

## 중고령 노인의 개인적 가치에 따른 라이프스타일 분류: 머신러닝을 활용한 상대적 중요도 분석

임승주\*, 박지혁\*\*

\*연세대학교 일반대학원 작업치료학과 박사과정 학생

\*\*연세대학교 소프트웨어디지털헬스케어융합대학 작업치료학과 교수

### 국문초록

**목적** : 노인의 건강한 삶의 방식으로서 라이프스타일에 대한 연구가 증가하고 있다. 라이프스타일이 개인의 가치와 삶의 태도를 반영하는 개념임에도 불구하고, 아직까지 개인의 어떠한 가치가 라이프스타일을 건강하게 유도하는지 파악한 연구는 부족한 실정이다. 이에 본 연구는 노인의 라이프스타일 유형을 두 가지로 분류하고, 머신러닝을 활용하여 어떠한 개인적 가치가 건강한 라이프스타일에 우선적으로 작용하는지 파악하고자 한다.

**연구방법** : 본 연구는 지역사회에 거주하는 55세 이상 중고령 노인 300명을 대상으로 횡단 연구를 수행하였다. 라이프스타일은 Yonsei Lifestyle Profile-Active, Balanced, Connected, Diverse (YLP-ABCD) 응답을 사용하여 잠재프로파일 분석을 통해 유형화하였다. 라이프스타일 유형을 예측하는 개인적 가치는 YLP-V (Values) 응답을 수집하여, 예측성능이 가장 높은 머신러닝 알고리즘을 선정 후 상대적 중요도를 파악하였다.

**결과** : 잠재프로파일 분석 결과, 라이프스타일은 건강한 라이프스타일 실천형(48.87%), 비실천형(51.13%)으로 분류되었다. 실천형에 속한 중고령 노인은 비실천형에 비해 사회관계가 활발한 특성을 나타내었다. 본 연구에 포함된 머신러닝 알고리즘 중 가장 우수한 성능을 보인 모델은 서포트 벡터 머신으로, 정확도 96%, Receiver Operating Characteristic (ROC) 영역 95%로 나타났다. 본 알고리즘을 바탕으로 개인적 가치의 상대적 중요도를 분석한 결과, 건강한 식단, 건강 매체, 여가활동, 건강 제품 및 머신러닝에 주의를 기울일수록, 해당 가치에 따라 중고령 노인은 건강한 라이프스타일을 실천하는 그룹에 속할 가능성이 큰 것으로 나타났다.

**결론** : 본 연구는 중고령 노인의 사회적 관계망을 포함한 건강한 라이프스타일을 유도하기 위해, 건강 식단, 매체, 여가, 제품 및 습관에 대한 가치 향상을 중점적으로 다루는 종합적인 프로그램 및 서비스의 필요성을 시사한다.

**주제어** : 개인적 가치, 노인, 라이프스타일, 머신러닝, 상대적 중요도

## I. 서론

노인의 건강 관리에 있어 치료적 접근보다 예방적 접근의 중요성이 증대되며(Michel & Sadana, 2017), 예방적 관점에서 노인의 건강을 관리하는 전략으로서 건강한 라이프스타일에 따른 효과가 다수 보고되고 있다(Ford et al., 2012; Loef & Walach, 2012). 건강한 라이프스타일이란 건강생활을 실천하는 습관으로(Lim & Park, 2020), 성공적 노화에 있어 변경 가능한 요인이자 융합적 건강 개념으로서 인식되고 있다(Depp & Jeste, 2006; Polidori et al., 2010). 라이프스타일은 한 가지 구성요소가 아닌 개인의 복합적인 건강 관련 습관으로 보고된다(Spring et al., 2012). 체계적 고찰 및 선행연구들에 따르면 신체활동(Sun et al., 2010), 건강한 식습관(Hu et al., 2001; Knoops et al., 2004), 활발한 사회적 관계망(Chen & Feeley, 2014; Hrubá & Sadilek, 2022), 활동참여(Cudjoe et al., 2020)는 라이프스타일의 주요한 핵심영역으로 종합되었다.

국내 지역사회 노인을 대상으로 라이프스타일을 유형화한 선행연구(Park, Yang, Won & Park, 2021)에 따르면, 신체활동의 경우 수동적/저강도/균형적 신체활동형으로 분류되고, 식습관의 경우 영양 불균형형과 균형형으로 분류되며, 활동참여의 경우 수동적/일상유지적/균형적 참여형으로 분류됨을 잠재프로파일 분석을 통해 밝혔다. 하지만, 해당 결과는 라이프스타일이 다면적, 복합적 개념임에도 불구하고 각 영역별로 나누어 유형화한 결과라는 점에서, 라이프스타일을 하나의 개념으로서 유형화하지 못했다는 한계가 있다. 라이프스타일의 전반적인 영역을 종합하여 분류한 선행연구(Park, Yoo, Kim, Hong, Lee & Park, 2021)가 존재하나, 해당 연구는 라이프스타일의 영역으로서 사회관계를 포함하지 않았다는 한계가 있다. 이에, 라이프스타일을 다면적으로 분류할 수 있는 유형화 연구가 필요하다.

머신러닝은 집단을 분류하고 예측하는 데에 활용되는 최신 기법으로, 데이터로부터 학습하고 패턴을 인식하여 예측, 분류, 인식 등의 작업을 수행할 수 있는 인공

지능의 한 분야이다(Zhou, 2021). 라이프스타일의 분류에 있어 여러 통계 방법들 가운데 머신러닝을 적용하여 유형화를 시도하는 연구들이 활발히 이루어지는 추세로서, 주로 로지스틱 회귀 모델(Logistic Regression: LR), 의사결정 알고리즘(Decision Tree algorithm: DT), 랜덤 포레스트(Random Forest: RF), 서포트 벡터 머신 알고리즘(Support Vector Machine, SVM) (Huang et al., 2022)이 노인의 라이프스타일을 효과적으로 분석하고 그룹화하는 데에 유용하다는 것을 보여주었다. 하지만, 머신러닝을 활용한 연구 결과들에서 높은 분류 성능을 보고하였음에도 불구하고, 결과의 해석이 특정 분포, 인종 및 민족의 일반화 우려가 있는 것으로 한계점을 밝혔다(Wang et al., 2023). 따라서 국내 지역사회 노인 인구 집단을 대상으로 분류 모델을 적용하기 위해서는, 국내 노인인구를 대상으로 다면적 라이프스타일을 분류할 수 있는 모형이 필요할 것이다.

또한, 노인의 라이프스타일을 분류함에 있어 개인의 가치에 기반한 관심사, 중요하다고 생각하는 활동 등의 영향을 고려할 필요가 있다. 선행연구에 따르면(Lim et al., 2023; Spring et al., 2012), 개인의 라이프스타일은 인간의 가치, 의견, 관심, 항목 등 맥락적 요인의 영향을 받는 것으로 보고하였다. 특히, 개인별 건강한 라이프스타일을 실천하는 데 있어 개인적 가치관과 관심사는 주요한 배경과 기반이 될 것으로(Ahmed et al., 2020; Lim et al., 2023), 노인집단의 건강한 라이프스타일 실천 유형을 분류하고 각 유형별 개인적 가치에 기반한 특성에 따라 맞춤형 라이프스타일 관리 전략을 제시할 수 있을 것이다. 하지만, 현재까지 국내 노인이 보유한 특정 가치가 라이프스타일에 미치는 영향을 탐구한 연구가 이루어지지 않았다. 이로써 국내 노인들의 개인적 가치와 라이프스타일 간의 관련성에 대한 이해가 부족한 상태이다. 이에 따라 본 연구는 노인의 라이프스타일 유형을 분류하고, 머신러닝을 활용하여 어떠한 개인적 가치가 건강한 라이프스타일 유형으로 분류되는지 파악하고자 한다.

