머신러닝 기술을 사용하여 비침습적으로 혈당 수치(Blood Glucose Level)를 추정하는 것을 목표로 한다. 그림 1과 같이 총 6개로 구성된 가시광선을 포함하는 근적외선 발광부와 이를 수신하는 수광부를 가지는 근적외선 센서를 사용한다. 근적외선 센서는 이전 연구[6]에서 제시한 시스템을 보완하여 수정 설계한 것으로 이 센서를 이용하여 인체의 특정 부위(예: 손가락 끝)에서 스펙트럼 데이터를 수집하였다. 수집한 데이터를 이용하여 혈당 수치에 따라 유의미한 차이가 있는지, 여러 가지 머신 러닝 알고리즘들을 통해 혈당 수치 추정을 시도하였다.

본 논문의 구성은 다음과 같다. 2장에서는 센서 데이터의 수집에 대해 설명하고, 3장에서는 혈당 수치 추정을 위한 알고리즘 적용에 대해 자세히 알아본다. 4장

에서는 알고리즘 별 평가 결과를 분석하고, 마지막으로 5장에서 결론 및 향후 연구 방향에 대해 논의한다.

## II. 데이터 수집 및 전처리

### 2.1 센서 데이터의 수집

센서 데이터의 수집을 위해서 그림 1과 같이 총 6개로 구성된 가시광선을 포함하는 근적외선 발광부와 이를 수신하는 수광부를 가지는 근적외선 센서를 사용한다.



그림 1. 혈당 측정기와 센서부  
Fig. 1 Blood glucose meter and sensor unit

센서 모듈은 광원 제어를 통해 6개의 광원으로부터 순차적인 단일 광원의 사용과 전체 광원을 사용함으로써 총 7가지의 광원 상태를 가진다. 1회 측정 시의 각 광원 상태는 약 7초간 유지하며, 전체 약 50초간의 스펙트럼 데이터를 수집한다. 그림 2는 광원 상태에 따른 스펙트럼 데이터의 예시를 나타내고 있다. 각 광원 별 신호세기의 꼭대기(peak) 부분에 해당하는 파장대역이 상이함을 알 수 있다.

각 센서 데이터의 취득 시에는 별도의 침습식 혈당 측정기(Accu-Chek)를 통해 취득 시점의 혈당 수치를 측정하여 데이터의 레이블(label)로 함께 기록한다.

