

라이프로그 데이터를 활용한 랜덤포레스트 및 SHAP 기반 인지기능 장애 예측 모델

이명진*¹, 이종언*², 이한준**

Cognitive Dysfunction Prediction Model with Lifelog Dataset based on Random Forest and SHAP

Myeongjin Lee*¹, Jongun Lee*², and Hanjun Lee**

요 약

본 연구에서는 174명의 인지기능 장애인과 비장애인을 대상으로 최소 35일부터 최대 120일 동안에 걸쳐 추적한 24시간 라이프로그 정보와 간이정신상태검사 데이터를 활용하여 인지기능 장애를 예측하는 머신러닝 모델을 구축하였다. 모델 구축에는 Random Forest 알고리즘을 활용하였고 79.3%의 정확도로 인지기능 장애 여부를 분류할 수 있었다. 또한 SHAP 분석을 통하여 수면 및 신체활동 관련 데이터가 인지기능 장애를 판단하는 중요 변수임을 확인하였다. 이러한 결과는 웨어러블 디바이스를 통해 수집된 데이터를 정량적으로 분석한 결과라는 점에서 응답자의 주관에 담긴 설문 기반의 기존 연구와 차별성을 갖는다. 본 연구의 결과는 인지기능 장애 조기 진단 및 추적관리에 활용될 수 있으며 장애 발생 가능자들의 수면 및 신체 활동 등 생활 습관 개선에도 도움을 줄 수 있을 것이다.

Abstract

In this study, we developed a machine learning model to predict cognitive dysfunction using 24-hour lifelog information tracked over a period ranging from a minimum of 35 days to a maximum of 120 days, as well as Mini-Mental State Examination data, involving 174 individuals with cognitive dysfunction and those without cognitive dysfunction. The model utilized the Random Forest algorithm and achieved an accuracy of 79.3% in classifying cognitive dysfunction status. Additionally, SHAP analysis confirmed that data related to sleep and physical activity were significant variables in determining cognitive dysfunction. These results are different from existing survey-based studies that contain respondents' opinions in that they were derived through quantitative analysis with data collected from wearable devices. The findings of this study can be applied to early diagnosis and tracking management of cognitive dysfunction, as well as potentially contributing to improvements for the individuals at risk of cognitive dysfunction in their lifestyle habits such as sleep and physical activity.

Keywords

lifelog, cognitive dysfunction, machine learning, random forest, AI-Hub

* 명지대학교 경영정보학과 학사과정
- ORCID¹: <https://orcid.org/0009-0009-6446-211X>
- ORCID²: <https://orcid.org/0009-0000-3496-0657>
** 명지대학교 경영정보학과 교수(교신저자)
- ORCID: <https://orcid.org/0000-0002-9005-3661>

• Received: Sep. 18, 2023, Revised: Oct. 18, 2023, Accepted: Oct. 21, 2023
• Corresponding Author: Hanjun Lee
Dept. of Management Information Systems, Myongji University, 34,
Geobukgol-ro, Seodaemun-gu, Seoul, 03674, Korea
Tel.:+82-2-300-0772, Email: hjlee1609@gmail.com

I. 서 론

인지기능 장애에 대한 위험성은 이미 널리 알려져 있다. 인지기능 장애는 인간이 일상생활을 유지하기 위한 능력을 상실시키고, 인간관계에 갈등을 유발시킨다. 또한 인지기능 장애는 방치할 시 치매로 이어져[1] 수발 가족에게 정신적, 경제적 부담을 안기기에 환자뿐만 아니라 가족까지 고통받는 질병이다. 그러므로 인지기능 장애의 정기 검진을 통한 조기 진단과 적극적 대처가 중요하다[2]. 이에 MMSE(Mini-Mental State Examination, 간이정신상태 검사) 등 환자의 인지기능 이상 유무를 선별하기 위한 도구들이 개발되었다. 그러나 이러한 검사는 환자의 현상태에 대한 진단만 가능할 뿐 환자의 생활습관과 관계된 문제점을 발견하거나 환자의 행동 및 생활환경 개선 등에 활용하기에는 제한이 있다.