## II. 연구 방법

### 1. 연구 대상

본 연구는 55세 이상인 자, 지역사회에 거주하는 자, 온라인 설문 조사가 가능한 자를 대상으로 하였다. 통상적으로 고령자의 기준을 65세 이상으로 설정하나, 중고령자의 라이프스타일 및 건강 관련 선행연구들에서는 중고령자를 55세 이상으로 분류하고 있다(Park, Yang, Won & Park, 2021; Park, Yoo, Kim, Hong, Lee & Park, 2021; Park et al., 2021b; Yang et al., 2020). 글을 읽고 이해하는데 어려움이 있는 자는 제외하여, 총 300명이 설문에 참여하였다.

### 2. 자료 수집

설문 기간은 2022년 12월 1일부터 2023년 1월 31일 로 2개월에 걸쳐 진행되었다. 코로나19의 영향으로 설문 조사의 어려움을 감안하여 전문 조사기관인 엠브레인에 조사를 의뢰하여 조사하였다. 일반적 특성으로서 대상자의 연령, 성별, 교육 수준, 거주 지역, 경제 수준, 결혼 여부, 주관적 건강 상태에 대한 응답을 수집하였다. 이어 라이프스타일과 개인적 가치를 파악하는 설문 응답을 수집하였다. 설문 결과, 총 300명의 유효한 데이터가 회수되었다.

### 3. 연구 도구

#### 1) 종속변수 : 라이프스타일 유형

고령자의 라이프스타일을 파악하고자 Yonsei Lifestyle Profile-Active, Balanced, Connected, Diverse (YLP-ABCD) 설문 도구를 사용하였다(Park et al., 2023). 본 설문 도구는 노인의 라이프스타일을 다면적으로 파악하기 위해 개발되었다. 설문 영역은 신체활동, 식습관, 사회관계, 활동참여로 구분되며, 각 영역은 긍정 영역과 부정 영역으로 나누어져, 총 8개의 영역과 72개의

문항으로 구성된다. 각 문항은 5점 리커트 척도(1점: 매우 그렇지 않다, 5점: 매우 그렇다)로 평가되며, 영역별 평균치를 계산하여 결과를 해석한다. 평균치가 높을수록 긍정 영역은 해당 영역이 건강함을, 부정 영역은 건강하지 않음을 의미한다. 본 설문 도구의 Cronbach's  $\alpha$ 는 .631로 나타났으며, 검사-재검사 신뢰도는 약 .692로 나타났다.

#### 2) 예측변수 : 개인적 가치

고령자의 개인적 가치를 파악하고자 Yonsei Lifestyle Profile-Values (YLP-V) 설문 도구를 사용하였다(Lim et al., 2023). 본 설문 도구는 Activity, Interest and Opinion (AIO) 기법(Plummer, 1974)을 적용하여, 활동(A), 관심사(I), 의견(O) 영역을 중심으로 노인의 라이프스타일에 대한 태도와 가치관을 평가하는 설문 도구로서 개발되었다. 평가도구는 총 3개의 영역과 24개의 문항으로 구성되어 있다. 개인이 선호하는 활동 영역 7개 문항, 개인의 관심사 영역 8개 문항, 개인의 의견 영역 9개 문항을 포함한다. 각 문항은 5점 리커트 척도(1점: 매우 그렇지 않다, 5점: 매우 그렇다)로 평가되며, 각 문항별 점수가 높을수록 해당 문항에 대한 가치가 높음을 의미한다.

YLP-V 영역 중 활동 영역의 경우 종속변수와 문항 구성이 유사하여, 본 연구에는 관심사 영역과 의견 영역에 해당되는 문항 응답 결과를 토대로 분석하였다. 본 설문 도구의 Cronbach's  $\alpha$ 는 .917로 나타났다.

### 4. 분석 절차

본 연구의 분석 절차는 총 5단계로 이루어졌다. 첫째, 연구 대상자들의 일반적 특성을 파악하고자 범주형 변수는 빈도, 연속형 변수는 평균치를 산출하였다. 둘째, 잠재프로파일 분석을 수행하여, 라이프스타일 유형화에 있어 최적의 모형적합도를 나타내는 잠재집단 수를 바탕으로 라이프스타일이 몇 개의 유형으로 구분되는지 도출하였다. 해당 결과로 도출된 유형은 머신러닝

모델의 종속변수와 같다. 셋째, 잠재프로파일 분석을 통해 도출된 각 유형별 라이프스타일 및 상대적 가치의 기술통계량과 집단 간 차이를 파악하고자 독립표본 T 검정을 수행하였다. 넷째, 도출된 라이프스타일 잠재집단 수를 종속변수로 하는 머신러닝 모형을 구축하고 모형별로 성능 평가를 실시하였다. 다섯째, 머신러닝 모형들 중 분류 성능이 가장 높은 머신러닝 알고리즘을 선정한 후, 라이프스타일 유형을 분류하는 개인적 가치 변수의 상대적 중요도를 도출하였다.

### 1) 잠재프로파일 분석

라이프스타일에 대한 잠재집단의 수를 결정하기 위한 통계적 기준으로 아카이케 정보기준지수인 Aikake's information criterion (AIC) (Akaike, 1974), 베이저언 정보기준지수인 Bayesian information criterion (BIC) (Schwarz, 1978), 표본크기보정 베이저언 정보기준지수인 Sample Size adjusted BIC (SABIC)를 이용하여 잠재계층의 수를 확인하였다. 해당 지수들은 수치가 낮을수록 적합도가 좋은 것을 의미한다. 또, 모형 비교 검증으로서 조정된 카이제곱차이검증인 Lo-Mendel-Ruben likelihood ratio test (LMRLRT) (Lo et al., 2001)와 모수적 부츠 트랩 우도비 검정인 Bootstrap Likelihood Ratio Test (BLRT) (Nylund et al., 2007) 검증을 통해 잠재집단이  $K$ 개인 모형이  $K-1$ 개인 모형보다 타당한지 유의성을 확인하였다. 또한, entropy 최소 .8 이상, 1에 가까울수록 좋은 모형 기준(Muthén, 2004)을 적용하여 분류의 질을 확인하였다. 집단별 소속 확률을 고려하여 1% 미만인 소속하는 잠재프로파일의 유무를 추가적으로 확인하였다(Jung & Wickrama, 2008). 본 연구에서 잠재프로파일 분석은 SAS version 9.4 (SAS Institute Inc.)와 Mplus version 8.0 (Muthén & Muthén) 통계프로그램을 활용하여 분석하였다.

### 2) 기술통계 분석

잠재프로파일 분석을 통해 도출된 각 유형별 라이프스타일의 기술통계량과 집단 간 차이는 독립표본 T-검

정을 통해 분석되었다. 통계적 유의수준은 .05로 설정하였으며, 양측검정이 수행되었다.

### 3) 머신러닝 모델 구축 및 성능 평가

머신러닝 모델의 구축 및 성능 평가 과정은 다음과 같다. 첫째, 머신러닝 모델의 예측변수(predictor)를 선정하였다. 예측변수는 YLP-V 응답자료를 활용하였다. 둘째, 머신러닝 모델은 로지스틱 회귀 모델, 의사결정 알고리즘, 랜덤 포레스트, 서포트 벡터 머신 알고리즘 4개 모델로 구축하였다. 머신러닝 모델이 훈련 데이터에서 과적합(Overfitting)이 나타날 경우 검증 데이터에서 오류가 증가할 수 있으므로 정규화 혹은 일반화 (Regularization) 절차가 필요하다고 보고(Feurer & Hutter, 2019)되는 바, 본 연구는 과적합을 방지하고자 하이퍼파라미터(Hyper-parameter)를 설정하였다.

#### (1) 로지스틱 회귀모형

로지스틱 회귀모형은 예측변수와 출력 변수 간의 관계를 모델링하여 변수의 값을 예측하는 데 사용되는 모형이다(Bisong, 2019). 예측변수와 출력변수 사이의 선형 관계를 통해 확률을 추정하며, 시그모이드(Sigmoid) 함수를 사용하여 출력을 제한한다. 모델은 학습 데이터를 기반으로 로그우도를 최대화하도록 모델의 파라미터(Parameter)를 조정할 수 있다.