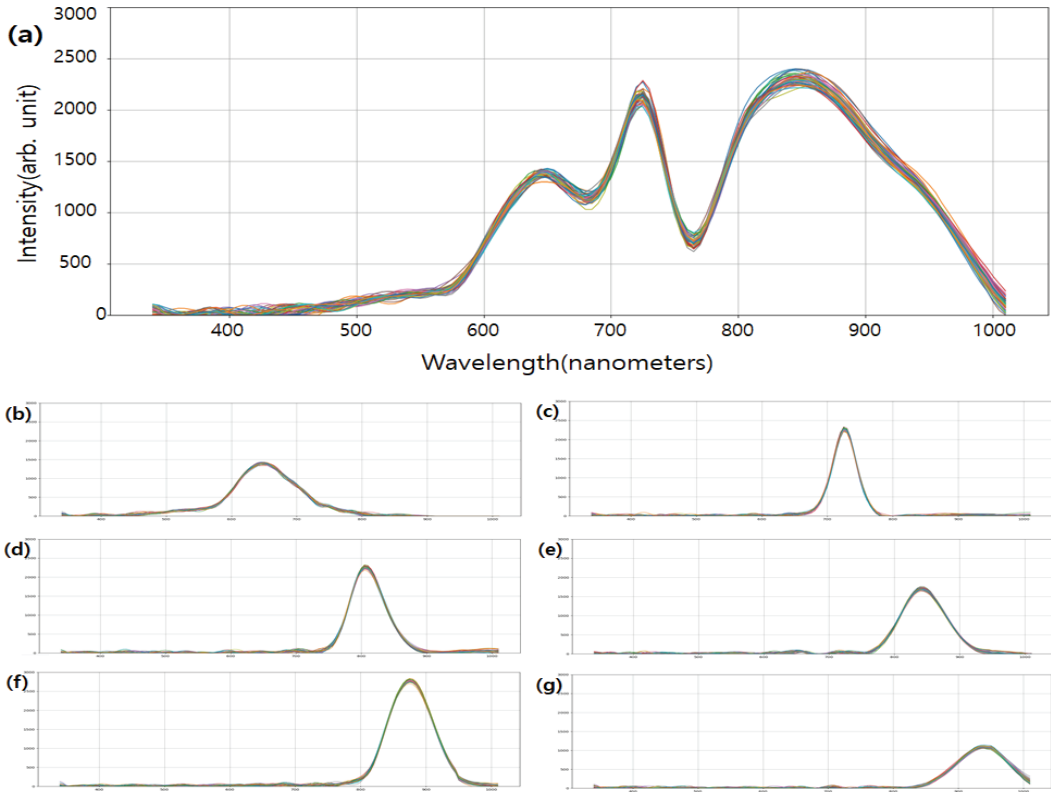


그림 2. 광원 상태에 따른 스펙트럼 데이터 예시  
 (a: All, b: White, c: Far-Red 735, d: NIR 810, e: NIR 850, f: NIR 880, g: NIR 940)  
 Fig. 2 Spectral data example according to light source status

### 2.2 데이터 전처리

기기에서 측정되는 스펙트럼은 대략 350~1010nm 범위에서 5nm간격의 해상도를 가진다. 이 스펙트럼 데이터는 각 광원의 상태에 따라 대상 물체로부터 반사/흡수되어 수신되는 광도가 일정 수준 이상의 것을 유효한 스펙트럼 범위로 제한하여 사용한다.

수집된 센서 데이터는 초당 5~10hz의 샘플링 데이터를 수집하게 되며, 일련의 배열 형태의 수치 데이터로 표현된다. 이 데이터를 각 광원 상태에 따른 구분된 데이터로 그룹화하고, 노이즈의 영향을 줄이기 위해 측정 데이터가 아닌 평균 스펙트럼 데이터를 일반적으로 사용한다. 또한, 광원의 제어에 따라 상태 전환 시 발생하는 왜곡된 데이터 부분이 있는 경우 해당 샘플링 데이터 구간은 제거한다.

센서 데이터는 단순 평균 스펙트럼 데이터를 그대로 사용하는 것 외에도, 스펙트럼 데이터의 전처리로 1차 미분, 2차 미분, SNV( Standard Normal Variate), MSC( Multiplicative Scatter Correction), OPLS( Orthogonal Projection to Latent Structures) 등의 적용을 시도하였으나, 최종적으로 식(1)과 같이 1차 미분을 적용하였다[7]. 이는 높은 주파수 성분을 제거하고 낮은 주파수 성분을 강화하여 스펙트럼 데이터에서의 노이즈를 감소시키고 신호 대 노이즈 비율(SNR)을 향상시킬 수 있다. 또한 데이터에서의 변화의 위치와 크기를 강조함으로써 어떤 특징이 있는지 더 잘 확인할 수 있다.

$$f(\lambda) = \frac{dI(d\lambda)}{d\lambda} \dots (1)$$

여기서  $d\lambda$ 는 주파수 영역에서의 주파수 변화를 나타내며,  $dI(d\lambda)$ 는 해당 주파수 변화에 따른 스펙트럼 데이터 값의 변화를 나타낸다.

### 2.3 특성 생성

가공된 센서 데이터는 각 광원 별 7초간의 평균 스펙트럼 데이터로, 조정된 유효한 파장 범위 내에서 5nm간격의 수치 정보를 나타낸다. 본 연구에서는 7가지 광원 상태에 대해서 전처리 과정을 거친 데이터로부터 총 390개의 특성(feature)을 생성하였다. 표 1은 광원에 따른 유효 파장 범위와 생성된 특성 개수를 나타내고 있다. 비교적 많은 수의 특성은 이후 머신러닝 알고리즘 적용 시점에 특성 선택 알고리즘을 통해 지정된 비율로 조정하여 사용된다.

