

## 직업적 특성에 따른 인지기능장애 예측 및 중요 변수 탐색

정 석 송\*

직업적 특성이 인지기능장애에 미치는 영향은 직업 활동과 뇌 기능 사이의 상호작용으로 인하여 나타난다. 본 연구는 특정 유형의 직업 활동과 특성이 개인의 인지기능에 어떠한 영향을 미치는지에 대한 이해와 과학적 근거 확보를 위해 수행되었다. 2004년 7차 한국노동패널 조사에 참여한 대상자를 기반으로 2022년까지의 인지기능장애 발생 패널데이터를 통하여 추적관찰이 수행되었다. XGboost 기법을 통하여 직업적 특성의 비선형적 변수중요도를 평가하였으며 로지스틱회귀모형을 통하여 선형적 독립예측인자를 평가하였다. 변수중요도는 직업의 유무, 정규직 여부, 고정근무시간 및 연소득 순서로 나타났으며 독립예측인자는 정규직 여부와 연소득으로 확인되었다. 이러한 결과는 직업적 특성이 인지기능장애와 비선형적 연관성이 있을 수 있음을 시사하며 직업적 특성의 개선을 통한 인지기능장애 발생 위험의 낮춤에 대한 가능성을 시사한다.

주요용어 : 인지기능장애, 직업, 정규직, 고정근무시간, 연소득, 한국노동패널(KLIPS)

### 1. 들어가는 글

직업적 특성과 질병 발생 사이에는 여러가지 요인들이 영향을 미칠 수 있다. 특정 작업 조건으로 인한 질병 발생뿐만 아니라 사회적 및 정신적 직업 특성도 영향을 미치는 것으로 알려졌다(Grotz, 2018). 인지기능장애는 인지 능력에 영향을 미치는 질병 또는 장애로 기억, 사고, 지각, 언어 등의 기능에 문제가 있는 상태이며 주변 사회와 환경, 기저질환 등으로 인하여 영향을 받는 것으로 알려지고 있다(Schubert, 2023; Xu, 2022). 특히, 돌봄 제공자들의 고용 상태가 경도인지기능장애와 유의미한 연관성을 보이는 것으로 알려졌다(Connors, 2019). 본 연구에서는 직업의 유무, 정규직 유무, 고정근로시간 및 연소득수준 등의 직업적 특성이 인지기능장애 발생에 가지는 예측력과 변수중요도를 평가하고자 한다.

---

\* 차의과학대학교 의학전문대학원 연구교수

## II. 실증자료

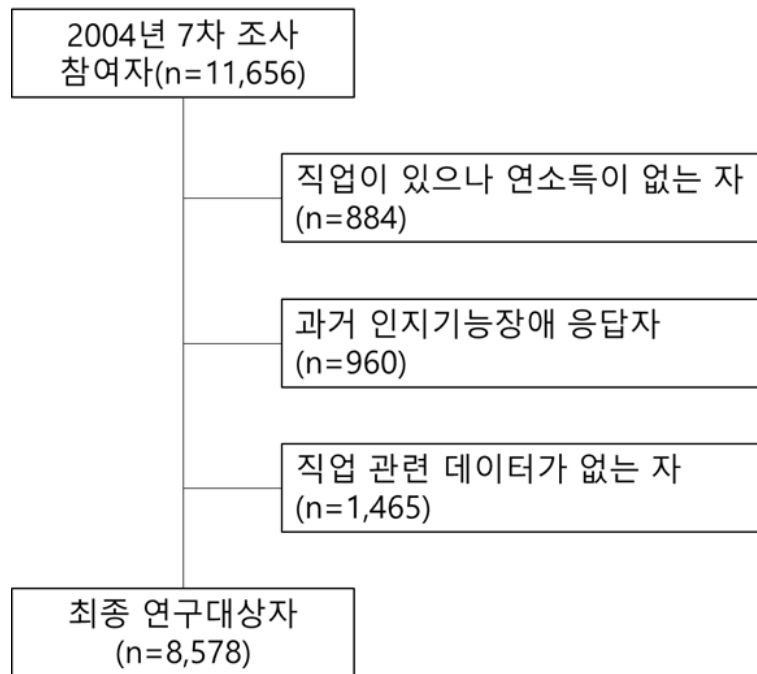
### 1. 한국노동패널

본 연구에 사용된 자료는 한국노동연구원에서 2003년에서 2022년까지 실시한 한국노동패널의 자료이다. 본 자료는 노동관련 가구패널조사로 횡단면 자료와 시계열 자료의 장점을 모두 가지고 있다.