#### (2) 의사결정 나무 알고리즘

의사결정 나무 알고리즘은 지도 학습 알고리즘으로서 데이터를 분류하는데 목적을 두며, 트리 구조를 사용하여 데이터를 분할하고 예측 모델을 만드는 모형이다(Charbuty & Abdulazeez, 2021; Navada et al., 2011). 각 내부의 노드는 하나의 특성(Attribute)을 기준으로 데이터를 분할하는 규칙을 가지며, 각 단말 노드(Leaf node)는 예측값을 출력한다. 분할 규칙은 정보 이득 (Information gain), 지니 불순도(Gini impurity), 엔트로피(Entropy) 등의 측도를 사용하여 정할 수 있다 (Charbuty & Abdulazeez, 2021; Navada et al., 2011).

본 연구에서는 C5.0 알고리즘을 바탕으로 성장중지 규칙(Growth stopping rules)을 적용한 랜덤서치(Randomsearch) 기법을 사용하였고, 노드를 분할하는데 필요한 최소 샘플 수의 경우 10, 최소 불순도 .001, 트리의 최대 깊이는 2로 설정한 최적의 하이퍼파라미터를 적용하여 분석하였다.

### (3) 랜덤 포레스트

랜덤 포레스트 모형은 앙상블 학습 기반의 알고리즘으로, 다수의 의사결정 트리를 결합하여 예측하는 모형이다(Livingston, 2005; Schonlau & Zou, 2020). 다수의 의사결정 트리를 독립적으로 학습하고, 각 트리의 예측 결과를 결합하여 최종 예측을 수행한다. 전체 모형의 평균을 계산하는 회귀모형과 과반수 투표(Voting)를 결정하는 분류모형을 기준으로 최적의 분류 결과를 도출하는 방식이 적용된다. 본 연구에서는 랜덤서치 기법을 적용하였고, 노드를 분할하는데 필요한 최소 샘플 수의 경우 4, 생성할 트리 개수의 경우 628로 최적의 하이퍼파라미터를 적용하여 분석하였다.

### (4) 서포트 벡터 머신

서포트 벡터 머신은 지도 학습 기반 알고리즘으로, 결정 경계(Decision Boundary)를 찾는 것에 주점을 두고 예측하는 모형이다(Somvanshi et al., 2016; Wang, 2005). 주어진 데이터를 고차원 특징 공간으로 매핑하고, 클래스 간 최대 마진(Margin)을 갖는 결정 경계를 찾으며, 데이터 포인트 중 일부만 사용되는 서포트 벡터(Support Vector)에 의해 결정 경계가 형성된다. 본 연구에서는 랜덤서치 기법을 적용하였고, 커널은 선형으로 지정하여 규제파라미터 C의 경우 10, 규제파라미터 감마(Gamma)의 경우 .01로 최적의 하이퍼파라미터를 적용하여 분석하였다.

### (5) 분류 성능 검증

마지막으로, 머신러닝 모형의 과적합을 방지하기 위해 교차검증을 활용하여 데이터를 5개의 폴드로 나누고,

각 폴드를 번갈아가며 검증 세트로 사용하였다. 머신러닝 모형의 성능은 정확도(Accuracy), 민감도(Sensitivity), 특이도(Specificity), Area Under the Curve (AUC) 수치를 바탕으로 평가하였다. 정확도는 모형이 올바르게 분류한 샘플의 비율, 민감도는 실제 양성인 데이터를 양성으로 예측한 비율(True positive rate), 특이도는 실제 음성인 데이터를 음성으로 예측하는 비율(True negative rate)을 의미한다(Glas et al., 2003). AUC 통계치는 민감도와 '1-특이도(False negative rate)'를 두 축으로 하는 Receiver Operating Characteristic curve (ROC curve) 아래 면적이다(Safari et al., 2016). AUC가 .7 이상일 경우 허용되는(Fair) 수준, .8 이상이면 양호한(Good) 수준, .9 이상이면 우수한(Excellent) 수준으로 구분된다(Safari et al., 2016).

### 4) 머신러닝 모형 예측변수의 상대적 중요성

머신러닝 모형에서 예측변수의 상대적 중요성을 분석하기 위하여, 분류 성능이 가장 좋은 머신러닝 모형을 선정하여 예측변수의 상대적 중요도를 분석하였다. 선형 커널을 사용하여 개인적 가치 변수의 중요도를 계산하였으며, 각 특성의 가중치(Coefficient)를 도출하여 각 변수가 분류 결정에 미치는 영향을 시각화하여 파악하였다. 본 연구의 모든 머신러닝 분석은 Google colab을 통해 분석되었다.

## III. 연구 결과

### 1. 일반적 특성

본 연구에 포함된 연구대상자의 특성은 Table 1과 같다. 연령은 최소 55세부터 최대 79세까지 평균 62.23세로 나타났다. 성별은 남녀가 절반으로 분포되었으며, 교육 수준은 고등학교 졸업 이하 32.33%, 대학교 졸업 57.33%, 대학원 졸업 10.33%로 대학교 졸업자가 가장 많은 것으로 나타났다, 거주 지역은 도시 거주자가

94.33%로 읍면부 거주자에 비해 압도적으로 많은 것으로 나타났다. 경제 수준은 57.33%가 보통으로 응답하였고, 기혼자가 82.67%이며, 주관적 건강 상태는 최소 1점 최대 5점 중 평균 3.76점으로 나타났다.

## 2. 라이프스타일 유형화 결과

### 1) 잠재프로파일 집단의 수 분석 결과

본 연구에 포함된 연구대상자의 라이프스타일 실천 유형에 대한 잠재집단 수를 결정하기 위해, 잠재프로파일의 수를 1개부터 5개까지 하나씩 증가시키며 분석하였다. 잠재프로파일 분석 결과는 Table 2와 같다.

연구 결과에 따르면, 정보지수 AIC, BIC, SABIC는 모두 잠재프로파일이 집단의 수가 증가할수록 값이 점점 낮아졌다. BLRT의 경우 모든 잠재프로파일의 수에서 유의미하였다. 분류의 질인 Entropy는 모든 잠재프로파일의 수에서 0.80 이상으로, Muthén (2004)이 제시한 기준을 충족하였다. 잠재프로파일의 수가 4개인 시점의 경우 LMRLRT값에 통계적 유의성이 나타나지 않았으며, 집단별 소속 확률이 1% 미만인 집단이 나타나 최종 모형에서 제외되었다. 또한, 잠재프로파일의 수가 3개인 모형의 경우 한 집단의 케이스가 49건으로 나타나, 작은 표본 크기에 따른 머신러닝 모델 분석의 한계 (Arlot & Celisse, 2010; Cawley & Talbot, 2010)에 따라

Table 1. Demographic Characteristics (N = 300)

Characteristics	n	%
Age <sup>a</sup>	62.23	5.56
Sex (Male, %)	150	50
Educational level (%)	≤ High school	32.33
	Bachelor's degree	57.33
	> Bachelor's degree	10.33
Residence (Urban, %)	283	94.33
Economic level (%)	Good	19
	Normal	57.33
	Bad	23.67
Marital status (Married, %)	248	82.67
Subjective health status (%)	3.76	.73

<sup>a</sup>Mean ± Standard Deviation

The sum of the percentages does not equal 100% because of rounding.

Table 2. Results of Latent Profile Analysis on Lifestyle Behavioral Patterns

Number of class	Log likelihood	AIC	BIC	SABIC	LMRLRT	BLRT	Entropy	Latent class distribution rate n (%)			
								1	2	3	4
1	-7313.451	14658.90	14718.16	14667.42	-	-	1.00	300 (100)	-	-	-
2	-7077.096	14204.19	14296.79	14217.50	463.678*	472.711*	.80	153 (51.11)	147 (51.11)		
3	-7015.796	14099.59	14225.52	14117.69	120.257*	122.599*	.80	49 (16.33)	88 (29.33)	163 (54.33)	
4	-6987.03	14060.06	14219.32	14082.95	56.434	57.534*	.84	153 (51.11)	3 (1.00)	50 (16.68)	94 (31.19)

AIC = Akaike's information criterion; BIC = Bayesian information criterion; BLRT = bootstrap likelihood ratio test; LMRLRT = Lo-Mendel-Ruben likelihood ratio test; SABIC = sample-size adjusted bayesian information criteria.