표 1. 유효 파장 범위와 특성 개수  
Table 1. Effective wavelength range and number of features

light source classification	identification mark	Effective wavelength range(nm)	Number of features
All	#all	400 ~ 1010	113
White	#1	400 ~ 740	68
Far-Red 735	#2	680 ~ 770	26
NIR 810	#3	755 ~ 875	42
NIR 850	#4	780 ~ 920	61
NIR 880	#5	800 ~ 970	44
NIR 940	#6	860 ~ 1010	36

### III. 실험 혈당 수치 추정을 위한 알고리즘 적용

본 연구에서는 센서 데이터로부터 생성한 특성 정보들에 대해 머신러닝 기법을 활용하여 혈당 수치를 추정하고자 다양한 알고리즘의 적용을 시도하였다.

알고리즘 적용을 위해 준비된 데이터 샘플은 약 1,200개이며, 이 중 약 8 : 2 비율로 학습/검증용 샘플과 테스트용 샘플로 구분하여 사용하였다. 이 샘플은 약 3개월에 걸쳐 성인 남성 실험 대상자의 혈당치 정보를 기록한 것이고, 각 센서 데이터에 대한 침습식 혈당측정기를 통해 혈당 수치가 함께 기록되어 있다. 표 2는 실험에 사용한 데이터의 측정일과 데이터 개수이다.

표 2. 실험 데이터  
Table 2. Experimental data

Date	Number of measured data
2021/7	145
2021/8	93
2021/9	605
Total	843

실험 데이터의 혈당 수치 구간별 분포도는 그림 3과 같다.

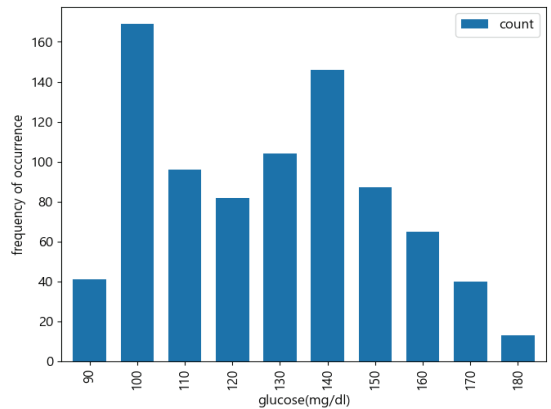


그림 3 실험 데이터의 혈당 수치 구간별 분포도  
Fig. 3 Distribution chart of blood glucose level by section of experimental data

### 3.1 학습용 샘플의 선정

학습용 샘플에 대해 지식 기반의 이상치(outlier) 제거 후 학습 적용을 시도하였다. 예를 들면, 혈당 수치가 유사한 샘플들에 비해 센서 데이터의 스펙트럼 레벨이 유사한 다른 샘플과는 크게 차이가 나는 경우 학습 샘플에서 제외할 수 있다. 또는 각 샘플의 생성된 특성 정보에 대해 이상탐지(Anomaly Detection) 알고리즘을 적용하여 유사한 혈당 수치를 가지는 샘플들 중에서 이상치를 가지는 샘플을 필터링할 수 있다.

이상탐지 알고리즘으로는 데이터 포인트들을 고립된 트리 구조로 표현하고 이상치는 더 짧은 경로로 고립될 가능성이 높다는 가정을 기반으로 하는

Isolation forest, 가우스 분포로부터 벗어난 데이터를 이상치로 판단하는 Elliptic envelope, 데이터 포인트 주변의 밀도를 측정하여 주변 데이터와 비교하여 밀도가 낮은 지역에 위치하면 이상치로 판단하는 LOF(Local Outlier Factor) 등이 있다[8]. 우리는 이상탐지 알고리즘으로 Isolation forest 알고리즘을 적용하였다. 이는 다른 알고리즘에 비해 빠른 속도로 이상치를 탐지할 수 있으며, 데이터 분포와 상관없이 이상치를 잘 감지하며, 고차원 데이터에도 적용이 가능하였다. 또한 주요 파라미터의 수가 적어 튜닝이 상대적으로 간단하다는 장점이 있다.

