



해석가능한 기계학습을 활용한 보행목적별 보행만족도 영향 요인 분석*

Analysis of Influencing Factors of Walking Satisfaction by Purpose Using Interpretable Machine Learning

박준상** · 이수기***

Park, Junsang · Lee, Sugie

Abstract

The uncoordinated development and expansion of cities along with industrialization and urbanization have caused various urban issues, such as traffic congestion, heavy energy consumption, and environmental pollution. To alleviate the negative impacts of the automobile-centered urban environment, pedestrian-oriented urban planning and design practices have been proposed and implemented for several decades in Seoul, Korea. However, the factors influencing the walking satisfaction by walking purpose have not been sufficiently investigated in urban planning and design literature. In addition, the majority of previous studies include attempts to diagnose only the linear relationships between the walking satisfaction and the built environment. With the recent development of an interpretable machine learning (IML) model, the nonlinear relationships between the walking satisfaction and the built environment is investigated in this study. Further, the interaction effects of the built environmental variables on the walking satisfaction are identified. The results indicate that the machine learning model shows a significantly higher explanatory power compared with the conventional model. In addition, it is confirmed that IML is a useful tool to understand the nonlinear relationships between the walking satisfaction and the built environment. The analysis results on the aforementioned relationship suggest that they can be used as important data to promote a pedestrian-friendly urban environment. Additionally, the policy implications of promoting walking satisfaction by purpose are presented in this study.

주제어 보행만족도, 해석 가능한 기계학습, 물리적 환경, 비선형관계, 상호작용 효과

Keywords Walking Satisfaction, Interpretable Machine Learning, Physical Environment, Nonlinear Relationship, Interaction Effect

1. 서론

산업화와 도시화에 따른 도시의 무분별한 개발과 외연적 팽창은 도시 내 공간 구조를 자동차 중심으로 변화시켰으며 이로 인한 교통체증, 에너지 소비, 환경오염은 주요 도시문제로 부상하

였다(조혜민·이수기, 2016). 이와 같은 자동차 중심의 도시구조가 미치는 부정적인 영향을 완화하기 위해 그동안 보행자 중심의 도시설계 방안과 공공정책이 제시되어 왔다. 보행자 중심의 보행 친화적 도시환경은 교통, 환경, 에너지 소비 패턴 변화에 영향을 미쳐 장기적으로 논의되고 있는 도시문제를 해결할 수 있을 것

* 이 논문은 2021년 10월 대한국토·도시계획학회 추계학술대회에서 발표하여 우수논문상 수상 논문을 수정·보완한 것이며, 2020년 대한민국 교육부와 한국연구재단의 지원을 받아 수행된 연구임(NRF-2020S1A5A2A01044573).

** Master's Student, Department of Urban Planning & Engineering, Hanyang University (First Author: june1264@hanyang.ac.kr)

*** Professor, Department of Urban Planning & Engineering, Hanyang University (Corresponding Author: sugielee@hanyang.ac.kr)

로 기대되기 때문이다.

미국의 도시학자 Jacobs(1961)는 ‘미국 대도시의 죽음과 삶(The Death and Life of Great America Cities)’에서 도시의 물리적 환경이 보행 활동을 포함한 도시의 활력(urban vitality)에 주요한 역할을 한다고 주장하였다. 구체적으로 Jacobs(1961)는 도시가 다양성과 활력을 유지하기 위해 혼합된 토지이용, 작은 블록, 오래된 건물, 적당하게 높은 밀도라는 네 가지 필수조건을 언급하였다. 또한, 보행 활동에 영향을 미치는 추가 요소로 접근성과 경계 공백 지대(border vacuums)를 제시하였다. 특히 물리적 경계를 구분 짓는 시설이나 자연조건은 경계 공백을 형성하며 보행을 단절시키고 도시 활력을 감소시킬 수 있다고 보았다. 또한, 대규모 단일용도 토지이용은 보행자들에게 보행의 목적을 단일목적의 활동으로 획일화시키기 때문에 그 주변 지역의 도시 활력을 줄일 가능성이 높다고 보았다.

도시계획 분야에서 도시의 물리적 환경요인과 보행 활동과의 관계를 분석하기 위한 연구는 지속적으로 이루어졌다(Ewing et al., 2014; Sallis et al., 2009; Frank et al., 2010; 성현근, 2014; 이경환·안건혁, 2007; 이수기 외, 2014; 박근덕 외, 2021). 이러한 맥락에서 서울시는 ‘자동차 중심 도시’에서 ‘걷는 도시’로 패러다임 전환을 위해 기존에 진행 중이던 보행 정책에 새로운 정책을 포함하여, 2016년에 「걷는 도시, 서울」 종합계획을 발표하였다(이신해·정상미, 2019).

이처럼 근린 환경이 보행 활동 및 보행환경에 미치는 영향에 대한 관심이 높아지자 관련 연구가 활발하게 진행되었으며, 최근에는 보행목적을 구분하여 보행 활동에 영향을 미치는 요인을 제시하였다(이경환·안건혁, 2007; 성현근 외, 2014; 조혜민·이수기, 2016). 또한, 보행 활동은 보행의 목적 및 가로의 다양한 물리적 환경요인의 영향을 받기 때문에 보행환경은 보행자들이 걸으며 경험하는 대상이기도 하면서 머물고자 하는 장소로서의 의미를 가지는 것이 중요하다고 보았다(이수기 외, 2014).