\*p < .01

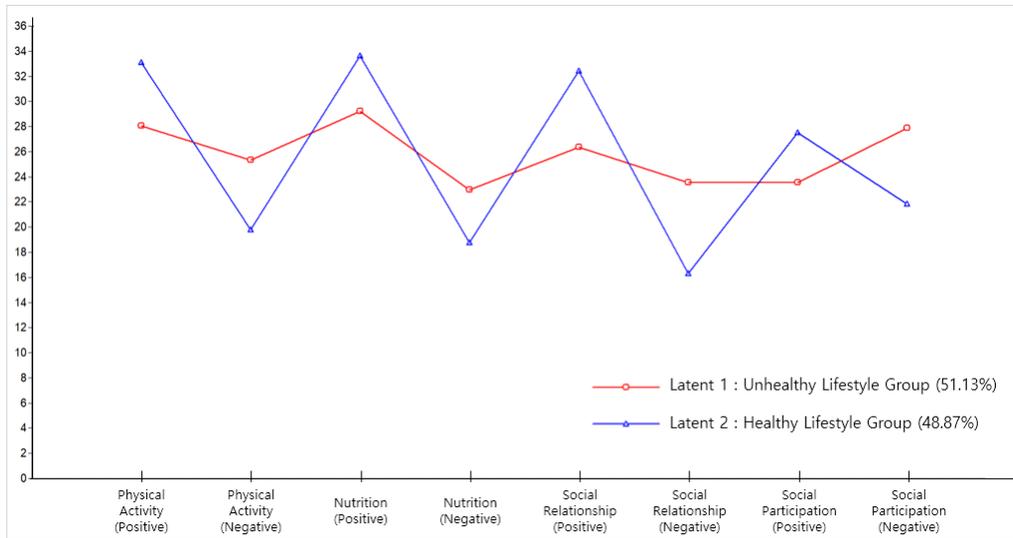


Figure 1. Results of Latent Profile Analysis on Lifestyle Behavioral Patterns

Table 3. The Descriptive Statistics for Latent Profile Analysis Results

Characteristics	HLG		UHLG		T statistics
	M	SD	M	SD	
Physical activity (Positive)	33.10	4.91	27.99	5.39	-8.58*
Physical activity (Negative)	19.77	4.99	25.39	4.58	10.16*
Nutrition (Positive)	33.66	3.62	29.16	4.04	-10.13
Nutrition (Negative)	18.70	3.58	23.06	3.96	9.99*
Social relationship (Positive)	32.46	4.14	26.26	4.73	-12.07*
Social relationship (Negative)	16.25	4.06	23.65	4.23	15.46*
Social participation (Positive)	27.57	4.75	23.48	5.01	- 7.25*
Social participation (Negative)	21.78	3.88	27.98	3.13	15.22*

HLG = Healthy Lifestyle Group; UHLG = Unhealthy Lifestyle Group.

\*  $p < .001$ .

제외되었다. 따라서 본 연구는 잠재프로파일의 수가 2 개인 모형을 최종 모형으로 결정하였다.

## 2) 잠재프로파일 집단의 특성 분석 결과

2개의 집단으로 분류되는 라이프스타일 실천 유형별 특성을 분석한 결과는 Figure 1과 같으며, 각 항목별 상세한 기술통계치와 통계적 유의성에 대한 결과는 Table 3과 같다. 두 잠재집단 간 긍정 영역과 부정 영역 모두 평균치에 두드러진 차이가 나타났다. 긍정 문항의 경우, 모든 값에서 잠재집단 2가 잠재집단 1에 비해 통계적으로 유의미하게 높았으며, 부정 문항은 긍정 문

항과 반대로 잠재집단 1이 잠재집단 2에 비해 통계적으로 유의미하게 높았다. 특히, 두 집단 간 평균 차이는 '사회관계 부정 영역'에서 가장 크게 나타난 것으로 분석되었다.

이에 따라, 잠재집단 2는 잠재집단 1에 비해 신체활동, 식습관, 사회관계, 활동참여 영역에 있어 긍정 영역의 비중이 높고, 부정 영역의 비중이 낮은 특성을 보여 '건강한 라이프스타일 실천군(이하 실천군)'으로 명명하였다. 해당 집단은 전체 대상자의 48.87%를 차지하였다. 반대로, 잠재집단 1은 라이프스타일 4가지 하위 영역에 있어 부정 영역의 비중이 높고 긍정 영역의 비중

이 낮은 특성을 보여 '건강한 라이프스타일 비실천군 (이하 비실천군)'으로 명명하였다. 해당 집단은 전체 대상자의 51.13%를 차지하였다.

### 3. 라이프스타일 유형별 상대적 가치 분석 결과

라이프스타일 잠재집단 간 예측변수에 유의미한 차이가 있는지 분석한 결과는 Table 4와 같다. 모든 예측변수에서 '실천군'이 '비실천군'에 비해 높은 평균치를 보였다. 모든 예측변수의 평균치는 두 잠재집단 간 유

Table 4. Differences in Predictive Variables Between Latent Groups

YLP-V items	HLG		UHLG		T statistics
	M	SD	M	SD	
Interest : 8 items					
1. Family I tend to be interested in family health.	4.15	.60	3.71	.71	-5.80*
2. Job I tend to be interested in productive activities (unpaid work or paid work).	3.74	.72	3.47	.82	-3.02*
3. Community I tend to be very interested in health and welfare issues in the area where I live.	3.50	.72	3.05	.80	-5.06*
4. Recreation I tend to be interested in hobbies or leisure activities.	3.94	.59	3.40	.78	-6.75*
5. Fashion I tend to be interested in the latest information regarding health.	3.97	.56	3.47	.77	-6.52*
6. Food I tend to be interested in a healthy diet.	3.95	.63	3.48	.71	-6.03*
7. Media I tend to be interested in video media or news (articles) on health topics.	3.89	.64	3.46	.74	-5.38*
8. Achievements I tend to plan and practice a regular life for health.	3.98	.61	3.45	.70	-6.94*
Opinion : 9 items					
1. Self-concept I value my own existence and health.	4.31	.59	3.86	.67	-6.13*
2. Health Issue I attach great importance to health problems that are directly related to daily life.	4.29	.54	3.76	.64	-7.73*
3. Politics I think health-related national policies are important.	4.24	.67	3.78	.70	-5.72*
4. Work-life balance I value work-life balance activities as long as my health allows.	4.19	.60	3.84	.62	-5.00*
5. Economics I tend to spare no consumption expenditures for health care.	3.95	.66	3.70	.66	-3.34*
6. Education I think it is important to acquire health-related knowledge and information.	4.07	.61	3.69	.64	-5.31*
7. Products I value the purchase of health-related food and products and the use of services.	3.95	.63	3.61	.69	-4.39*
8. Future I believe that lifestyle changes are important for health.	4.20	.57	3.91	.60	-4.26*
9. Culture I value health management practices in line with changing health information.	4.22	.58	3.92	.68	-4.06*

HLG = Healthy Lifestyle Group; UHLG = Unhealthy Lifestyle Group.

\* $p < .001$ .

의미한 차이를 보였다.

#### 4. 머신러닝 모델 구축 및 분류 성능

라이프스타일 유형 분류의 성능을 검증하기 위해 17개 문항의 예측변수를 로지스틱 회귀, 의사결정 알고리즘, 랜덤 포레스트, 서포트 벡터 머신 모델에 투입하였다. 4개의 머신러닝 모델을 구축하고 모델별 성능을 비교한 결과는 Table 5와 같다.