#### 가. 표본의 선정

본 연구는 2004년 7차 조사에 참여한 대상자 중 개인 자료가 있는 11,656명에서 직업은 있으나 연소득이 없는 자(n=884), 2003년-2004년 사이 인지기능장애 응답자(n=960) 및 직업 관련 데이터가 없는 자(n=1,465)를 제외한 8,578명을 대상으로 수행되었다(그림 1). 최종 연구대상자 중 2022년까지 인지기능장애 관련 문항에 대한 답변을 하지 않은 자는 없었다.

[그림 1] 연구대상자 선정과정



### III. 연구 방법

#### 1. 독립변수와 종속변수

본 연구의 독립변수는 나이(연속변수; 세), 성별(범주변수; 남성 및 여성), 직업 유무(범주변수; 있음 및 없음), 정규직 유무(범주변수; 직업 없음, 정규직 및 비정규직), 고정근무시간(범주변수; 직업 없음, 고정근무 및 비고정근무) 및 연소득(범주변수; <1000만원, 1000-2999만원, 3000-4999만원 및 5000만원 이상)임. 종속변수는 다음과 같은 설문에 기반하였음: 배우기, 기억하기 및 집중하기에 어려움을 느낀다(예, 인지기능장애 있음; 아니오, 인지기능장애 없음).

#### 2. 통계 분석

본 연구는 Extreme Gradient Boosting(XGBoost) 모델을 사용하였다. 이 모델은 결정 트리 기반 앙상블을 통한 머신러닝 학습 알고리즘 중 하나로 구조화된 데이터의 처리에 뛰어난 성능을 보인다. 변수중요도는 Shapley Additive Explanations(SHAP) 패키지를 사용하였다. 변수중요도의 해석을 위해 로지스틱회귀모형 기반의 오즈비를 계산하였다. 모든 데이터마이닝과 통계분석은 Jupyter Notebook 환경의 Python에서 이루어졌다.

### IV. 연구 결과

#### 1. 기본 특성

연구대상자들의 평균 나이는 39.5세였으며 그중 남성이 47.7%였다. 직업이 있는 사람과 정규직인 사람의 비율은 각각 44.6%와 33.4%였다(표 1).

<표 1> 한국노동패널 연구대상자 기본 특성

기본 특성	대상자(n=8,578)
나이, mean (SD)	39.5 (16.3)
성별, n (%)	
남성	4,008 (46.7)
여성	4,570 (53.3)
직업, n (%)	
있음	3,826 (44.6)
없음	4,752 (55.4)

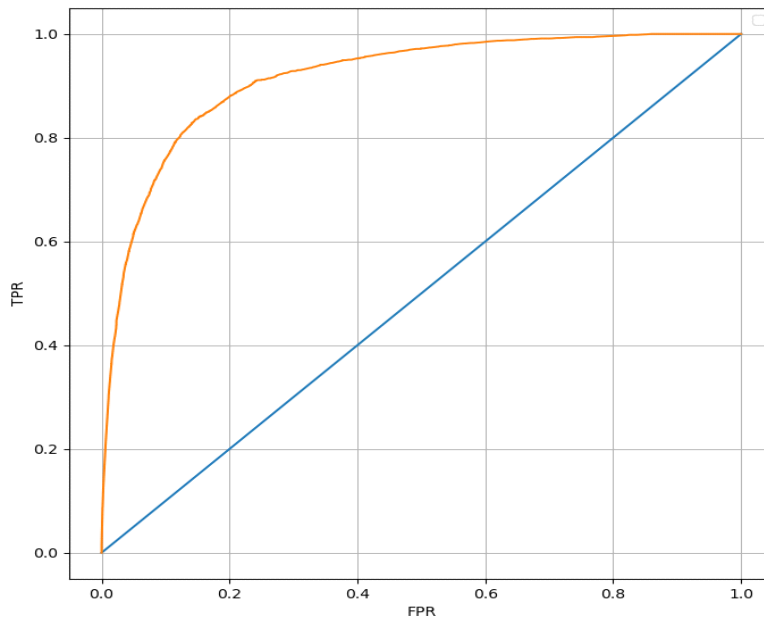
기본 특성	대상자(n=8,578)
정규직 유무, n (%)	
직업 없음	4,752 (55.4)
정규직	2,861 (33.4)
비정규직	965 (11.2)
고정 근무시간, n (%)	
직업 없음	4,752 (55.4)
고정 근무	3,249 (37.9)
비고정 근무	577 (6.7)
연소득, n (%)	
<1000만원	5,519 (64.3)
1000-2999만원	2,416 (28.2)
3000-4999만원	539 (6.3)
≥5000만원	104 (1.2)