한편 인지기능 장애와 신체활동(운동) 및 수면은 밀접한 연관이 있는 것으로 연구되어왔다[3][4]. 최근 웨어러블 디바이스 기술의 발전으로 활동량이나, 수면 정보, 체중 변화, 체질량 등 사용자의 다양한 라이프로그를 수집할 수 있게 되었다. 이에 본 연구에서는 신체활동과 라이프로그 데이터를 활용하여 인지기능 장애 여부를 예측하는 모형을 개발하고자 한다. 최근 인지기능 장애 예측을 목적으로 여러 선행연구에서 머신러닝 기법을 활용한 바 있다[5]-[9]. 본 연구에서도 머신러닝 알고리즘을 기반으로 예측 모델을 제안할 것이다. 또한 제안한 모델의 주요 변수를 도출하여 인지기능 장애에 영향을 미치는 주요 요인들을 식별하고자 한다.

본 연구의 구성은 다음과 같다. 2장에서는 본 연구와 관련된 선행연구 사례들을 소개하고, 3장에서는 본 연구의 접근방법에 대하여 설명한다. 4장에서는 연구 결과를 설명하며, 5장에서 결론과 함께 시사점을 제시하고자 한다.

II. 선행연구

2.1 머신러닝 기반 인지기능 예측 연구

인지기능의 저하는 치매의 직접적인 위험요인으로 알려져있다[1]. 정상과 치매의 중간단계인 인지

기능 장애(MCI, Mild Cognitive Impairment)의 경우 일상생활에는 큰 어려움이 없으나 동년배의 정상 노인 대비 치매 진단율이 적게는 5배에서 많게는 15배에 이른다[4]. 이에 인지기능 장애를 예측하고 인지기능 장애와 관련된 요인들을 찾기 위한 시도들이 있어왔으며 최근에는 머신러닝 기법을 활용한 연구들이 늘고 있다[5]-[8].

경기도 광주시 65세 이상 인지기능 장애 경험 노인을 대상으로 시행된 설문조사를 활용한 선행연구에서는 인지기능 변화곡선 기울기값을 근거로 인지기능저하를 예측하는 의사결정나무 및 로지스틱 회귀모델을 제안하였다[6]. 해당 연구의 모형은 특이도 92.5%, 민감도 69.5%의 성능을 보였으며 가구원수와 신체활동, 성별 및 주관적 기억력을 주요 예측인자로 식별하였다.

한국고용정보원에서 시행하는 고령화연구패널조사 자료를 활용한 연구에서는 로지스틱 회귀분석과 의사결정나무분석을 기반으로 45세 이상 중노년층의 경도인지장애를 예측하는 모델을 제안하였다[7]. 모델의 정확도는 80.5%였는데 연령과 교육수준 변수가 모델의 성능에 큰 영향을 미친 것으로 나타났다.

보건사회연구에서 수집한 공공데이터인 2020년 노인실태조사를 활용한 연구에서는 의사결정나무를 활용하여 65세 이상 노인인구의 인지기능 차이를 예측하는 모델을 제안하였다[8]. 모델의 정확도는 68.6%였으며 흡연, 운동, 식이 및 인간관계 등이 인지 기능에 유의미한 영향을 미치는 것을 확인하였다.

이상에서와 같이 인지기능을 예측하는 머신러닝 기반의 연구들은 주로 설문조사 결과를 데이터로 활용한 연구들이 대부분이다. 설문조사의 경우 피 설문자들의 주관적인 응답을 기준으로 작성되어 자료의 정확성과 신뢰도에 제한점이 있다는 점은 한계점으로 지적된다[8]. 또한 결과의 해석을 위해 불가피하게 의사결정나무 모형을 사용함에 따라 예측 성능 측면에서 개선의 여지가 있다. 이에 본 연구에서는 대상자의 실제 생활 정보를 객관적으로 수집한 라이프로그 데이터를 기반으로 인지기능 장애를 예측하고자 한다. 그리고 머신러닝 알고리즘을 활용하여 인지기능 예측에 최적의 모델을 구축하고 XAI(eXplainable AI) 기법을 활용함으로써 인지기능 장애에 영향을 미치는 주요 요인들을 식별하고자 한다.

2.2 라이프로그 분석에 사용한 알고리즘

본 연구에서는 라이프로그 데이터를 분석하기 위하여 머신러닝 알고리즘 중 Random Forest 알고리즘을 사용한다. Random Forest는 앙상블 학습 방법으로서 복수 개의 의사결정나무를 생성하고 각 의사결정나무가 개별적으로 학습을 수행한 뒤 결과를 취합하여 결정하는 방식이다[10]. 일반적으로 의사결정나무를 이용하여 학습하게 되면 결과해석이 용이한 장점은 있으나 성능의 변동 폭이 크고 과적합의 우려가 크다는 결점을 가진다. 반면 Random Forest 기법에서는 여러 의사결정나무를 만들어 투표(Voting)를 통하여 최종 결과값을 도출함으로써 과적합을 방지하며 안정적이고 일반화된 결과를 얻을 수 있다. 또한 단일 의사결정나무 모델 대비 대체로 우수한 성능을 보인다.