실천군과 비실천군을 구분하는 민감도는 의사결정

모형, 로지스틱 회귀모형, 랜덤 포레스트모형, 서포트 벡터 머신 순서로 각각 .73(.59~.86), .88(.79~.97), .94(.86~.99), .96(.89~.99)으로 서포트 벡터 머신이 가장 높은 것으로 나타났다. 일반적으로 민감도와 특이도는 반비례함을 고려해볼 때, 본 연구의 4가지 머신러닝 모델의 특이도는 .80 이상으로 허용되는 수준으로 나타났다. 민감도와 특이도를 모두 고려한 AUC값은 서포트 벡터 머신이 .95(.90~.99)로 가장 높으며, 정확도 역시 서포트 벡터 머신이 .96(.90~.99)으로 가장 높았다. 민감도, 특이도, AUC값, 정확도를 모두 종합했을 때 가장 성능이

Table 5. Classification Performance of Machine Learning Model

Model	AUC (95% CI)	Accuracy (95% CI)	Specificity (95% CI)	Sensitivity (95% CI)
LR	.94 (.89~.99)	.93 (.86~.99)	.99 (.99~.99)	.88 (.79~.97)
DT	.79 (.68~.88)	.79 (.68~.86)	.84 (.70~.96)	.73 (.59~.86)
RF	.91 (.85~.97)	.92 (.86~.97)	.95 (.77~.99)	.94 (.86~.99)
SVM	.95 (.90~.99)	.96 (.90~.99)	.95 (.85~.99)	.96 (.89~.99)

AUC = Area under the ROC curve; CI = Confidence Interval; DT = Decision Tree algorithm; LR = Logistic Regression; RF = Random Forest; SVM = Support Vector Machine.

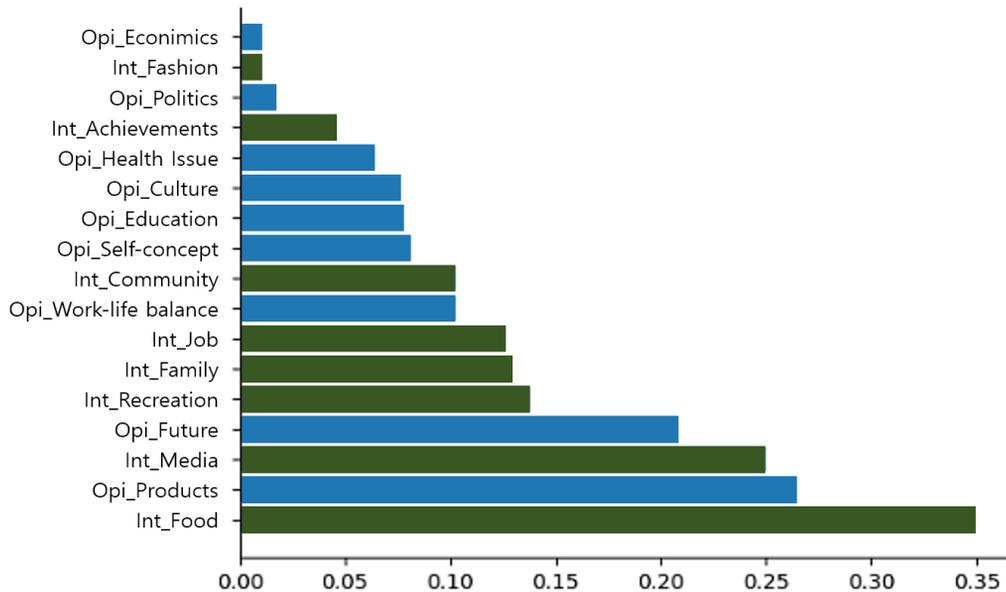


Figure 2. Relative Importance of Predictive Variables

Opi = Opinion items of YLP-V; Int = Interest items of YLP-V.

좋은 집단은 서포트 벡터 머신 모형을 확인하였다.

## 5. 예측변수의 상대적 중요도

4가지 분류모델에서 가장 높은 성능을 보였던 서포트 벡터 머신 모델을 바탕으로 예측변수의 상대적 중요도를 시각화한 결과 Figure 2와 같다. 분류모형에 높은 중요도를 보인 예측변수는 관심사 영역의 “나는 고른 영양소 섭취를 위해 조절된 식단과 건강한 식품에 대해 관심을 갖는 편이다.”, “나는 건강정보를 얻기 위해 영상 매체나 뉴스(기사) 등에 관심을 갖는 편이다.”, “나는 취미 혹은 여가, 문화생활, 스포츠 등에 관심을 갖는 편이다.” 문항으로 나타났다. 의견 영역에서는 “나는 건강관련 식품 및 제품, 이용 가능한 서비스를 중요하게 생각한다.”, “나는 건강을 위해 생활습관의 변화가 중요하다고 생각한다.” 문항이 높은 중요도를 보인 것으로 나타났다.

## IV. 고찰

본 연구는 국내 중고령 노인의 라이프스타일을 두가지로 유형화하고, 머신러닝을 활용하여 어떠한 개인적 가치가 건강한 라이프스타일에 우선적으로 작용하는지 파악하였다. 주요 결과로서, 라이프스타일 유형은 ‘실천군’과 ‘비실천군’으로 분류되었으며, 실천군과 비실천군 간 가장 큰 평균 차이는 사회관계 부정 영역에서 나타났다. 분류 성능이 가장 높은 모형인 서포트 벡터 머신을 통해 개인적 가치 요인의 상대적 중요도를 파악한 결과, 건강식단, 건강 매체, 여가활동에 대한 관심도, 건강 제품 및 건강 습관에 대한 중요도가 라이프스타일 분류에 가장 높게 작용하는 변수임을 확인하였다. 주요 연구결과에 대한 상세한 고찰은 다음과 같다.

첫째, 잠재프로파일 분석을 통해 라이프스타일 잠재집단은 ‘실천군(48.87%)’과 ‘비실천군(51.13%)’으로 명명된 2가지 유형으로 분류되었다. 잠재프로파일 분석을 통해 라이프스타일을 분류한 선행연구(Park et al,

2021b)에 따르면, 수동적이며 불균형적인 라이프스타일군, 기초적 라이프스타일군, 적극적이며 균형적인 라이프스타일군 3가지 잠재집단으로 구분되었다. 하지만, 해당 연구는 라이프스타일을 신체활동, 식습관, 활동참여 3가지 영역으로 구성한 바, 본 연구는 라이프스타일 영역으로서 사회관계를 포함하여 보다 다면적인 라이프스타일을 해석할 수 있다. 본 연구에서 ‘사회관계 부정 영역’은 ‘실천군’과 ‘비실천군’ 간에 가장 주요한 유의미한 평균 차이를 보인 핵심적인 영역으로 나타났다. 다시 말해, 노인의 건강한 라이프스타일을 고려할 때 사회적 연결성이 필수적인 요소임을 시사할 수 있다. 선행연구에 따르면 건강하지 않은 라이프스타일을 보유한 중고령 노인일수록 사회적지지 상태가 낮아, 주변 사람들로부터 도움이 필요한 상황에서 정신적, 재정적, 경제적 지원을 받기 어려움을 보고하였으며(Han, 2023), 동시에 사회적 적응 능력도 저하되어 지역사회 보건시설, 복지시설, 여가시설의 접근, 지역사회 활동 등이 저조한 특성을 보이는 것으로 보고하였다(Han, 2023). 이에 따라, 노인의 라이프스타일 실천 시 사회관계를 필수적인 요소로 반드시 포함하며, 사회관계망을 확장함과 동시에 빈번하고 다양한 사회적 네트워킹을 독려할 필요가 있다.