한편, 해외를 포함한 국내 대부분의 정량적 연구에서 보행 활동과 도시의 물리적 환경과의 관계는 선형회귀모형을 통해 이루어졌다. 그러나 도시의 물리적 환경과 보행 활동과의 관계는 선형 관계만으로 진단하기에는 한계가 있다(Riggs, 2014). 이러한 문제점에 기반하여 성현근(2014)의 연구는 주거지역의 건조환경과 보행 활동과의 비선형 연관성에 대한 분석을 진행하였으며 도시 물리적 환경에 대한 비선형 분석의 중요성을 강조하였다.

다른 한편으로, 최근 도시계획 분야에 활용되기 시작한 기계학습 모형은 높은 설명력과 비선형 관계 분석의 장점에도 불구하고 적용사례가 미비한 실정이다. 기계학습 방법론은 일반적으로 기존의 전통적인 회귀모형보다 높은 설명력을 가지고 있다. 그러나 알고리즘을 통해 훈련된 모형은 블랙박스로 간주하여 각 설명변수가 종속변수의 예측에 얼마나 기여하였는지 알 수 없다는 단점이 있었다. 하지만 최근 들어 딥러닝 모형을 포함한 해석 가능

한 기계학습(interpretable machine learning) 모형의 발전으로 블랙박스를 어느정도 해석할 수 있게 되면서 도시 및 교통계획 분야에서 활용되기 시작하였다(윤준호·이수기, 2019; 조월 외, 2021). 기계학습 모형의 블랙박스를 해석하는 방법에는 여러 가지가 있다. 그중 Partial Dependence Plot(PDP)은 모형에 포함된 모든 변수의 평균적인 영향을 고려하여 종속변수에 대한 설명변수의 한계효과(marginal effect)를 그래프로 설명하는 방법이다(Chung, 2013; Friedman, 2001). PDP 방법론은 설명변수와 종속변수 사이의 선형관계 또는 비선형 관계를 분석할 수 있다는 장점이 있다.

따라서 본 연구는 전통적인 회귀분석 모형과 비교하여 보행만족도 분석 모형으로 기계학습 모형의 우수성을 살펴보고, 해석 가능한 기계학습을 활용하여 보행만족도에 영향을 미치는 도시의 물리적 환경 요인을 보행목적별로 도출한다. 나아가 보행만족도에 영향을 미치는 변수의 상호작용 효과를 분석하고 보행만족도 증진을 위한 정책적 시사점을 도출한다.

II. 선행연구 고찰

1. 보행만족도 영향요인 연구

보행 활동과 도시의 물리적 환경에 대한 다양한 연구에 비교했을 때 보행만족도에 관한 연구는 국내외적으로 많지 않다(이수기 외, 2016). 보행만족도에 관한 분석을 진행한 연구 중 서울 서베이 2013 자료를 중심으로 근린환경 특성이 보행만족도에 미치는 영향을 분석한 이수기 외(2016)의 연구에서는 보행만족도에 부정적인 영향을 미치는 물리적 환경으로 Jacobs의 경계 공백변수인 철도와 고속도로의 통과 여부를 제시하였으며, 경계 공백으로 작용하는 물리적 환경이 도시민의 보행만족도에 부정적인 영향을 가지고 있음을 보였다.

다중 선형회귀모형을 활용하여 Jacobs가 강조한 물리적 환경과 보행 활동 데이터 간의 관계를 분석한 Sung et al.(2015)의 연구에서는 Jacobs가 주장하였던 도시 활력에 영향을 미치는 요소 중 일부는 시간에 따라 영향을 미치는 정도가 다름을 주장하였다. 또한, 분석 결과를 바탕으로 주로 서구 국가에서 연구되고 적용되어왔던 Jacobs의 이론이 우리나라에서도 적용될 수 있음을 주장하였다.

보행목적별 보행만족도에 영향을 미치는 근린 건조환경에 대해 분석한 김창국 외(2016)는 통근 통학 목적의 보행만족도가 그 외 목적의 보행만족도보다 높다는 결과를 도출하였으며, 도시민이 만족하는 보행환경의 특성이 보행의 목적에 따라 다르다는 것을 확인하였다. 이와 비슷한 맥락에서 보행자의 지점 방문목적별 보행만족도에 영향을 미치는 가로환경 구성요소에 대해 분석한 김창국(2015)은 보행만족도에 영향을 미치는 요인이 보행의 목적

에 따라 통계적으로 유의한 차이가 있다고 주장하였다. 추가적으로, 보행의 목적이 명확한 모형이 통합적인 목적의 모형보다 설명력이 높게 나타나므로 보행의 성격이 공통될수록 보행만족도에 영향을 미치는 변수들이 더 명확하게 나타난다고 주장하였다.

또한, 딥러닝과 구글 Street View API를 활용하여 가로 이미지의 보행만족도를 분석한 박근덕 외(2021)의 연구에서는 의미론적 분할 분석을 활용하여 가로 이미지에서 나타난 구성요소를 도출하고, 각 요소들이 보행만족도와 어떠한 상관관계를 보이는지 분석하였다. 분석 결과로는 가로 이미지에서 하늘, 나무, 식물 등의 비율은 보행만족도와 양(+)의 상관관계가 나타났으며, 반대로 건물, 벽, 차양 등은 보행만족도와 음(-)의 상관관계가 나타났다.