한편 최근 활발히 활용되고 있는 XAI는 머신러닝 모델이 제시한 결과에 대한 이유를 인간이 이해할 수 있는 방식으로 제시하는 방법론이다. SHAP 기법은 대표적인 XAI 기법 중 하나로써 Shapley Value를 기반으로 모형의 예측 결과에 대한 변수들의 기여도를 보여준다[11]. Shapley Value는 변수들의 조합을 구성하고 해당 변수의 유무에 따른 평균적 변화를 통하여 값을 계산하며 예측 모형의 종류와 관계없이 계산 가능하다. 본 연구에서는 Random Forest 기반의 예측 모델을 구축한 후 사후분석을 위하여 SHAP 기법을 활용함으로써 인지기능 장애 예측에 중요한 변수들을 식별하고자 한다.

III. 연구 방법

3.1 데이터 수집

본 연구에서는 AI-Hub(<https://www.aihub.or.kr/>)에서 제공하는 데이터를 분석 대상으로 한다. AI-Hub는 한국지능정보사회진흥원이 운영하는 AI 통합 플랫폼으로서 AI 서비스 개발을 지원하고자 헬스케어 관련 포함 여섯 개 분야 학습용 데이터를 제공하고 있다. 본 연구에서는 그 중 “치매 고위험군 웨어러블 라이프로그”라는 제목의 데이터셋을 활용한다. 해당 데이터셋은 전문의의 정밀 진단을 바탕으로

선정한 광주광역시 55세 이상의 정상인, 경도인지장애 및 치매 환자 등 총 300명에게 반지 형태의 웨어러블 디바이스를 착용시켜 수집된 데이터로 구축되었다. 대상자 선정 과정에서 병리진단 기반(아밀로이드 PET 및 전문 임상 의 진단) 알츠하이머병(AD) 고위험군을 우선 연구대상자로 선정하였다. 또한 데이터 수집 후 장비(반지) 착용 기간이 35일 미만인 연구 대상자를 제외하는 등 검수를 진행하고 비식별화 및 정규화를 진행하였다. 최종 배포된 데이터셋에는 인지기능 정상(111건), 경도인지장애(51건) 및 치매(12건) 단계의 총 174명의 대상자에 대한 정보가 기록되어 있다.

데이터셋은 라이프로그 데이터와 간이인지검사(MMSE) 데이터로 구분된다. 먼저 라이프로그 데이터는 착용자의 수면 데이터(수면시작 및 종료시간까지의 혈압, 심장 박동수, 심박 변이도, 호흡, 체온, 렘수면시간 등)와 활동 데이터(수면 이외의 활동 시간에 대한 움직임, 심장 박동수, 심박 변이도, 칼로리 소모량 등)를 최소 35일부터 최대 120일 동안에 걸쳐 추적한 24시간 라이프로그 정보가 총 12,183건 담겨있다. 데이터 유형에 따라 값의 형태가 다른데 예를 들어 수면시작시간 데이터는 타임스탬프 형태로 측정되어 있었고 수면시간 데이터는 수치값 형태로, 수면깊이 데이터는 0부터 5까지 정수값으로 기록되어 있었다.

다음으로 간이인지검사(MMSE) 데이터는 연구 대상자의 기억력, 언어기능 등 특정 기능 장애와 뇌영역 중 손상 영역을 파악하는 19개의 문항과 하위 문항으로 구성된 설문 기록으로 이루어져 있다. 여기에는 연구대상자가 설문에 대하여 제대로 수행하였으면 1점(혹은 2점), 수행하지 못하면 0점(혹은 1점)을 부여한 정보가 담겨있다.