### 3.2 특성 선택

모델 학습에서 불필요한 특성을 제거하면 모델이 더 잘 일반화될 수 있어, 적절한 특성의 선택은 모델의 성능을 향상시킬 수 있다. 학습에 사용되는 특성의 개수에 비해 준비된 샘플 수는 머신 러닝 학습에 사용하기에는 비교적 적은 수에 해당한다. 학습 샘플 수의 제약으로 인해 변별력 있는 특성들의 선택이 필요하여 특성 개수를 줄여서 학습이 진행되도록 시도하였다.

특성 선택을 위한 방법으로는 단순히 분산이 낮은 특성을 제거하는 방법 외에도 다양한 알고리즘들이 존재한다[9]. 우리는 특성 선택 알고리즘으로 RFE(Recursive Feature Elimination)를 적용하였다. RFE는 모델 기반의 재귀적인 특성 선택 알고리즘으로, 여기서는 외부 추정기로 Random Forest Regressor를 사용하여 각 특성에 대한 가중치가 적은 특성들을 재귀적으로 제거하게 된다. 이는 특성 선택에 대한 최적의 결과가 산출되지 않거나 계산 비용이 높은 편이지만, 자동화된 특성 선택으로 시간을 절약하고 모델 튜닝에 집중할 수 있게 하는 장점이 있다[10].

### 3.3 학습 알고리즘

레이블이 지정된 샘플들에서 특징 벡터로부터 혈당 수치를 추정하는 함수를 추론하는 알고리즘에 대해 설명한다. 학습 알고리즘으로 생성되는 모델은 예측 정확도가 높고, 입력되는 데이터에 내제되어 있는 노이즈에 강해야 하며, 결과를 예측하는데 리소스 및 계산 비용이 낮아 휴대용 기기에 적용하기 적합한 것이

어야 한다. 또한 성능 향상을 위한 학습 데이터의 구성의 일부 변화에도 성능 차이가 크게 발생하지 않아야 한다.

학습은 전체 샘플에서 약 80%에 해당하는 샘플을 학습용 샘플로 사용했으며, 나머지 약 20%는 테스트 목적의 평가용 샘플로 사용하였다. 학습용 샘플은 내부적으로 교차 검증용을 위해 훈련용과 검증용으로 구분하여 사용된다.

본 연구에서는 특성이 다른 5종류의 예측 알고리즘을 테스트하였다. 테스트를 위해 회귀 분석 기법으로 선형 회귀(LR: Linear Regression), 랜덤 포레스트 회귀(RFR: Random Forest Regression), 서포트 벡터 회귀(SVR: Support Vector Regression), 다층 퍼셉트론 회귀(MLPR: Multi-Layer Perceptron Regression), 부분 최소 제곱법 회귀(PLSR: Partial Least Squares Regression) 다섯 가지 알고리즘을 적용하였다.

선형 회귀는 가장 기본적이면서 간단한 회귀 분석 기법 중의 하나로 특성과 레이블에 해당하는 혈당 수치의 관계를 이해하고 설명하기 쉽다는 장점을 가진다. 그러나 관계를 선형으로 가정하기 때문에 비선형 관계인 경우 문제의 적용이 어려울 수 있다. 또한 이상치에 민감하며 변수가 너무 많은 경우 과적합 문제가 있을 수 있다. 이에 특성 선택 알고리즘에서 특성의 수가 너무 많지 않도록 조정하며, 과적합의 문제를 해결하기 위해 릿지(Ridge)나 라쏘(Lasso)와 같은 규제 기법의 적용을 시도하였다[11-12]. 이 기법에서는 특성 수의 비율을 약 0.05로 설정하고 릿지 규제 기법을 적용했을 때 가장 좋은 결과를 보였다.