약어 : SD, standard deviation.

## 2. XGBoost 모델 기반 인지기능장애 예측

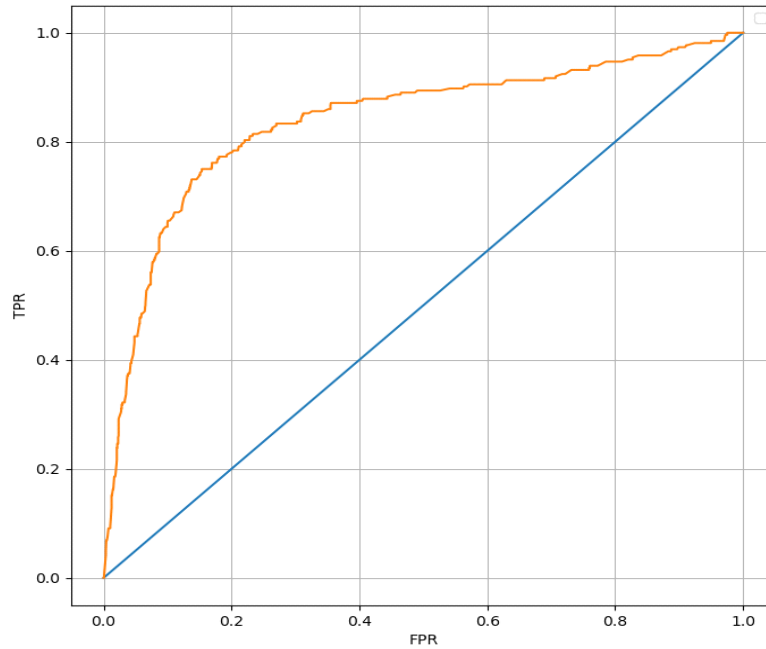
연구대상자 중 20%를 테스트 데이터로 나눈 후 학습률과 최대 깊이를 각각 0.1과 3으로 설정하였다. 회귀 함수로 더 큰 오차에 더 큰 가중치를 주는 Squared Error 기법을 사용한 결과 수신자 조작 특성 곡선에서 곡선 아래 면적 값이 0.916으로 훈련되었다(그림 2).

[그림 2] 훈련 모델 수신자 조작 특성 곡선



테스트 데이터를 통한 검증 결과 수신자 조작 특성 곡선에서 곡선 아래 면적 값이 0.838로 검증 되었다(그림 3).