둘째, 라이프스타일 잠재집단 유형에 따라 4가지 머신러닝 알고리즘에 투입하여 분류 성능을 확인한 결과, 서포트 벡터 머신, 로지스틱 회귀 모형, 랜덤 포레스트 모형, 의사결정 트리 알고리즘 순서로 분류 성능이 좋은 것으로 나타났다. 서포트 벡터 머신은 선형 커널 또는 비선형 커널을 사용하여 이진 분류 문제에 모두 적용 가능한 모형이다(Cortes & Vapnik, 1995). 분석한 결과 4가지 머신러닝 모형 중 서포트 벡터 머신이 가장 좋은 성능을 보였기 때문에, 서포트 벡터 머신의 결정 경계가 주어진 데이터를 가장 잘 분류하며, 이에 따라 일반화 성능이 가장 높은 것으로 해석할 수 있다. 두 번째로 분류 성능이 높았던 로지스틱 회귀모형은 선형 분류 모델로서, 데이터의 선형적 경향성을 잘 반영할 수 있는 모형이다(Pavlyshenko, 2016). 이에, 서포트 벡터 머

신과 로지스틱 회귀 분석이 4가지 머신러닝 모형 중 높은 순위를 차지함에 따라, 두 모형의 공통적인 특성으로서 적용된 서포트 벡터 머신의 선형 커널, 로지스틱 회귀 모형의 선형 추론성이 높은 분류 성능을 야기했을 것으로 사료된다. 이에, 본 연구의 표본으로부터 수집된 데이터가 비선형적 형태보다는 선형적으로 더 잘 구분될 수 있을 것으로 해석된다. 또한, 랜덤 포레스트 역시 AUC값이 .90 이상으로 우수한 성능을 보였으며, 다수의 트리 모델을 결합한 앙상블 방법 역시 라이프스타일 분류에 좋은 성능을 보임을 알 수 있다. 본 연구에서 의사결정 트리 모델은 상대적으로 다른 모델에 비해 낮은 성능을 보여, 해당 데이터셋에서는 의사결정 트리의 단일 모델이 적합하지 않은 것으로 해석된다. 따라서 본 결과는 새로운 데이터를 활용하여 노인의 라이프스타일 잠재 유형을 예측하고 분류할 때 서포트 벡터 머신을 가장 우선적으로 고려할 필요가 있음을 보여준다. 또한, 개발된 머신러닝의 분류모델의 성능이 우수한 수준임을 고려해 볼 때, 본 연구에 투입된 변수들은 노인의 라이프스타일을 분류함에 있어 설명력이 뛰어난 변수로 해석할 수 있다.

셋째, 노인의 라이프스타일 유형을 예측하는 변수 중 상대적 중요도가 높은 가치 변수는 관심사 영역에서 건강 식단, 건강 매체, 여가활동에 대한 관심도, 의견 영역에서 건강 제품 및 건강 습관에 대한 중요도로 나타났다. 이는 다시 말해, 개인이 건강한 식단, 건강 매체, 여가활동에 관심을 두고 있는 정도 및 건강 제품 및 건강 습관에 중요성을 두고 있는 가치가 실천군 혹은 비실천군으로 분류되는 중요한 예측기준이 된다는 것을 의미한다. 체계적 고찰 및 메타분석 연구들을 살펴보면 영양관리 및 식습관은 건강한 생활습관의 일환으로서 노인의 기능저하를 예방하고(Mareschal et al., 2020), 허약성을 낮추는데에 긍정적으로 작용하는 것으로 보고되어(Moraes et al., 2021) 본 연구 결과와 같은 맥락을 나타낸다. 건강매체에 대한 관심도는 건강정보 이해능력과 연관지어, 건강정보 이해능력이 높을수록 건강에 대한 정보를 파악하고 적용하는 능력에 따라 건강한 라

이프스타일에 가치를 두고, 건강한 생활을 실천하고 유지하는 데에 도움을 주는 것으로 보고되었다(Uemura et al., 2021). 또한, 여가활동에 대한 관심도는 본 연구의 유형화 결과와 같은 맥락으로, 사회관계 및 사회참여에 관심을 두고 다양한 활동에 가치를 두는 것이 건강한 라이프스타일과 밀접한 연관이 있음을 알 수 있다. 선행연구에 따르면 건강에 대한 중요한 인식 수준이 개인의 건강 및 삶의 질에 기본적으로 작용하는 요소임이 보고된 바(Divine & Lepisto, 2005), 개인적 가치에 따라 개인의 라이프스타일 유형은 민감하게 작용받는 것으로 해석할 수 있다.

본 연구는 다음과 같은 제한점이 있다. 첫째, 본 연구는 노인의 라이프스타일을 예측할 수 있는 변인으로서 AIO에 기반한 노인의 활동(A), 관심사(I), 의견(O) 영역을 조사하는 YLP-V 평가도구를 사용하였으나, 활동(A) 영역의 경우 종속변수와 상관이 높아 정보 누설의 위험성(Kohavi & John, 1997), 과적합의 가능성(Hastie et al., 2009), 모델의 편향 문제 등의 가능성으로 본 분석에서 제외하였다. 향후 예측요인의 범위를 넓혀 더욱 다양한 개인적 요인과 환경적 요인을 포괄하여 예측할 필요가 있다. 둘째, 본 연구의 경우 YLP-V 문항의 개인적 가치 중 가장 우선순위가 높은 문항들을 탐색하기 위해 변수 선정(feature selection) 과정을 거치지 않아, 모형의 과적합 문제를 가지고 있다. 변수 선정 과정은 모형의 과적합을 방지하고 불필요한 정보로 인한 모델의 복잡성을 줄이는 데에 기여하는 것으로 보고된다(Kotsiantis, 2011; Salam et al., 2021). 이에, 향후 모형의 과적합을 방지하고 타당한 결과 추론을 위해 연구에서 탐색된 상대적 중요도 결과를 토대로 체계적 변수 선정 과정을 거칠 필요가 있다. 셋째, 본 연구에 사용된 모든 변인은 온라인 설문 조사 자료로, 웹기반 온라인 설문 조사에 따른 응답의 신뢰성에 대한 문제를 배제할 수 없다(Bethlehem, 2010; Gosling et al., 2004). 따라서, 추후 연구에서는 체계화된 표집방법과 전문조사자를 활용한 데이터를 활용할 필요가 있을 것이다.

이와 같은 제한점들에도 불구하고 본 연구는 다음과

같은 주요 강점이 있다. 본 연구의 잠재프로파일 분석을 통해 구분된 라이프스타일 잠재집단을 머신러닝으로 분류하고, 분류에 우선적으로 작용하는 요인을 도출한 최초의 연구이다. 기존 노인의 건강관련 종속변수와 독립변수 간의 관계에 초점을 맞추었던 선행연구들과는 달리 사람의 가치에 기초하여 노인의 라이프스타일 실천 유형을 분류하였다는 점에서 의의가 있다. 이를 통해 노인의 라이프스타일을 연구 분야에 있어 머신러닝을 바탕으로 노인의 건강 수준을 분별할 수 있는 기반을 마련하였다는 점에서 주요 의의가 있다.

## V. 결론

본 연구는 국내 지역사회에 거주하는 55세 이상 중고령 노인을 대상으로, 개인의 가치가 건강한 라이프스타일의 실천 유형을 예측할 수 있는지 파악하였다. 연구 결과, 실천군에 속한 중고령자는 비실천군에 비해 사회적 관계가 활발하고, 건강식단, 건강매체, 여가활동에 관심도 및 건강 제품 및 건강 습관에 대한 인식이 높은 것을 확인하였다. 이를 바탕으로 본 연구는 중고령자의 라이프스타일이 단순히 신체활동, 식습관, 활동참여로만 이루어진 것이 아니라, 사회적 관계망도 중요한 영역임을 강조하며, 노인의 건강한 라이프스타일을 고려할 때 사회적 연결성이 필수적인 요소임을 시사한다. 또한, 다양한 개인적 가치요인 가운데 건강한 식단, 건강매체, 여가활동, 건강 제품 및 습관에 높은 주의와 중요성을 부여할수록 건강한 라이프스타일을 실천할 가능성이 크게 높아짐을 나타내었다. 이에 따라, 노인의 사회적 관계성을 포함한 건강한 라이프스타일을 유도하기 위해서는 건강가치에 대한 인식 개선을 중점적으로 다루는 종합적인 프로그램과 서비스가 필요할 것이다.

## Conflicts of interest

No potential conflict of interest relevant to this article was reported.