3.2 데이터 전처리 및 변수 추출

수집된 데이터는 174명을 대상으로 하며 대상자별 인지기능 장애 여부를 분류하는 것이 본 연구의 목표이다. 따라서 대상자별 복수 개의 값을 갖는 라이프로그 데이터들은 각 측정 항목의 대표성을 유지하면서 값을 요약할 필요가 있었다. 예를 들어 수면깊이 데이터는 수면 기간 중 0부터 5까지의 정수

값이 5분 단위로 측정되어 있으므로 0 이하, 1 이하, 2 이하, 3 이상, 4 이상, 5 이상의 수치가 나온 비율과 개수를 변수화하였다. 일별 수면시간과 같이 측정일만큼의 데이터가 있는 경우 최소, 최대, 평균, 표준편차, 중간값을 변수화하였다. 수면시작시간과 같은 타임스탬프 데이터는 ‘2020-12-26T03:59:59+09:00’와 같은 표기형식을 따르고 있었는데 시간 정보만을 추출하여 0에서 24 사이의 실수값을 갖도록 변환하고 측정일 중의 최소, 최대, 평균, 표준편차, 중간값을 변수화하였다. 그림 1은 전처리후의 데이터 샘플을 보여준다.

예측하고자 하는 진단결과 변수는 연구 대상자의 진단결과 CN(정상), MCI(인지기능 장애), Dem(치매)의 세 가지 중 하나의 값을 가진다. 그러나 정상으로 분류된 데이터가 과반 이상이고 본 연구의 목적이 인지기능 장애 여부를 예측하는 것이므로 인지기능 장애의 심화단계라고 할 수 있는 치매를 인지기능 장애와 통합하여 처리하였다. 결과적으로 CN(정상) 111건과 MCI(인지기능 장애) 63건으로 진단결과 변수를 구성하였다.

최종 전처리된 데이터셋은 174행 330열의 크기를 가진다. 174행은 연구 대상자의 수와 일치하며, 330열에는 검사 기간의 라이프로그가 전처리된 데이터와 연구 대상자별 간이인지검사 결과 항목을 포함한다.

3.3 분류 모델 구축과 최적화

모델 구축에 앞서 계층적 샘플링(Stratified shuffle split) 기법을 사용하여 각 클래스에서 균등하게 샘플을 추출하였고, 훈련 데이터와 테스트 데이터를 7:3 비율로 분배하였다. 정상 데이터 건수가 인지기능 장애에 비해 크기 때문에 훈련 데이터에 SMOTE(Synthetic Minority Oversampling TEchnique)

오버샘플링을 진행하여 클래스 간 불균형 문제를 해소하였다.

독립변수에 대해서는 스케일링 적용에 따른 모델 성능 영향을 확인하고자 총 다섯 가지 종류의 스케일러를 적용하여 모델을 만들었다. 스케일러를 적용하지 않는 모델과 Standard Scaler, MinMaxScaler, MaxAbs Scaler, Robust Scaler를 각각 적용한 총 다섯 가지 종류의 모델을 만들었다.

다음으로, Random Forest 하이퍼파라미터 튜닝을 통하여 모델을 최적화하였다. 이를 위해 훈련 데이터를 계층적 K겹 교차검증(Stratified K-Fold) 기법을 사용하여 4개의 학습 폴드와 1개의 검증 폴드로 나누었고, 검증 폴드를 바꿔가며 5개의 데이터 세트를 생성하였다. 이후 4개의 학습 폴드로 모델을 훈련시키고 1개의 검증 폴드로 검증하는 작업을 세트마다 진행하여 5개의 교차검증에서 정확도의 평균값이 가장 높았을 때의 하이퍼파라미터값을 탐색하였다.

먼저, 최적 트리 개수를 찾기 위해 정확도값을 기준으로 최적 트리 개수를 10개부터 400개까지, 최대 깊이 값은 2부터 20까지 늘려가면서 Random Forest 모델에서 최적의 성능을 보이는 하이퍼파라미터값을 찾았다. 탐색 시 Random Forest 모델의 분류 기준은 엔트로피값으로 하였다. 모델 간 성능 비교는 정확도 값을 근거로 판단하였으며 정확도 값이 같을 경우 AUC 값을 근거로 판단하였다. 결과적으로 스케일링을 적용하지 않고, 표 1과 같이 하이퍼파라미터를 설정하였을 때 가장 우수한 성능을 보였다.