랜덤 포레스트 회귀는 앙상블 학습 방법 중 하나로, 여러 결정 트리(Decision Tree) 모델을 조합하여 강력한 예측 성능을 제공한다[13]. 각각 다른 데이터 부분 집합(부트스트랩 샘플)과 무작위로 선택된 변수로 다수의 결정 트리를 학습하므로 모델의 다양성과 안정성이 높으며, 변수의 중요도를 평가하여 중요한 변수를 기반으로 예측을 수행한다. 특성이 많은 고차원 데이터에서도 잘 동작하며 이상치와 노이즈에 대해서도 강한 편이다. 그러나 너무 깊은 트리의 경우 과적합 문제가 발생할 수 있으며, 변수 중요도를 제공하지만 왜 그런 예측을 한 근거에 대한 설명력이 부족할 수 있다. 이 기법에서는 특성 수의 비율을 약

0.5로 설정하고 트리의 최대 깊이를 7로 제한하여 수행했다.

서포트 벡터 회귀는 SVM(Support Vector Machine)에서 회귀 문제를 해결하기 위한 확장된 형태의 머신러닝 알고리즘으로 데이터 포인트와 모델 사이의 마진(간격)을 최대화하는 것을 목표로 한다[14]. 또한 다양한 커널 함수를 사용하여 데이터의 비선형 관계를 모델링할 수 있다. 그러나 매개변수의 선택과 모델 해석의 어려움이 있으며, 큰 데이터셋에서는 계산 비용이 높을 수 있다. 이 기법에서는 특성 수의 비율을 약 0.11로 설정하고, 커널(kernel)은 linear, 오류 허용 정도를 제어하는 Cost는 약 0.67로 설정한 경우 가장 좋은 성능을 보였다.

다층 퍼셉트론 회귀는 인공 신경망을 기반으로 한 회귀 모델로 여러 개의 은닉층을 가진 다층 신경망으로 구성되며 이 다층 구조를 통해 복잡한 비선형 관계를 모델링할 수 있다[15]. 각 은닉층의 뉴런에는 비선형 활성화 함수 (예: ReLU, Sigmoid)가 적용되어 다양한 함수 형태를 학습할 수 있다. 입력 데이터로부터 중요한 특성들을 자동으로 추출하고 학습해서 고차원 데이터의 복잡한 패턴을 모델링할 수 있다. 그러나 은닉층과 뉴런 수, 학습률 및 규제 파라미터를 잘못 설정하면 모델이 과적합 될 수 있고 작은 데이터셋에서도 과적합의 위험이 높을 수 있다. 이 기법은 입력 데이터의 스케일링이 중요하다. 따라서 여기서는 표준화(standardization) 스케일링을 적용하였다. 특성 수의 비율은 1.0으로 설정하고 다층 퍼셉트론 구조(architecture)로는 3개의 히든 레이어(hidden layer)를 사용하였으며, 활성화 함수는 ReLU를 사용하였다.

부분 최소 제곱법 회귀는 회귀 분석을 위한 다변량 통계 기법 중 하나로, 주로 고차원 데이터셋의 변수 선택 및 차원 축소에 사용된다[16]. 독립 변수 사이의 다중 공선성 문제를 효과적으로 다룰 수 있어 모델의 안정성을 향상시키고 과적합을 방지할 수 있다. 또한 변수 간 상관관계 및 변수의 중요성에 대한 정보를 제공하기 때문에 결과를 해석하기가 비교적 쉽다. 이 기법은 데이터의 스케일링과 노이즈 처리에 민감하므로 데이터 전처리가 중요하며, 하이퍼 파라미터 설정이 부적절할 경우 과적합 문제가 발생할 수 있다. 이 기법은 일반적인 다른 회귀 기법에 비해 다중 공선성 문제를 효과적으로 처리하려면 더 많은 데이터가 필

요할 수 있다. 또한 고차원 데이터셋에서 변수의 수에 비해 데이터 포인트 수가 적으면 차원 축소가 어려울 수 있다. 이 기법으로 학습 시 하이퍼 파라미터로 특성 수의 비율을 약 0.29로 설정하고, 주성분에 해당하는 컴포넌트 개수를 2로 설정하였을 때 가장 좋은 성능을 나타냈다.