마지막으로, 머신러닝 및 컴퓨터 비전 기법을 활용하여 보행만족도에 영향을 미치는 가로의 물리적 특성을 분석한 이지윤(2021)의 연구에서는 보행만족도에 영향을 많이 미치는 가로경관의 시각적 요소로는 도로 면적 비율, 보도 면적 비율, 가로시설물 면적 비율 등이 나타났다. 따라서 보행만족도가 높은 가로를 조성하기 위해서는 보행자의 눈높이에서 경험하는 시각적 물리적 환경의 중요성을 알 수 있다.

2. 기계학습 방법론

기계학습은 머신러닝(machine learning)이라고도 불리며, 인공지능의 한 분야로 데이터의 패턴을 분석하여 알고리즘의 예측 오류를 최소화하는 수학 기법을 적용하여, 신뢰도 있게 결과를 예측하는 기법이다(권철민, 2019).

머신러닝에서는 전통적인 통계학으로 분석하기 힘든 특징의도 없이 수집되거나 다양한 형태를 가진 자료를 분석할 수 있으며, 비선형을 보이는 변수 사이의 복잡한 상호작용을 분석할 수 있다는 장점이 있다(유진은, 2017). 따라서 최근 들어 빅데이터 시대를 맞이해 비정형 데이터를 포함한 데이터 분석을 위해 많은 연구자들이 머신러닝 알고리즘을 활용한 데이터 해석을 진행하고 있다(이지윤, 2021).

이제승·이현희(2019)의 연구는 기계학습모형을 활용하여 보행만족도 예측 모형을 제시하였다. 로지스틱 회귀모형, 인공신경망 모형, Random Forest 모형 세 개를 비교·분석한 결과 Random Forest의 예측력이 가장 좋은 것으로 제시하였다. 나아가 Random Forest 모형 변수 중요도 분석 결과를 통해 경사도, 보도 폭, 차선 수 등이 보행만족도에 중요한 요인임을 제시하였다. Parsa et al.(2020)은 머신러닝 모형을 활용하여 교통사고의 발생을 예측하고, SHapley Additive exPlanation(SHAP) 지수로 해석을 하였으며, 99%의 모형 정확도를 달성하였다. 분석 결과로는 교통 특성, 인구통계학적 특성, 토지이용 특성 등의 순으로 사고 발생률에 영향을 미친다는 것이 도출되었다. 또한, Yang et al.(2021)의 연구에서는 머신러닝 모형 중 하나인 Gradient Boosting

Decision Tree(GBDT)를 활용하여 도시 활력과 건축환경 사이의 비선형 관계를 분석하였다. 분석 결과로는 도시 활력과 지하철역까지의 거리는 비선형적인 관계가 있으며, 약 460m를 기준으로 그보다 적은 경우에는 도시 활력이 높아지지만, 이후에는 도시 활력에 영향력이 없다는 결과가 도출되었다.

이처럼 머신러닝 방법론은 비선형 관계에 대한 데이터를 효과적으로 해석하는 것이 가능하다. 또한, 최근 들어 해석 가능한 기계학습 방법론의 활용을 통해 기계학습의 블랙박스를 해석하는 것으로 연구의 영역이 넓어지고 있다. 전통 회귀 모형과 다양한 종류의 머신러닝 모형을 비교하고, R-Squared 값과 표준절대오차(MAE), 표준제곱오차(MSE)를 비교하여 가장 우수한 머신러닝 모형을 선정하여 분석에 활용하고 있다. 나아가, 해석 가능한 기계학습을 활용하여 최종적으로 선정된 모형의 블랙박스를 해석함으로써 모형의 예측력에 영향을 미치는 요소들을 도출하고 있다. 따라서 해석 가능한 기계학습 방법론은 도시의 물리적 환경을 분석하는 데 있어 활용 가능성이 높다는 것을 알 수 있다.

3. 연구의 차별성

선행연구 고찰 결과에 따른 한계점은 다음과 같다. 먼저, 기존의 전통적인 분석 방법은 기계학습에 비해 설명력이 낮음에도 불구하고 모형의 블랙박스를 해석할 수 없다는 한계점으로 인해 설명력이 낮은 전통적인 회귀분석을 활용한 연구가 주를 이루었다. 둘째, 보행 활동은 물리적 환경과 선형의 관계를 가정하기에는 더욱 복잡한 비선형 관계가 있음에도 불구하고 선행연구의 대부분은 보행만족도에 미치는 영향요인을 다중 회귀분석을 활용하여 선형관계를 전제한 상태에서 분석이 진행되었다. 셋째, 보행만족도에 영향을 미치는 요인을 보행목적을 구분하지 않고 통합적인 보행만족도에 미치는 연구가 주를 이루었으며, 보행 목적을 구분하여 진행한 연구는 부족한 실정이다. 마지막으로 대부분의 선행연구에서는 하나의 독립변수가 보행만족도에 미치는 영향을 분석한 연구가 대부분이며, 독립변수의 상호작용 효과를 고려하여 분석한 연구는 부족한 실정이다.