표 1. Random forest 모델의 최적 하이퍼파라미터 값
Table 1. Optimized hyperparameter values of the random forest model

Parameter	Value	Parameter	Value
n_estimators	75	min_samples_leaf	4
max_depth	4	min_samples_split	9
max_features	0.2	max_samples	0.8

	mean_activity_rest	mean_activity_high	mean_activity_score	mean_sleep_bedtime_start	mean_sleep_deep	mean_sleep_total	Q01	Q02	Q03
0	618.063492	0.793651	58.682540	20.645847	8256.666667	23096.190476	1	2	1
1	521.247788	0.256637	54.522124	18.934142	4681.592920	21382.300885	2	2	2
2	434.086207	0.637931	65.758621	20.195814	3031.034483	26979.827586	1	2	2
3	556.114754	0.836066	91.590164	15.525387	7519.180328	23163.196721	2	2	2

그림 1. 전처리후 데이터 샘플
Fig. 1. Examples of the preprocessed data

이후 후진 제거법(Backward elimination)을 활용하여 모델에 사용하는 변수를 최적화하였다. 이를 위하여 모델의 변수중요도를 계산하고 변수중요도가 낮은 변수부터 30개씩 제거하면서 모델을 다시 학습하는 작업을 반복하였다. 변수중요도 기준별로 생성한 모델들의 성능 비교는 표 2에서 보는 바와 같다.

변수중요도 기준 상위 30개 변수를 사용할 때 정확도 79.2%, AUC 0.797로서 가장 우수한 성능을 보였다. 이에 해당 모델을 최종 모델로 선정하였다.

표 2. 후진 제거법 기반의 변수 최적화
Table 2. Backward elimination based feature optimization

Number of features	Accuracy	AUC
329	0.7547	0.7352
300	0.7547	0.7770
270	0.7735	0.7616
240	0.7735	0.7863
210	0.7169	0.7244
180	0.7547	0.7322
150	0.7358	0.7832
120	0.7358	0.7801
90	0.7169	0.7894
60	0.7547	0.7678
30	0.7924	0.7972
20	0.7735	0.7585
10	0.7924	0.7786
5	0.6415	0.5557

3.4 성능 비교를 위한 추가 모델 구축

본 연구에서 제시한 모델의 성능에 대한 객관적인 비교를 위하여 다른 알고리즘을 사용한 모델을 추가적으로 구축하였다. 과적합에 강하고 작은 크기의 데이터셋에서 좋은 성능을 보이는 것으로 알려진 SVM(Support Vector Machine)과 부스팅 계열에서 최근 좋은 성능을 보이는 것으로 널리 알려진 XGBoost(Extreme Gradient Boosting) 알고리즘을 활용하여 모델 구축 및 최적화 과정을 진행하였다.

표 3. SVM 모델의 최적 하이퍼파라미터 값
Table 3. Optimized hyperparameter values of the SVM model

Parameter	Value	Parameter	Value
C	1000	gamma	0.01
kernel	rbf		

SVM 모델과 XGBoost 모델의 최종 하이퍼파라미터 값은 각각 표 3과 표 4에서 보는 바와 같다.

표 4. XGBoost 모델의 최적 하이퍼파라미터 값
Table 4. Optimized hyperparameter values of the XGBoost model

Parameter	Value	Parameter	Value
booster	gbtree	colsample_bytree	1
min_child_weight	1	eval_metric	logloss
n_estimators	25	gamma	0
objective	binary:logistic	learning_rate	0.1
reg_alpha	0	max_depth	6
reg_lambda	1	scale_pos_weight	1
subsample	1		

IV. 연구 결과

본 연구에서 최종 제안한 모델의 성능에 대한 평가 지표는 정확도와 정밀도, 재현율 및 F1값을 사용하였다. 표 5에서 보는 바와 같이 모델의 정확도는 Random Forest 모델이 SVM 및 XGboost 모델 대비 우수한 성능을 나타냈으며 정확도는 79.3%, 정밀도와 재현율은 각각 78%와 75.7%였다.

표 5. 모델 간 성능 비교
Table 5. Performance comparison among the models

model	Accuracy	Precision	Recall	F1 score
RF	0.7924	0.7804	0.7570	0.7654
SVM	0.7169	0.7677	0.6168	0.6074
XGboost	0.7429	0.7500	0.4615	0.5714

최종 모델에 사용된 변수중요도 기준 상위 20개 변수들의 결과에 대한 각각의 영향 정도를 파악하기 위하여 변수 중요도를 분석하였다. 각 변수들의 중요도에 따라 그래프로 나타낸 결과는 그림 2에서 보는 바와 같다. 20개 변수는 “오늘이 며칠입니까?” 등 인지기능검사 문항 4개(검은색)와 수면 중간점 시간 델타값의 평균 등 수면 관련 변수 14개(파란색), 저장도 활동 시간의 표준편차 등 활동 관련 변수 2개(빨간색)로 구성되어 있었다. 수면과 신체 활동 관련 데이터가 인지기능 장애 예측에 중요한 비중을 차지하는 것을 확인할 수 있다. 각 변수들에 대한 목록은 표 6에서 확인할 수 있다.