## Acknowledgements

- 이 논문 또는 저서는 2021년 대한민국 교육부와 한국연구재단의 지원을 받아 수행된 연구임(NRF-2021S1A3A2A02096338)
- 미래융합연구원(ICONS), 연세대학교

## References

- Ahmed, M., Zehou, S., Raza, S. A., Qureshi, M. A., & Yousufi, S. Q. (2020). Impact of CSR and environmental triggers on employee green behavior: The mediating effect of employee well-being. *Corporate Social Responsibility and Environmental Management*, 27(5), 2225-2239. <https://doi.org/10.1002/csr.1960>
- Akaike, H. (1974). A new look at the statistical model identification. *Institute of Electrical and Electronics Engineers Transactions on Automatic Control*, 19(6), 716-723. <http://doi.org/10.1109/TAC.1974.1100705>
- Arlot, S., & Celisse, A. (2010). A survey of cross-validation procedures for model selection. *Statistics Surveys*, 4, 40-79. <https://doi.org/10.1214/09-SS054>
- Bethlehem, J. (2010). Selection bias in web surveys. *International Statistical Review*, 78(2), 161-188. <https://doi.org/10.1111/j.1751-5823.2010.00112.x>
- Bisong, E. (2019). Building machine learning and deep learning models on google cloud platform: A comprehensive guide for beginners. *Apress*, 243-250. <https://doi.org/10.1007/978-1-4842-4470-8>
- Cawley, G. C., & Talbot, N. L. C. (2010). On over-fitting in model selection and subsequent selection bias in performance evaluation. *Journal of Machine Learning Research*, 11, 2079-2107.
- Charbuty, B., & Abdulazeez, A. (2021). Classification based

- on decision tree algorithm for machine learning. *Journal of Applied Science and Technology Trends*, 2(1), 20-28. <https://doi.org/10.38094/jastt20165>
- Chen, Y., & Feeley, T. H. (2014). Social support, social strain, loneliness, and well-being among older adults: An analysis of the health and retirement study. *Journal of Social and Personal Relationships*, 31(2), 141-161. <https://doi.org/10.1177/0265407513488728>
- Cortes, C., & Vapnik, V. (1995). Support-vector networks. *Machine Learning*, 20(3), 273-297. <https://doi.org/10.1007/BF00994018>
- Cudjoe, T. K. M., Roth, D. L., Szanton, S. L., Wolff, J. L., Boyd, C. M., & Thorpe, R. J. (2020). The epidemiology of social isolation: National health and aging trends study. *Journals of Gerontology: Series B*, 75(1), 107-113. <https://doi.org/10.1093/geronb/gby037>
- Depp, C. A., & Jeste, D. V. (2006). Definitions and predictors of successful aging: A comprehensive review of larger quantitative studies. *American Journal of Geriatric Psychiatry*, 14(1), 6-20. <https://doi.org/10.1097/01.JGP.0000192501.03069.bc>
- Divine, R. L., & Lepisto, L. (2005). Analysis of the healthy lifestyle consumer. *Journal of Consumer Marketing*, 22(5), 275-283. <https://doi.org/10.1108/07363760510611707>
- Feurer, M., & Hutter, F. (2019). Hyperparameter optimization. In F. Hutter, L. Kotthoff, & J. Vanschoren (Eds.), *Automated machine learning: Methods, systems, challenges* (pp. 3-33). Springer. [https://doi.org/10.1007/978-3-030-05318-5\\_1](https://doi.org/10.1007/978-3-030-05318-5_1)
- Ford, E. S., Bergmann, M. M., Boeing, H., Li, C., & Capewell, S. (2012). Healthy lifestyle behaviors and all-cause mortality among adults in the United States. *Preventive Medicine*, 55(1), 23-27. <https://doi.org/10.1016/j.ypmed.2012.04.016>
- Glas, A. S., Lijmer, J. G., Prins, M. H., Bonsel, G. J., & Bossuyt, P. M. (2003). The diagnostic odds ratio: A single indicator of test performance. *Journal of Clinical Epidemiology*, 56(11), 1129-1135. [https://doi.org/10.1016/S0895-4356\(03\)00177-X](https://doi.org/10.1016/S0895-4356(03)00177-X)
- Gosling, S. D., Vazire, S., Srivastava, S., & John, O. P. (2004). Should we trust web-based studies? A comparative analysis of six preconceptions about internet questionnaires. *American Psychologist*, 59(2), 93-104. <https://doi.org/10.1037/0003-066X.59.2.93>
- Han, Y. (2023). *A study of the lifestyle effects on social health of Korean older adults: Focused on structural equation modeling* (Master's thesis). Yonsei University.
- Hastie, T., Tibshirani, R., & Friedman, J. H. (2009). *The elements of statistical learning: Data mining, inference, and prediction*. Springer. <https://doi.org/10.1007/978-0-387-84858-7>
- Hrubá, R., & Sadilek, T. (2022). Social relations to food, the meaning of the farmers' market and healthy lifestyle: The case of Czechia and Ukraine. *Journal of International Food and Agribusiness Marketing*, 35(4), 1-19. <https://doi.org/10.1080/08974438.2022.2056558>
- Hu, F. B., Manson, J. E., Stampfer, M. J., Colditz, G., Liu, S., Solomon, C. G., Walter, C., & Willett, W. C. (2001). Diet, lifestyle, and the risk of type 2 diabetes mellitus in women. *New England Journal of Medicine*, 345(11), 790-797. <https://doi.org/10.1056/NEJMoa010492>
- Huang, W., Ying, T. W., Chin, W. L. C., Baskaran, L., Marcus, O. E. H., Yeo, K. K., & Kiong, N. S. (2022). Application of ensemble machine learning algorithms on lifestyle factors and wearables for cardiovascular risk prediction. *Scientific Reports*, 12(1), 1033. <https://doi.org/10.1038/s41598-021-04649-y>
- Jung, T., & Wickrama, K. A. (2008). An introduction to latent class growth analysis and growth mixture modeling. *Social and Personality Psychology Compass*, 2(1), 302-317. <https://doi.org/10.1111/j.1751-9004.2007.00054.x>
- Knoops, K. T., de Groot, L. C., Kromhout, D., Perrin, A. E., Moreiras-Varela, O., Menotti, A., & Van Staveren, W. A. (2004). Mediterranean diet, lifestyle factors, and 10-year mortality in elderly European men and women: The HALE project. *Journal of the American Medical Association*, 292(12), 1433-1439. <https://doi.org/10.1001/jama.292.12.1433>
- Kohavi, R., & John, G. H. (1997). Wrappers for feature subset selection. *Artificial Intelligence*, 97(1-2), 273-324. [https://doi.org/10.1016/S0004-3702\(97\)00043-X](https://doi.org/10.1016/S0004-3702(97)00043-X)
- Kotsiantis, S. (2011). Feature selection for machine learning classification problems: A recent overview. *Artificial Intelligence Review*, 42(1), 157-176. <https://doi.org/10.1007/s10462-011-9230-1>
- Lim, S. J., & Park, J. H. (2020). The study of the convergent factors of the lifestyle on the cognitive decline among elderly. *Journal of the Korea Convergence Society*, 11(8), 229-236. <https://doi.org/10.15207/JKCS.2020.11.8.229>
- Lim, Y. M., Park, K. H., & Park, J. H. (2023). Development of lifestyle evaluation tools based on the values of