위의 각 회귀 기법들에 적용되는 하이퍼 파라미터는 다양하게 있으나, 이들을 수동으로 설정하여 확인하기에는 너무 많은 시도와 경험이 필요하다. 특히 다층 퍼셉트론 회귀의 경우 최적의 구조를 생성하는 것은 어려운 작업이다. 따라서 최적의 하이퍼 파라미터 조합을 찾기 위해 그리드 서치(Grid Search)나 랜덤 서치(Random Search), 베이저안 최적화(Bayesian Optimization)의 방법을 사용해 볼 수 있다[17].

### 3.4 평가 결과

회귀 분석 모델의 성능 평가로는 여러 가지 평가 지표가 있다. 일반적으로 사용되는 지표로는 평균 제곱 오차(Mean Squared Error), 평균 절대 오차(Mean Absolute Error),  $R^2$ (R-squared) 등을 사용하여 평가한다[18]. 그러나 여기에서는 혈당 측정 시스템의 정확도에 관한 국제 표준인 ISO 15197에서 제시하는 조건에 따라 평가하였다[19]. 각 샘플에서 레이블 값으로 가지는 혈당 수치와 알고리즘을 통해 추정된 혈당 수치에 대해, 혈당 농도가  $\pm 15\%$  범위 내에 있어야 하며, 농도가 100 mg/dL미만인 경우에는  $\pm 15$  mg/dL 내에 있어야 하는 조건을 만족하는지 여부의 정확도로 평가하였다. 추가 조건으로 Consensus Error Grid Analysis를 사용하여 혈당 측정 값이 오차 그리드의 A와 B Zone 내에 99%이상 포함되는지 여부를 확인하였다. 또한 추가적인 성능 평가 지표로 MARD(Mean Absolute Relative Difference)[20]에 대한 평가도 추가 산출하였다. 표 3에서는 각 알고리즘 별로 평가한 최종 결과를 보여준다.

표 3. 각 알고리즘별 평가 결과  
Table 3. Evaluation results for each algorithm

Algorithm	ISO 15197:2013 (%)	Consensus Error Grid Analysis A&B Zone distribution map(%)	MARD(%)
LR	76.8	100.0	10.9
RFR	74.7	100.0	11.4
SVR	82.8	100.0	9.4
MLPR	72.5	100.0	12.4
PLSR	80.3	100.0	9.8

시스템의 알고리즘의 최종 선택은 ISO 15197에서 제시된 혈당 농도 기준의 정확도를 주된 평가 척도로 선정하였으며, 여기서는 표 3에서와 같이 가장 성능이 좋은 서포트 벡터 회귀 알고리즘이 적용되었다. 이 알고리즘은 혈당 농도가 ±15% 범위 내에서는 82.8%, ±20% 범위 내에서는 94.4%, ±40% 범위 내에서는 99.6%의 성능을 나타내었으며, 오차 그리드(Error Grid) 분석 결과는 표 4와 같다.

표 4. SVR 알고리즘 오차 그리드 분석 결과  
Table 4. Error grid analysis results of SVR algorithm

algorithm	A zone	B zone	C zone	D zone	E zone
SVR	97.85%	2.15%	0.00%	0.00%	0.00%

Parkes 오류 그리드 분석은 혈액 내 실제 및 예측 혈당 수준 모니터링을 위해 사용되는 것으로, 각 영역(A, B, C, D, E)은 측정된 혈당값의 오류로 인해 부정적인 결과가 발생할 위험 정도를 나타낸다[21]. 그림 4는 Parkes 오차 그리드 분석을 사용하여 혈당 측정값이 오차 그리드의 A와 B Zone 내에 99% 이상 포함되는지를 확인한 것이다.