이에 따른 본 연구의 차별성은 다음과 같다. 첫째, 기존의 전통적인 분석 방법에 비해 설명력이 높은 해석 가능한 기계학습을 활용하여 분석을 진행하였다. 둘째, 선형 관계에 국한되어 있는 대부분 선행연구의 한계점을 해결하기 위해 기계학습 모형의 블랙박스를 해석하여 설명변수와 보행만족도 간의 비선형적인 관계를 분석하였다. 셋째, 보행의 목적에 따라 도시의 물리적 환경이 미치는 영향력이 다르기 때문에 보행 목적을 일상보행과 운동보행으로 구분하여 보행만족도에 영향을 미치는 물리적 환경변수를 분석하였다. 마지막으로 독립변수 간의 상호작용 효과가 보행만족도에 미치는 영향을 분석하여 보행목적별 보행만족도에 미치는 시사점을 도출하였다.

III. 방법론

1. 분석자료

본 연구의 공간적 범위는 서울특별시이며 종속변수를 포함한 주요 변수는 2016년 9월 조혜민·이수기(2017)의 연구에서 조사한 “서울시민 사회적 자본과 보행활동 설문조사” 자료를 2차 가공하여 활용하였다. 설문조사는 서울특별시에 2년 이상 거주한 만 19세 이상 65세 미만의 성인 남녀를 대상으로 전문적인 설문조사 기관을 통해 온라인 설문조사 방식으로 실행되었다. <그림 1>은 연구의 공간적 범위인 서울특별시와 설문조사 응답자의 거주지를 분포를 나타낸다. 설문조사 응답자는 일상보행만족도의 경우 “귀하께서는 주거지역 인근에서 일상으로 걷기가 얼마나 만족스러우십니까?”라는 질문에서 1점의 경우 “매우 불만족”, 7점의 경우 “매우 만족”으로 구분하여 응답하였고, 운동보행만족도의 경우에는 “귀하께서는 주거지역 인근에서 운동으로 걷기가 얼마나 만족스러우십니까?”라는 질문의 7점 척도로 응답을 하였다.

분석단위는 보행만족도에 미치는 도시의 물리적 환경요인이라는 본 연구의 목적에 따라 도시 물리적 환경 분석을 위한 반경은 500m 이내 보행거리 또는 보행으로 10분 정도라는 선행연구를 참고하였다(Leslie et al., 2010; 최창규 외, 2013). 따라서 설문조사 응답자의 거주지 위치로부터 반경 500m를 기준으로 한 면적과 설문조사 지점으로부터 해당 시설물까지의 최단 거리를 기반으로 설정하였다. <표 1>은 분석에 활용된 변수에 대한 정의이며, 독립 변수로는 개인 특성변수, 가로환경 특성변수, 토지이용 특성변수, 접근성 변수, 경계 공백 지대 변수로 구분하여 구축하였다. 먼저 개인 특성 변수는 설문 응답자의 성별, 나이, 일상생활 보행시간, 운동보행시간으로 구축하였다.

가로환경 특성변수는 가로환경 요소의 시각적 특성과 보행만족도와와의 관계를 분석한 박근덕 외(2021)의 연구를 참고하여 보

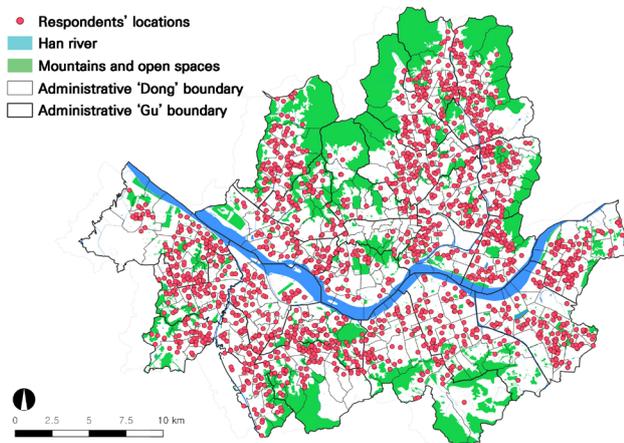


그림 1. 연구의 공간적 범위 및 설문 응답자 주소지 위치

Figure 1. The case study area and respondents' address locations

행만족도와 밀접한 관련이 있는 것으로 도출된 몇 가지 요인 중 가로환경의 건물 비율과 녹지 비율을 구축하여 활용하였다. 또한, 하늘의 비율이 보행만족도에 영향을 미칠 수 있을 것이라고 판단하여 분석 변수에 추가하였다. 가로환경의 요소별 비율을 산출하기 위해 서울시 가로경관 빅데이터 세트인 2016년 Naver Street View(NSV) 파노라마 이미지(Panorama Image) 41,140개를 구축하였다. NSV 이미지 구축은 설문조사 년도인 서울시의 2016년 도로망을 기준으로 50m 간격으로 지점을 생성하고, 각 지점별 파노라마 이미지를 구축하여 분석에 활용하였다.

이후, 이미지 내의 가로환경 요소를 추출하기 위해 컴퓨터 비전 기법의 하나인 딥러닝 의미론적 분할 기법을 활용하였다. 의미론적 분할 기법은 건물, 녹지, 하늘, 차량 등 이미지 내에 존재하는 요소를 픽셀 단위로 구분해주는 기법이다. 이러한 의미론적 분할의 모형으로는 Deeplabv3+, U-Net, PSPNet, ReSeg, FCN8s 등이 있으며, 본 연구에서는 2015년 구글의 개발 이후로 현재까지 지속적인 발전을 통해 우수한 성능을 나타내는 Deeplabv3+ 모형을 활용하였다(Chen et al., 2018). Deeplabv3+ 모형을 활용하여 의미론적 분할을 진행한 예시는 <그림 2>와 같다.