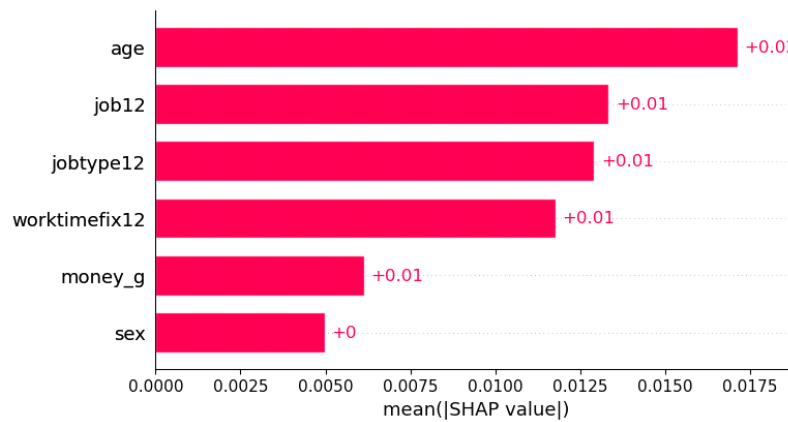
[그림 3] 테스트 데이터 수신자 조작 특성 곡선



### 3. SHAP 기반 변수중요도

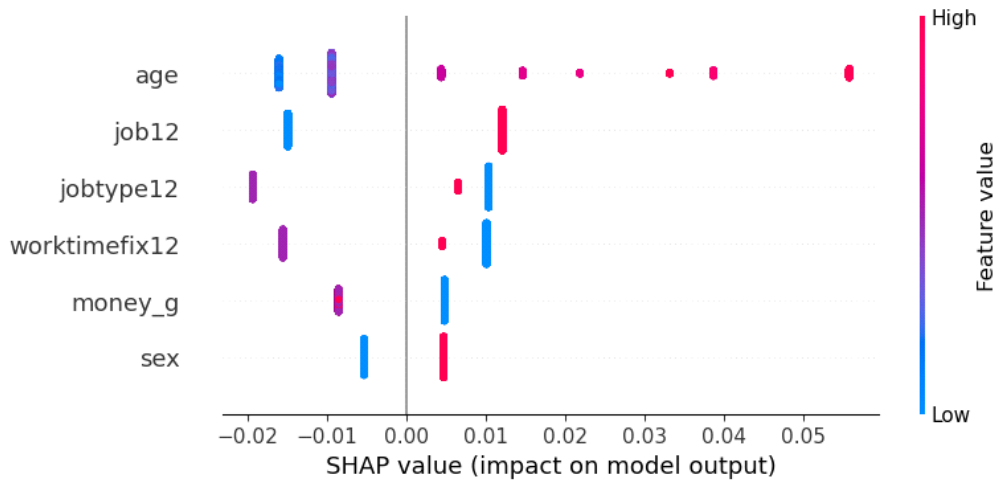
SHAP 기법을 통한 변수중요도 평가 결과 나이, 직업 유무, 정규직 유무, 고정근무시간, 연소득 및 성별 순으로 중요도가 나타났다(그림 4). 아래 그림에서 모든 변수는 양의 영향을 미쳤으며 각각의 막대그래프는 SHAP 값의 절대값을 나타낸다.

[그림 4] 막대그래프를 통한 SHAP의 시각화



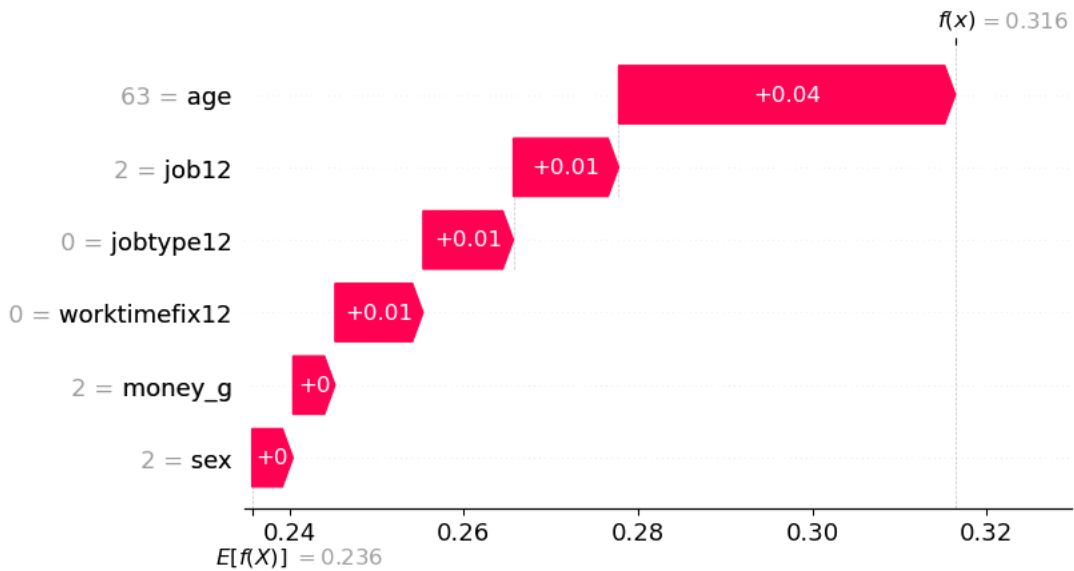
Beeswarm Plot을 통한 SHAP 값들의 분포는 다음과 같았다(그림 5).