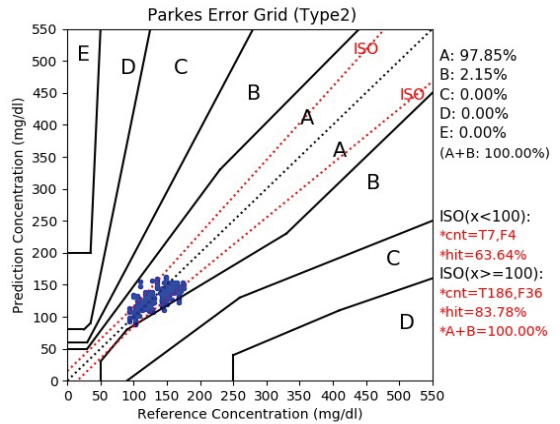


그림 4. Parkes 오차 그리드 분석  
Fig. 4 Parkes error grid analysis

#### IV. 결론 및 향후 연구

근적외선 센서로부터 수집한 스펙트럼 데이터에 대해 머신 러닝 기술을 사용하여 비침습적으로 혈당 수치를 보다 나은 정확도로 추정하였다. 이를 위해, 이상치 데이터를 제거하기 위한 샘플의 선정, 불필요한 특성을 제거하여 모델의 일반화가 잘 생성될 수 있도록 다양한 방법들을 시도해 보았다. 동일한 샘플에 대해 머신 러닝의 5가지 알고리즘을 적용하고 다양한 하이퍼 파라미터를 조정한 결과, 서포트 벡터 회귀 알고리즘이 가장 좋은 성능을 나타냄을 확인하였다.

앞으로는 근적외선 센서 수집 단계에서 인체의 특정 부위에 대한 스펙트럼 측정 시, 보다 정교하고 안정적인 측정을 위해 기구의 구조 개선 및 추가적인 정보(예: 온도, 압력 등)의 입력이 가능하도록 센서의 확장을 시도해 볼 수 있다. 또한 추정 알고리즘에 대한 안정적인 학습 모델을 생성하기 위해 보다 많은 샘플 데이터의 취득, 즉 다양한 실험자의 샘플 수집과 당뇨병 환자에 대한 고혈당에 대한 샘플의 확보 후 알고리즘의 성능을 추가로 평가할 예정이다.