해당 모형에서는 차량, 보행자, 하늘, 녹지, 펜스, 벽, 빌딩, 보행로 도로 등 19개의 항목을 분류하였지만, 최종적으로는 앞서 설명한 바와 같이 건물, 녹지, 하늘의 비율을 분석에 활용하였다. 분할이 진행된 이미지를 대상으로 이미지 전체 픽셀 수 대비 해당 가로환경요인 이미지 픽셀 수로 계산하여 변수를 구축하였으며, 활용된 수식은 다음과 같다.

$$\text{가로환경요인 비율} = \frac{\text{가로환경요인으로분류된 픽셀 수}}{\text{전체 이미지 픽셀 수}} \times 100$$

즉, 해당 가로환경요인 비율은 분류된 픽셀이 존재하지 않을 경우 0, 모든 픽셀이 해당 가로환경요인으로 분류될 경우 최대 100의 값을 가지며, 값이 클수록 해당하는 지점에서 시각적으로 보이는 요인의 비율이 높다는 것으로 해석할 수 있다(기동환 외, 2021).

토지이용 특성변수로는 주거시설 연면적, 상업시설 연면적, 업

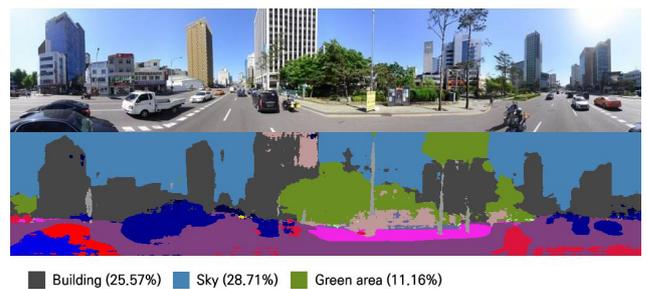


그림 2. 2016년 NSV 이미지 의미론적 분할 결과 예시

Figure 2. An example of the 2016 NSV image semantic segmentation results

표 1. 변수의 정의 Table 1. Description of variables

Variables		Description of variables	Sources	
종속변수 Dependent variable	일상보행만족도 Daily walking satisfaction	Daily walking satisfaction (1-7 points scale)		
	운동보행만족도 Exercise walking satisfaction	Exercise walking satisfaction (1-7 points scale)		
개인특성 Personal characteristics	일상보행시간 Daily walking time (min.)	Average daily walking time (minute) per day × Average number of daily walking days per week	Survey (2016)	
	운동보행시간 Exercise Walking time (min.)	Average exercise walking time (minute) per day × Average number of exercise walking days per week		
	성별 Gender	Male (0), Female (1)		
	나이 Age	Age of survey respondents		
가로환경특성 Street environment characteristics	건물비율 Percentage of buildings	Percentage (%) of buildings within a 500 m radius from a survey location	Deep Learning semantic segmentation using Naver street view images (2016)	
	하늘비율 Percentage of sky	Percentage (%) of sky within a 500 m radius from a survey location		
	녹지비율 Percentage of green area	Percentage (%) of green area within a 500 m radius from a survey location		
독립변수 Independent variable	토지이용혼합도 Land use mix	Entropy index of residential, commercial, and office facilities within a 500 m radius from a survey location	National spatial information portal (2017)	
	주거시설 연면적(km ²) Total floor area of residential facilities (km ²)	Total floor area (km ²) of residential facilities within a 500 m radius from a survey location		
	상업시설 연면적(km ²) Total floor area of commercial facilities (km ²)	Total floor area (km ²) of commercial facilities within a 500 m radius from a survey location		
	오피스시설 연면적(km ²) Total floor area of office facilities (km ²)	Total floor area (km ²) of office facilities within a 500 m radius from a survey location		
	교차로 개수 Number of intersections	Number of intersections within a 500 m radius from a survey location		National transportation DB (2016)
	건물연령 다양성 Diversity of building age	Standard deviation of building age within a 500 m radius from a survey location		National spatial information portal (2017)
	평균 건물 연령 Average building age	Average building age within a 500 m radius from a survey location		
접근성특성 Accessibility characteristics	버스정류장 접근성 Accessibility to bus stop	Distance (m) from the survey location to the nearest bus stop	Seoul open data (2018)	
	지하철역 접근성 Accessibility to subway station	Distance (m) from the survey location to the nearest subway station	New address DB (2016)	
경계공백특성 Border vacuum characteristics	대규모 건물 연면적(km ²) Total floor area of large buildings (km ²)	Total floor area (km ²) of large buildings (top 15% large buildings in Seoul) within a 500 m radius of a survey location	National spatial information portal (2017)	
	대규모 공원 유무 Existence of large parks	Existence of large parks (top 15% of park size in Seoul) within a 500 m radius of a survey location	National spatial information portal (2016)	
	하천(강) 유무 Existence of rivers	Existence of rivers within a 500 m radius of survey location		
	고속도로 유무 Existence of highways	Existence of highways within a 500 m radius of survey location	New address DB (2016)	

무시설 연면적을 포함하였다. 이와 더불어 주거시설, 상업시설, 업무시설의 혼합도 또한, 보행만족도에 영향을 미칠 수 있기 때문에 세 용도의 혼합도를 엔트로피 지수로 계산하여 토지이용 혼합도 변수를 구축하였다. 추가적으로 Jacobs가 도시 활력에 영향을 주는 필수적인 요소로 제시한 블록 크기도 보행만족도에 영향을 미칠 수 있기 때문에 블록의 크기를 나타낼 수 있는 교차로의 개수를 변수로 구축하여 추가하였다. 마찬가지로 다양한 연령의 건물 또한, Jane Jacobs가 언급한 도시 활력에 영향을 주는 필수적인 요소 중 하나였기 때문에 건물 연령의 표준편차와 평균값을 계산하여 변수에 추가하였다.