[그림 5] Beeswarm Plot을 통한 SHAP의 시각화



Waterfall Plot을 통한 누적 SHAP 값들은 다음과 같았다(그림 6).

[그림 6] Beeswarm Plot을 통한 SHAP의 시각화



#### 4. 로지스틱회귀모형 기반 변수중요도

마지막으로 해석력을 높이기 위한 로지스틱회귀모형 기반 각 변수들의 보정오즈비는 다음과 같았다. 보정오즈비는 나이, 성별, 직업 유무, 정규직 여부, 고정근무 여부 및 연소득이 포함된 다변량

로지스틱회귀모형을 기반으로 평가되었으며 평가 결과 나이, 정규직 여부 및 연소득이 인지기능장애에 대한 독립예측인자로 평가되었다.

<표 2> 로지스틱회귀모형 기반 인지기능장애에 대한 변수들의 오즈비 평가

변수	aOR (95% CI)	P value	P for trend
나이(1세 증가 당)	1.10 (1.10-1.11)	<0.001	
남성(vs 여성)	0.94 (0.80-1.11)	0.473	
직업 있음(vs 없음)	1.20 (0.91-1.58)	0.200	
정규직(vs 비정규직)	0.70 (0.54-0.91)	0.008	
고정근무(vs 비고정근무)	0.81 (0.61-1.10)	0.173	
연소득			0.045
<1000만원	1.00 (reference)		
1000-2999만원	0.81 (0.64-1.02)	0.070	
3000-4999만원	0.59 (0.40-0.89)	0.012	
≥5000만원	0.58 (0.28-1.20)	0.142	

## V. 결 론

본 연구에서는 머신러닝 기법을 통하여 직업적 특성 중 직업의 유무, 정규직 여부, 고정근무시간 여부 및 연소득 순서로 인지기능장애의 발생과 연관성이 있음을 확인하였다. 이러한 직업적 특성과 인지기능장애의 연관성은 비선형성으로 로지스틱회귀모형에서는 다른 결과로 나타났다. 로지스틱회귀모형에서는 정규직 여부와 연소득이 독립예측인자로 확인되었다. 머신러닝 모델에서 가장 낮은 변수중요도를 보였던 연소득이 로지스틱회귀모형에서 유의미한 독립예측인자로 확인된 것으로 보아 본 연구에 사용된 모든 직업적 특성이 인지기능발생과 비선형적 연관성이 있을 가능성을 시사한다. 즉, 직업적 특성의 개선은 인지기능장애 발생 위험을 낮출 가능성이 있다.

## 참고문헌

- Connors, M. H., Seeher, K., Teixeira-Pinto, A., Woodward, M., Ames, D., & Brodaty, H (2019). Mild Cognitive Impairment and Caregiver Burden: A 3-Year-Longitudinal Study. *American Journal of Geriatric Psychiatry*. 27(11):1206-1215.
- Grotz, C., Meillon, C., Amieva, H., et, al. (2018) Occupational social and mental stimulation and cognitive decline with advancing age. *Age and Ageing*. 47(1):101-106.
- Schubert, C. R., Fischer, M. E., Pinto, A. A., et, al (2023). Inflammation, metabolic dysregulation and environmental neurotoxins and risk of cognitive decline and impairment in midlife. *Neurological Sciences*. 44(1):149-157.
- Xu, L., Sun, H., Qu, C., et, al. (2022) The environmental enrichment ameliorates chronic unpredictable mild stress-induced depressive-like behaviors and cognitive decline by inducing autophagy-mediated inflammation inhibition. *Brain Research Bulletin*. 187:98-110.