- the elderly: A delphi study. *Sage Open*, 13(3), 1-9. <https://doi.org/10.1177/21582440231194802>
- Livingston, F. (2005). Implementation of Breiman's random forest machine learning algorithm. *ECE591Q Machine Learning Journal Paper*, 1-13.
- Lo, Y., Mendell, N. R., & Rubin, D. B. (2001). Testing the number of components in a normal mixture. *Biometrika*, 88(3), 767-778. <https://doi.org/10.1093/biomet/88.3.767>
- Loef, M., & Walach, H. (2012). The combined effects of healthy lifestyle behaviors on all cause mortality: A systematic review and meta-analysis. *Preventive Medicine*, 55(3), 163-170. <https://doi.org/10.1016/j.ypmed.2012.06.017>
- Mareschal, J., Genton, L., Collet, T. H., & Graf, C. (2020). Nutritional intervention to prevent the functional decline in community-dwelling older adults: A systematic review. *Nutrients*, 12(9), 2820. <https://doi.org/10.3390/nu12092820>
- Michel, J. P., & Sadana, R. (2017). "Healthy aging" concepts and measures. *Journal of the American Medical Directors Association*, 18(6), 460-464. <https://doi.org/10.1016/j.jamda.2017.03.008>
- Moraes, M. B. D., Avgerinou, C., Fukushima, F. B., & Vidal, E. I. (2021). Nutritional interventions for the management of frailty in older adults: Systematic review and meta-analysis of randomized clinical trials. *Nutrition Reviews*, 79(8), 889-913. <https://doi.org/10.1093/nutrit/nuaa101>
- Muthén, B. (2004). Latent variable analysis. *The Sage Handbook of Quantitative Methodology for the Social Sciences*, 345(368), 106-109. <https://doi.org/10.4135/9781412986311>
- Navada, A., Ansari, A. N., Patil, S., & Sonkamble, B. A. (2011). Overview of use of decision tree algorithms in machine learning. *Proceedings of the 2011 Institute of Electrical and Electronics Engineers Control and System Graduate Research Colloquium*, 37-42. <https://doi.org/10.1109/ICSGRC.2011.5991826>
- Nylund, K. L., Asparouhov, T., & Muthén, B. O. (2007). Deciding on the number of classes in latent class analysis and growth mixture modeling: A monte carlo simulation study. *Structural Equation Modeling: A Multidisciplinary Journal*, 14(4), 535-569. <http://doi.org/10.1080/10705510701575396>
- Park, J. H., Park, H. Y., Hong, I., Han, D. S., Lim, Y. M., Kim, A. R., Nam, S., Park, K. H., Lim, S., Bae, S., & Jin, Y. (2023). Conceptual model of establishing lifestyle (Lifestyle-DEPER [Decision, execution, personal factor, environment, resources]) and lifestyle intervention strategies. *Therapeutic Science for Rehabilitation*, 12(4), 9-22. <https://doi.org/10.22683/tsnr.2023.12.4.009>
- Park, K. H., Yang, M. A., Won, K. A., & Park, J. H. (2021). Predicting healthy lifestyle patterns in older community dwelling adults: A latent profile analysis. *Therapeutic Science for Rehabilitation*, 10(2), 75-93. <https://doi.org/10.22683/tsnr.2022.11.2.075>
- Park, K. H., Yoo, E. Y., Kim, J., Hong, I., Lee, J. S., & Park, J. H. (2021). Applying latent profile analysis to identify lifestyle profiles and their association with loneliness and quality of life among community-dwelling middle-and older-aged adults in South Korea. *International Journal of Environmental Research and Public Health*, 18(23), 12374. <https://doi.org/10.3390/ijerph182312374>
- Pavlyshenko, B. (2016). Machine learning, linear and bayesian models for logistic regression in failure detection problems. *Proceedings of the 2016 Institute of Electrical and Electronics Engineers International Conference on Big Data*, 2046-2050. <http://doi.org/10.1109/bigdata.2016.7840828>
- Plummer, J. T. (1974). The concept and application of life style segmentation: The combination of two useful concepts provides a unique and important view of the market. *Journal of Marketing*, 38(1), 33-37. <https://doi.org/10.2307/1250164>
- Polidori, M. C., Nelles, G., & Pientka, L. (2010). Prevention of dementia: Focus on lifestyle. *International Journal of Alzheimer's Disease*, 2010, 1-9. <https://doi.org/10.4061/2010/393579>
- Safari, S., Baratloo, A., Elfil, M., & Negida, A. (2016). Evidence based emergency medicine: Part 5 receiver operating curve and area under the curve. *Emergency*, 4(2), 111-113. <https://doi.org/10.22037/emergency.v4i2.11903>
- Salam, M. A., Azar, A. T., Elgendy, M. S., & Fouad, K. M. (2021). The effect of different dimensionality reduction techniques on machine learning overfitting problem. *International Journal of Advanced Computer Science and Applications*, 12(4), 641-655. <https://doi.org/10.14569/IJACSA.2021.0120480>
- Schonlau, M., & Zou, R. Y. (2020). The random forest algorithm for statistical learning. *Stata Journal*, 20(1), 3-29. <https://doi.org/10.1177/1536867X20909688>

- Schwarz, G. (1978). Estimating the dimension of a model. *Annals of Statistics*, 461-464. <https://doi.org/10.1214/aos/1176344136>
- Somvanshi, M., Chavan, P., Tambade, S., & Shinde, S. V. (2016). A review of machine learning techniques using decision tree and support vector machine. *Proceedings of the 2016 International Conference on Computing Communication Control and Automation*, 1-7. <https://doi.org/10.1109/iccubea.2016.7860040>
- Spring, B., Moller, A. C., & Coons, M. J. (2012). Multiple health behaviours: Overview and implications. *Journal of Public Health*, 34(suppl\_1), i3-i10. <https://doi.org/10.1093/pubmed/fdr111>
- Sun, Q., Townsend, M. K., Okereke, O. I., Franco, O. H., Hu, F. B., & Grodstein, F. (2010). Physical activity at midlife in relation to successful survival in women at age 70 years or older. *Archives of Internal Medicine*, 170(2), 194-201. <https://doi.org/10.1001/archinternmed.2009.503>
- Uemura, K., Yamada, M., & Okamoto, H. (2021). The effectiveness of an active learning program in promoting a healthy lifestyle among older adults with low health literacy: A randomized controlled trial. *Gerontology*, 67(1), 25-35. <https://doi.org/10.1159/000511357>
- Wang, L. (Ed.). (2005). *Support vector machines: Theory and applications* (STUDFUZZ, Vol. 177). Springer.
- Wang, R., Chaudhari, P., & Davatzikos, C. (2023). Reply to Holm et al.: Careful training is a good practice. *Proceedings of the National Academy of Sciences*, 120(23), e2305855120. <https://doi.org/10.1073/pnas.2305855120>
- Yang, M. A., Lim, S., Han, D., & Park, J. (2020). A policy proposal for improving the quality of life for the middle-aged people: Focusing on lifestyle. *Journal of the Korean Gerontology Society*, 40(3), 531-541. <https://doi.org/10.31888/JKGS.2020.40.3.531>
- Zhou, Z. H. (2021). *Machine learning*. Springer Nature.

## Abstract

### Identifying Personal Values Influencing the Lifestyle of Older Adults: Insights From Relative Importance Analysis Using Machine Learning

Lim, Seungju\*, M.S., O.T., Park, Ji-Hyuk\*\*, Ph.D., O.T.

\*Dept. of Occupational Therapy, Graduate School, Yonsei University, Doctoral Student

\*\*Dept. of Occupational Therapy, College of Software and Digital Healthcare Convergence, Yonsei University, Professor

**Objective :** This study aimed to categorize the lifestyles of older adults into two types - healthy and unhealthy, and use machine learning to identify the personal values that influence these lifestyles.

**Methods :** This cross-sectional study targeting middle-aged and older adults (55 years and above) living in local communities in South Korea. Data were collected from 300 participants through online surveys. Lifestyle types were dichotomized by the Yonsei Lifestyle Profile (YLP)-Active, Balanced, Connected, and Diverse (ABCD) responses using latent profile analysis. Personal value information was collected using YLP-Values (YLP-V) and analyzed using machine learning to identify the relative importance of personal values on lifestyle types.

**Results :** The lifestyle of older adults was categorized into healthy (48.87%) and unhealthy (51.13%). These two types showed the most significant difference in social relationship characteristics. Among the machine learning models used in this study, the support vector machine showed the highest classification performance, achieving 96% accuracy and 95% area under the receiver operating characteristic (ROC) curve. The model indicated that individuals who prioritized a healthy diet, sought health information, and engaged in hobbies or cultural activities were more likely to have a healthy lifestyle.

**Conclusion :** This study suggests the need to encourage the expansion of social networks among older adults. Furthermore, it highlights the necessity to comprehensively intervene in individuals' perceptions and values that primarily influence lifestyle adherence.

**Keywords :** Lifestyle, Machine learning, Older adults, Personal values, Relative importance