#### References

[1] L. Tang, S.-J. Chang, C.-J. Chen and J.-T. Liu,

- "Non-Invasive Blood Glucose Monitoring Technology: A Review," *Sensors*, vol. 20, 2020, pp. 6925.
- [2] A. Hina and W. Saadeh, "Noninvasive Blood Glucose Monitoring Systems Using Near-Infrared Technology-A Review," *Sensors*, vol. 22, 2022, pp. 4855.
- [3] M. Syafrudin, G. Alfian, N. L. Fitriyani, I. Fahrurrozi, M. Anshari, and J. Rhee, "A Personalized Blood Glucose Prediction Model Using Random Forest Regression," In *Proc. 2022 ASU Int. Conf. in Emerging Technologies for Sustainability and Intelligent Systems*, Manama, Bahrain, June 2022, pp. 295-299.
- [4] M.-M. Enric, "Non-invasive estimate of blood glucose and blood pressure from a photoplethysmograph by means of machine learning techniques," *Artificial Intelligence in Medicine*, vol. 53, no. 2, Oct. 2011, pp. 127-38.
- [5] J. Baek, N. Kang, Y.Woo, and H. Kim, "Compensation of Variation from Long-Term Spectral Measurement for Non-invasive Blood Glucose in Mouse by Near-Infrared Spectroscopy," *J. of the Pharmaceutical Society of Korea*, vol. 48. no. 3, pp. 177-181.
- [6] Y. Kang and S. Han, "Non-invasive Blood Glucose Detection Sensor System Based on Near-Infrared Spectroscopy," *J. of the Korea Institute of Electronic Communication Sciences*, vol. 16, no. 5, 2021, pp. 991-1000.
- [7] R. Åsmund, F. Berg, and S. Engelsen, "Review of the most common pre-processing techniques for near-infrared spectra," *TrAC Trends in Analytical Chemistry*, vol. 28, no. 10, 2009, pp. 1201-1222.
- [8] F. Jansson, "Comparing machine learning algorithms for detecting behavioural anomalies," Master's Thesis, *Blekinge Institute of Technology, Faculty of Computing*, 2023.
- [9] U. Khaire and R. Dhanalakshmi, "Stability of feature selection algorithm: A review," *J. of King Saud University-Computer and Information Sciences*, vol. 34, no. 4, 2022, pp. 1060-1073.
- [10] K. Kim, S. Han, T. Kim, S. Jung, S. Kim, H. Ceong, and J. Park, "Pattern Analysis of Clinical Signs in Cultured Olive Flounder, *Paralichthys Olivaceus*, with Edwardsielosis using the Decision Tree Technique," *J. of the Korea Institute of Electronic Communication Sciences*, vol. 16, no. 4, 2021, pp. 661-674.
- [11] A. E. Hoerl and R. W. Kennard. "Ridge Regression: Biased Estimation for Nonorthogonal Problems," *Technometrics*, vol. 42, no. 1, 2000, pp. 80-86.
- [12] R. Tibshirani, "Regression Shrinkage and Selection via the Lasso," *J. of the Royal Statistical Society. Series B (Methodological)*, vol. 58, no. 1, 1996, pp. 267-288.
- [13] L. Breiman, "Random Forests," *Machine Learning, Springer Nature*, vol. 45, Oct. 2001, pp. 5-32.
- [14] A. J. Smola and B. Schölkopf, "A tutorial on support vector regression," *Statistics and Computing*, vol. 14, 2004, pp. 199-222.
- [15] H. Kurt, S. Maxwell and W. Halbert, "Multilayer feedforward networks are universal approximators," *Neural Networks*, vol. 2, no. 5, 1989, pp. 359-366.
- [16] G. Geladi and B. R. Kowalski, "Partial least-squares regression: a tutorial," *Analytica Chimica Acta*, vol. 185, 1986, pp. 1-17.
- [17] J. Bergstra and Y. Bengio, "Random search for hyper-parameter optimization," *J. of Machine Learning Research*. vol. 13, 2012, pp. 281-305.
- [18] D. Chicco, M.-J. Warrens, and G. Jurman, "The coefficient of determination R-squared is more informative than SMAPE, MAE, MAPE, MSE and RMSE in regression analysis evaluation," *PeerJ Computer Science*. vol. 7, 2021, pp. e623.
- [19] ISO/TC 212/SC, Int. Std. ISO 15197, *In vitro diagnostic test systems*, Geneva, Switzerland: ISO, 2003.
- [20] G. Freckmann, J. Mende, S. Pleus, D. Waldenmaier, A. Baumstark, N. Jendrike, and



C. Haug, "Mean Absolute Relative Difference of Blood Glucose Monitoring Systems and Relationship to ISO 15197," *J. of diabetes science and technology*, vol. 16, no. 5, 2022, pp. 1089-1095.

- [21] D. Nimmagadda and B. Potluri, "Assessment of Parkes Error Grid through Machine learning techniques," Master's Thesis, *Blekinge Institute of Technology, Faculty of Computing*, 2020.

### 저자 소개



#### **강영만(Young-Man Kang)**

1985년 광운대학교 전자계산학과 졸업(공학사)

1987년 광운대학교 대학원 전자계산학과 졸업(이학석사)

1987 ~1992년 한국전자통신연구원

1992년 전남대학교 문화콘텐츠학부 교수

※ 관심분야 : 방송통신, 무선통신, 헬스케어, 영상처리(카메라),



#### **한순희(Soon-Hee Han)**

1983년 경북대학교 전자공학과 졸업(공학사)

1985년 광운대학교 전자계산학과 졸업(이학석사)

1993년 광운대학교 전자계산학과 졸업(이학박사)

1989년~현재 전남대학교 문화콘텐츠학부 교수

※ 관심분야 : 이동통신, 임베디드시스템, ICT융합