표 6. 변수중요도 목록

Table 6. Feature importance list

Variable	Type	Description
Q03	C	What day is today?
Q13_2	C	The second correct answer out of three object names memorized before the test
TOTAL	C	Total score of cognitive function test
mean_sleep_midpoint_at_delta	S	Average sleep midpoint time delta
std_activity_low	A	Standard deviation of low-intensity activity time
mean_sleep_score	S	Mean value of sleep quality score
min_sleep_bedtime_start	S	Minimum value of sleep onset time
min_sleep_total	S	Minimum value of sleep duration
mean_sleep_score_total	S	Mean value of sleep duration contribution score
std_sleep_rem	S	Standard deviation of REM sleep duration
mean_sleep_period_id	S	Mean value of sleep identification ID
median_sleep_rmssd	S	Median value of average heart rate variability
mean_sleep_rmssd	S	Average of average heart rate variability
std_sleep_rmssd	S	Standard deviation of average heart rate variability

std_sleep_light	S	Standard deviation of light sleep duration
Q13_3	C	The last correct answer out of three object names memorized before the test
median_sleep_light	S	Median value of light sleep time
min_sleep_score_latency	S	Minimum value of sleep latency score
std_activity_steps	A	Standard deviation of daily step count
min_sleep_score	S	Minimum value of sleep quality score

S: Sleep, C: Cognitive function, A: Activity

변수들에 대한 보다 자세한 분석을 위하여 SHAP(SHapley Additive exPlanation) 분석을 진행하였다. SHAP 분석 결과는 그림 3에서 보는 바와 같다. X축은 Shapley Value 값으로서 특정 변수의 Shapley Value 분포 범위가 오른쪽으로 넓을수록 목표변수인 인지기능 장애 변수에 대한 양(+)의 영향력이 크고 왼쪽으로 넓을수록 음(-)의 영향력이 큰 것으로 해석할 수 있다. 이 때 색상은 빨간색에 가까울수록 해당 변수의 값이 크고 파란색에 가까울수록 작은 것을 의미한다. 전반적으로 특정 변수의 색상 구분이 선명할수록 해당 변수가 예측에 유용한 것으로 볼 수 있다.