다음으로 일상 및 운동보행만족도에 있어 대중교통과의 접근성도 영향을 미칠 수 있다. 이수기 외(2014)의 연구는 설문조사자의 거주 지점으로부터 근거리에서 위치한 다양한 활동 공간은 보행만족도에 영향을 미칠 수 있음을 시사하였다. 따라서 대표적인 대중교통인 버스정류장과 지하철역 접근성 변수를 분석에 포함하였다. 이때 접근성은 설문조사자의 거주 지점으로부터 가장 근접한 해당 교통시설과의 거리(m)를 측정하였다. 대표적인 대중교통수단인 버스정류장, 지하철 변수를 통해 도시 공간 내 주요 교통시설과의 접근성을 분석하는 데 목적이 있다.

마지막으로 경계 공백 지대(border vacuums) 특성을 설명하고자 사용된 4가지 변수는 다음과 같은 이유로 선정하였다. 우선 대규모 건물(건물 면적 상위 15%) 유무와 대규모 공원(공원 면적 상위 15%) 유무 변수는 Jacobs가 지적한 도시 내 대규모 단일용도 토지이용을 측정하는 데 목적이 있다. Jacobs에 따르면 대규모 단일용도 건물, 대규모 공원 등은 해당 지역뿐만 아니라 인접한 공간들의 보행 목적을 단일로 획일화시킬 가능성이 높기 때문에 대도시에서 이 같은 경계 공백 지대를 지양해야 함을 주장했다.

경계 공백 지대가 소매업 매출에 미치는 영향에 대해 분석한 이동엽·성현곤(2020)의 연구를 참고하여 서울시 전체 건축물들의 연면적 중 상위 15% 건축물들이 설문조사 지점으로부터 500m 반경에 있는지, 서울시 전체 공원 중 상위 15% 공원들이 설문조사 지점으로부터 500m 반경에 있는지 여부를 더미 변수로 구축하여 활용하였다.

또한, 경계 공백 지대 특성을 분석하는 데 있어 하천 또는 강은 보행자의 통행이 어려운 편도 4차선 이상의 도로, 고속도로, 지상철도와 함께 보행을 단절시키는 요인으로 작용할 수 있다(이동엽·성현곤, 2020). 따라서 설문조사 지점 반경 500m 이내 하천이 존재하는지의 여부를 더미 변수로 구축하여 분석에 활용하였다. 이와 마찬가지로 고속도로 접근성의 경우에도 보행의 단절시키는 요인으로 작용할 수 있어, 2016년 도로망을 기준으로 고속도로, 도시 고속화도로, 고속도로 연결 램프 도로를 추출하여 설문조사 지점으로부터 반경 500m 내 유무를 더미 변수로 구축하여 활용하였다.

2. 분석 방법론

도시의 물리적 환경이 도시민의 보행만족도에 미치는 영향에 대한 대부분의 기존 선행연구 방법은 가장 보편적인 방법으로서 최소제곱법(OLS) 회귀모형을 사용하였다(이수기 외, 2016). 하지만 앞서 언급했듯이 도시의 물리적 환경은 선형의 관계로써 진단하기에는 간단하지 않다(Riggs, 2014). 따라서 보행만족도와 도시 물리적 환경과의 비선형 관계를 분석하기 위하여 인공지능의 주요 분야 중 하나인 기계학습(machine learning) 방법론이 유용하게 활용될 수 있다.

본 연구에서는 머신러닝의 여러 모형 중 보행만족도를 가장 잘 예측하는 모형을 선정하기 위해 현재 가장 많이 사용되고 있는 몇 가지 모형을 선정해 각각의 성능을 전통 회귀모형과 비교하였다. 선정된 모형은 Random Forest, Light Gradient Boost, CAT-Boost, Extreme Gradient Boost(XGBoost), Gradient Boost 모형이다. 선정된 모형들을 평가하기 위한 방법으로서 선행연구를 참고하여 주로 사용하고 있는 평가지표인 결정계수(R^2), 평균제곱오차(Mean Squared Error), 평균절대오차(Mean Absolute Error)의 3가지를 활용하여 모형을 평가하였다(Chen et al., 2020; 조월 외, 2021).

이후, 각 변수의 중요성을 확인하기 위해 SHAP 방법론을 활용하여 Shapley Value(기여도)를 도출하였다. 변수의 기여도는 Global Shapley Value와 Local Shapely Value로 나누어진다. 먼저, Global Shapely Value의 경우 변수의 절댓값 크기가 큰 순서대로 차례로 배치되어 있다. 각 변수별로 Shapley 값이 양수인 경우에는 붉은색, 음수일 경우에는 푸른색으로 표현된다.