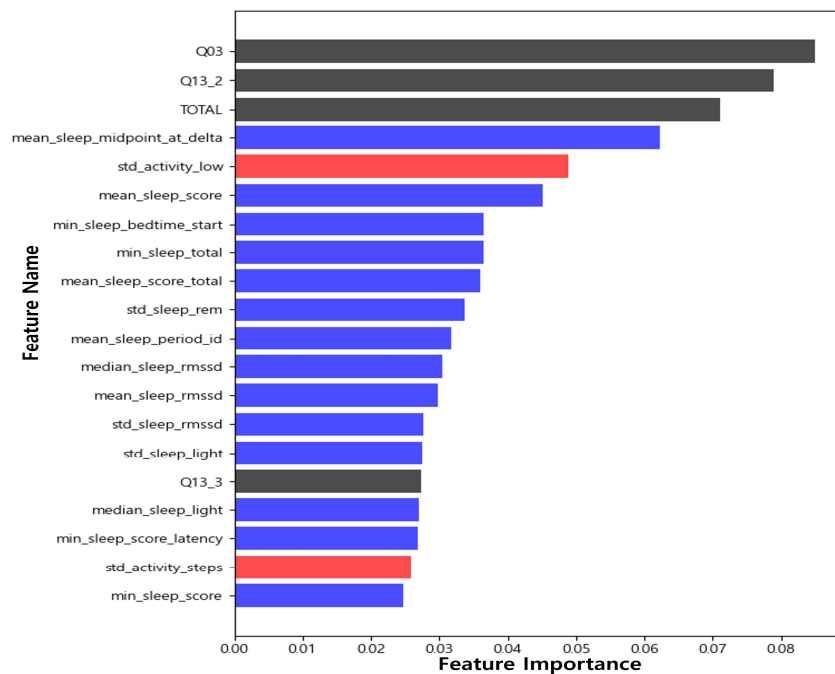


그림 2. 변수중요도 분석 결과
Fig. 2. Result of feature importance analysis

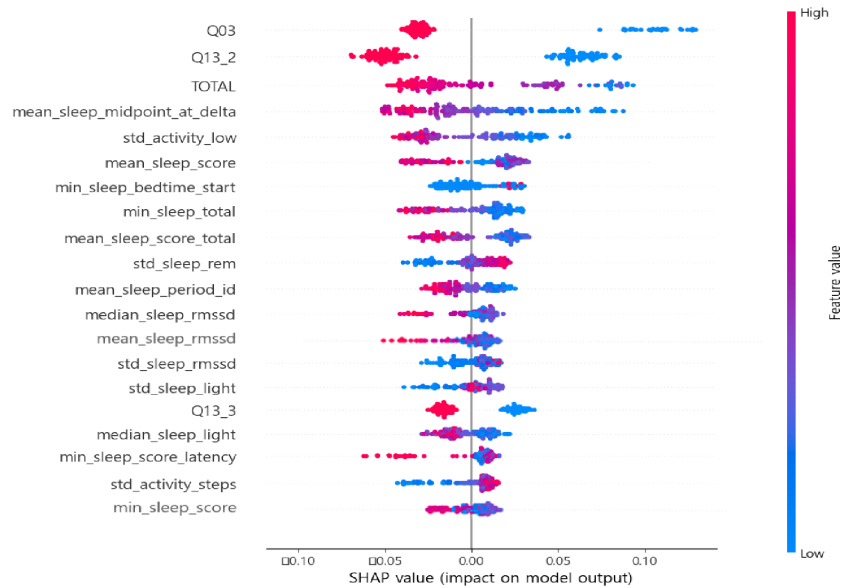


그림 3. SHAP 분석 결과
Fig. 3. Result of SHAP analysis

결과에 따르면 수면 관련 변수 중에서는 수면 중간점 시간 델타값의 평균이 낮을수록, 수면 종합 점수의 평균이 낮을수록, 수면 시작시간의 평균이 빠를수록 인지 기능 장애 위험이 큰 것으로 나타났다. 활동 관련 변수 중에는 저항도 활동 시간의 표준편차가 높을수록, 매일 걸음 수의 표준편차가 높을수록 인지기능 장애 위험이 적은 것으로 나타났다.

V. 결론 및 시사점

본 연구에서는 라이프로그 데이터를 기반으로 정상과 인지기능 장애를 분류할 수 있는 예측 모델을 제시하였다. 또한 사후분석을 통하여 수면 및 신체 활동 관련 데이터가 인지기능 장애를 판단하는 중요 변수가 됨을 확인하였으며 이러한 결과는 설문 기반으로 진행된 과거 선행연구들과 달리 웨어러블 디바이스를 통해 수집된 데이터를 정량적으로 분석한 결과라는 점에서 의미를 갖는다. 본 연구에서 제안한 모델을 웨어러블 디바이스에 적용한다면 검사 비용과 시간을 절약할 수 있으며 인지기능 장애 조기 진단에도 유용할 것이다. 또한 수면과 신체 활동 등 생활 습관 측면에서 개선방안을 안내함으로써 피검사자 스스로 인지기능 장애예방을 위한 생활 개선방안을 쉽게 실천할 수 있다는 장점이 있다.

본 연구에서는 정부 예산으로 구축된 공개 데이

터를 활용하였다는 점에서 의미가 있다. 그러나 생체정보를 장기간에 걸쳐 수집해야 하는 어려움으로 인해 최초 데이터 구축시 수집대상 인원의 수를 늘리는데 한계가 있었던 것으로 보인다. 이처럼 샘플의 수가 적다는 것은 본 연구의 제한점이자 추후 더 많은 인원을 대상으로 후속연구의 여지를 남긴다. 또한 본 연구는 이미 정상과 인지기능 장애로 각각 판정된 사람들을 대상으로 수집된 데이터를 활용한 횡단연구이지만 향후 정상인이 인지기능이 저하되는 과정을 추적하며 수집한 데이터를 기반으로 종단연구가 진행된다면 인지기능장애의 발병과정 과정에 대한 더욱 의미있는 시사점을 얻을 수 있을 것으로 판단된다.