Local Shapley Value의 경우에는 Global Shapley Value를 보완하여 모형의 예측력에 대해 변수의 영향 정도를 방향 및 크기로 표현한다(Lundberg et al., 2018). 변수의 순서는 Global Shapley Value와 마찬가지로 모형에 영향을 미치는 크기순으로 배열되며, 변수의 값이 높을수록 붉은색, 변수의 값이 낮을수록 푸른색의 점으로 나타난다. 이후 X축의 0 값을 기준으로 오른쪽은 양(+), 왼쪽은 음(-)의 관계 방향으로 모형의 성능 평가에 기여한다.

하지만 SHAP 방법론만으로는 물리적 환경변수가 가지는 복잡한 비선형 관계는 구체적으로 설명할 수 없다는 한계가 존재한다. 따라서 PDP를 활용하여 각각의 독립변수가 보행만족도와 어떠한 연관성을 가지고 있는지 분석하였다. 따라서 기계학습모형을 기반으로 변수의 PDP 그래프를 생성하고 해당 독립변수와 보행만족도 사이의 관계를 분석하였다.

IV. 분석결과

1. 기술 통계 분석

본 연구에서 사용된 변수들의 기술 통계량은 <표 2>와 같다. 종속변수는 일상보행만족도, 운동보행만족도이며 독립변수는 개인 특성, 가로환경 특성, 토지이용 특성, 접근성 특성이 해당한다. 기초통계량은 각 변수에 대한 평균, 표준편차, 최솟값, 최댓값으로 구성되어 있다. 종속변수는 일상보행만족도와 운동보행만족도이며 설문조사 항목의 1점부터 7점까지의 척도를 의미한다. 일상보행만족도의 평균은 4.71점, 운동보행만족도는 4.78점으로 두 변수 간 평균점수는 차이는 크지 않은 것으로 나타났다.

개인 특성 중 일상보행시간과 운동 시 보행시간은 보행목적별

로 일주일 동안 보행한 시간(분)을 의미하며 일상생활 시 보행에 소요하는 평균 시간은 약 260분으로 운동 시 평균적으로 소요하는 평균 시간인 약 181분보다 높은 것으로 나타났다. 이는 운동 시 보행하는 시간보다 일상생활에서 보행하는 시간이 평균적으로 높다는 것으로 해석된다. 가로환경 특성 중 가로 빌딩의 비율이 평균적으로 가장 높게 나타났으며 녹지 비율이 가장 낮은 것으로 나타났다. 이는 본 연구의 공간적 범위가 대도시인 서울이라는 점이 반영된 것으로 보인다.

토지이용 특성 중 토지이용 혼합도는 주거, 상업, 업무지역의 토지이용 혼합도를 측정된 것으로 기술통계 분석표에 따르면 토지이용 혼합도의 평균값은 약 0.49로 나타났으며 주거건물 연면적의 평균값이 상업과 업무건물의 연면적보다 높은 것으로 나타났다. 접근성 특성은 설문조사 지점으로부터 가장 근접한 교통시

표 2. 기술 통계 분석 Table 2. Descriptive statistical analysis

Variables	Obs.	Mean, %	S.D	Min.	Max.	
종속변수 Dependent variable	일상보행만족도 Daily walking satisfaction	2,501	4.71	1.17	1	7
	운동보행만족도 Exercise walking satisfaction	2,501	4.78	1.22	1	7
개인특성 Personal characteristics	일상보행시간 Daily walking time (min.)	2,501	260.54	290.69	0	4,200
	운동보행시간 Exercise walking time (min.)	2,501	180.99	288.93	105	6,000
	성별 Gender	2,501	Female (1) 59.78% Male (0) 40.22%	-	0	1
	나이 Age	2,501	44.33	10.20	20	64
가로환경 특성 Street environment characteristics	건물비율 Percentage of buildings	2,501	27.66	17.55	0.00	73.22
	하늘비율 Percentage of sky	2,501	9.61	8.29	0.00	45.07
	녹지비율 Percentage of green area	2,501	7.62	6.03	0.00	35.61
토지이용 특성 Landuse characteristics	토지이용혼합도 Land use mix	2,501	0.49	0.23	0.00	1.00
	주거시설 연면적(km ²) Total floor area of residential facilities (km ²)	2,501	0.67	0.29	0.00	1.80
	상업시설 연면적(km ²) Total floor area of commercial facilities (km ²)	2,501	0.12	0.10	0.00	1.10
	오피스시설 연면적(km ²) Total floor area of office facilities (km ²)	2,501	0.07	0.15	0.00	1.37
	교차로 개수 Number of intersections	2,501	429.37	294.62	2	1,604
	건물연령 다양성 Diversity of building age	2,501	11.31	3.18	1.03	23.14
	평균 건물 연령 Average building age	2,501	25.68	4.91	2.20	43.32

(Continue on next page)

Variables		Obs.	Mean, %	S.D	Min.	Max.
접근성 특성 Accessibility characteristics	버스정류장 접근성 Accessibility to bus stop	2,501	120.44	73.49	4.95	606.89
	지하철역 접근성 Accessibility to subway station	2,501	577.66	361.97	46.51	2,919.70
대규모 건물 연면적(km ²) Total floor area of large buildings (km ²)		2,501	0.90	0.43	0.00	2.91
경계 공백 특성 Border vacuum characteristics	대규모 공원 유무 Existence of large parks	2,501	Yes (1) 28.47%	-	0	1
			No (0) 71.53%			
	하천(강) 유무 Existence of rivers	2,501	Yes (1) 18.31%	-	0	1
			No (0) 81.69%			
고속도로 유무 Existence of highways	2,501	Yes (1) 45.62%	-	0	1	
		No (0) 54.38%				

설과의 거리(m)를 의미한다. 분석 결과로는 도시 내 주요 대중교통 시설인 버스정류장의 접근성의 평균거리는 120.44m로 나타났는데, 이는 지하철 접근성 평균 거리인 577.66m보다 짧아 버스정류장으로의 접근성이 더 좋은 것으로 해석된다.