References

- [1] T. C. Peng, W. L. Chen, L. W. Wu, Y. W. Chang, and T. W. Kao, "Sarcopenia and Cognitive Impairment: A Systematic Review and Meta-Analysis", *Clinical Nutrition*, Vol. 39, No. 9, pp. 2695-2701, Sep. 2020. <https://doi.org/10.1016/j.clnu.2019.12.014>.
- [2] O. Yanhong, M. Chandra, and D. Venkatsh, "Mild Cognitive Impairment in Adult: A Neuropsychological Review", *Annals of Indian*

Academy of Neurology, Vol. 16, No. 3, pp. 310-318, Jul. 2013. <https://doi.org/10.4103/0972-2327.116907>.

[3] S. J. Lee, "Effects of Exercise Program on Cognitive Function and Balance in the Elderly with Dementia Through Meta-analysis", The Journal of Humanities and Social Science, Vol. 12, No. 5, pp. 425-436, Oct. 2021. <http://dx.doi.org/10.22143/HSS21.12.5.30>.

[4] O. M. Bubu, et al., "Sleep, Cognitive Impairment, and Alzheimer's Disease: A Systematic Review and Meta-analysis", Sleep, Vol. 40, No. 1, Jan. 2017. <https://doi.org/10.1093/sleep/zsw032>.

[5] W. Li, L. Zeng, S. Yuan, Y. Shang, W. Zhuang, Z. Chen, and J. Lyu, "Machine learning for the prediction of cognitive impairment in older adults", Frontiers in Neuroscience, Vol. 17, Apr. 2023. <https://doi.org/10.3389/fnins.2023.1158141>.

[6] J. H. Back, "Prediction Model for Cognitive Decline in Older Adults Living in the Community", Ajou University, Doctoral Dissertation, Jan. 2013.

[7] H. Park and J. Ha, "Prediction Models of Mild Cognitive Impairment Using the Korea Longitudinal Study of Ageing", Journal of Korean Academy Nursing, Vol. 50, No. 2, pp. 191-199, Apr. 2020. <https://doi.org/10.4040/jkan.2020.50.2.191>.

[8] M. Han, "Cognitive Function Prediction Model for the Elderly using Decision Tree Analysis: 2020 National Survey of the Living Condition and Welfare Needs of Older Koreans", Journal of the Korea Academia-Industrial cooperation Society, Vol. 23, No. 9, pp. 462-475, Sep. 2022. <https://doi.org/10.5762/KAIS.2022.23.9.462>.

[9] K. K. Yeo and D. S. Kang, "CNN-based Alzheimer's Disease Image Learning and Accuracy Measurement Using F-Florbetaben PET Image Editted Database", Journal of Korean Institute of Information Technology, Vol. 15, No. 10, pp. 1-8, Oct. 2017. <https://doi.org/10.14801/jkiit.2017.15.10.1>.

[10] I. H. Sarker, "Machine Learning: Algorithms, Real-

world Applications and Research Directions", SN Computer Science, Vol. 2, No. 160, pp. 1-21, Mar. 2021. <https://doi.org/10.1007/s42979-021-00592-x>.

[11] S. S. Fatima, M. Wooldridge, and N. R. Jennings, "A Linear Approximation Method for the Shapley Value", Artificial Intelligence, Vol. 172, No. 14, pp. 1673-1699, Sep. 2008. <https://doi.org/10.1016/j.artint.2008.05.003>.

저자소개

이 명 진 (Myeongjin Lee)



2017년 3월 ~ 현재 : 명지대학교
경영정보학과 학사과정
2023년 8월 ~ 현재 : AK아이에스
사원
관심분야 : RPA, 데이터 분석,
머신러닝, 딥러닝

이 종 언 (Jongun Lee)



2017년 3월 ~ 현재 : 명지대학교
일어일문학과 학사과정
관심분야 : 머신러닝, SAP ERP
운영, ABAP 개발, UCC콘텐츠
제작

이 한 준 (Hanjun Lee)



2001년 2월 : 서울대학교
컴퓨터공학과(공학사)
2004년 2월 : 서울대학교
컴퓨터공학과(공학석사)
2016년 8월 : 고려대학교 경영학과
MIS 전공(경영학박사)
2007년 7월 ~ 2018년 2월 :

한국국방연구원 선임연구원

2018년 3월 ~ 2020년 2월 : 한남대학교 조교수

2020년 3월 ~ 현재 : 명지대학교 부교수

관심분야 : 머신러닝, 자연어 처리, 정보시스템, 정보화 정책