마지막으로, 경계 공백 지대 변수는 고속도로의 경우에는 약 3:7의 비율로 반경 500m 내 고속도로가 존재하지 않은 경우가 훨씬 많은 것으로 나타났다. 이와 마찬가지로 대규모 공원의 경우에도 약 2:8 정도의 비율로 반경 500m 내 대규모 공원이 존재하지 않는 경우가 훨씬 많은 것으로 나타났다. 하지만 하천의 경우에는 하천이 있을 경우 45.62%, 없을 경우 54.38%로 하천의 유무가 보행만족도에 미치는 영향을 잘 설명할 수 있을 것으로 판단된다. 마지막으로 대규모 건물 연면적의 경우에는 반경 500m 이 내 평균 약 0.9km²의 대규모 건물이 존재하는 것으로 나타났다.

2. 기계학습 모형

기계학습 모형 간의 설명력을 객관적으로 평가하기 위해 수집된 데이터 2,510개 중 이상치에 해당하는 일주일 평균 보행시간이 6,000분을 초과하는 데이터 6개와 가로 건물 비율이 약 94.5%인

데이터 3개를 제거한 2,501개의 데이터를 학습 데이터와 평가 데이터로 분할하였다. 이때, 학습 표본은 전체의 80%인 2,001개, 평가 데이터는 20%인 500개로 할당되었다. 이후, 과적합을 방지하기 위해 각각 모형의 학습에 영향을 미치는 주요 매개변수를 조정하였다. 최종적으로 학습 데이터를 통해 각 모형을 추정하도록 하였으며, 측정된 모형들에 대한 평가 데이터의 독립변수를 관찰한 후 종속변수인 보행만족도를 예측하도록 하였다. <표 3>은 앞서 언급한 5가지 기계학습 모형과 전통회귀모형을 평가한 결과이다.

<표 3>의 결과에 따르면 기계학습 모형의 평가 결과는 일반 선형회귀 모형에 비해 좋은 성능을 보여주고 있는 것으로 나타났다. 또한, 5가지 기계학습 모형 중 Gradient Boost 모형은 일반 회귀 모형의 예측력인 10%, 15%에 비해 43%, 52%로 모형 설명력이 향상된 것을 알 수 있다. 따라서 Gradient Boost 모형이 도시 물리적 환경으로 도시민의 보행만족도를 예측할 때 가장 적절한 것으로 나타났다. Gradient Boost 모형은 경사 하강법을 활용한 모형으로서 손실함수를 정량화해주는 역할을 한다. 이러한 손실함수를 파라미터로 미분하여 기울기 값을 산출하고, 이 손실(loss) 값이 작아지는 방향으로 파라미터를 움직이게 하는 방법이다. 즉, Gradient Boost 모형은 전 단계에서 학습된 모형

표 3. 모형 평가 결과 Table 3. Model evaluation results

모형 Model	일상보행만족도 Daily walking satisfaction			운동보행만족도 Exercise walking satisfaction		
	R ²	MSE	MAE	R ²	MSE	MAE
Linear Regression	0.10	1.261	0.888	0.15	1.228	0.875
Random Forest	0.23	1.102	0.818	0.28	1.033	0.781
Light Gradient Boost	0.28	1.026	0.727	0.30	1.044	0.712
CATBoost	0.38	0.886	0.700	0.44	0.811	0.672
XGBoost	0.40	0.863	0.640	0.45	0.797	0.606
Gradient Boost	0.43	0.823	0.613	0.52	0.695	0.546

의 잔차를 다음 모형에서 학습하여 보완을 하는 방식의 모형이다 (Friedman, 2001). 따라서 본 연구에서는 Gradient Boost 모형이 가장 적합한 모형이라고 판단하고 최종 분석에 활용하였다. Gradient Boost 모형의 주요 매개변수로는 모형의 최대 깊이를 설정하는 max_depth, weak learner의 개수를 조정하는 n_estimators, 모형이 학습을 진행할 때마다 학습률 조정을 위해 적용하는 learning_rate, weak learner가 학습에 사용하는 데이터 샘플링의 비율을 조정하는 subsample 등이 있다. 본 연구에서 최적화된 매개변수의 값을 찾기 위해 Grid Search의 방식을 활용하였다. 조정한 매개변수의 값은 일상 및 운동 보행만족도 모두 max_depth = 15, n_estimators = 1000, learning_rate = 0.01, subsample = 0.6으로 설정하였다.

3. 변수의 기여도 분석

〈그림 3〉과 〈그림 4〉는 본 연구에서 활용한 Gradient Boost 모형에 대한 Global Shapley Value와 Local Shapley Value에 대한 결과이다. Global Shapley Value에서 붉은색으로 표현된 변수일 경우에는 보행만족도와 양(+)의 관계를 가지고 있음을 뜻하며, 반대로 푸른색으로 표현된 변수의 경우에는 보행만족도와 음(-)의 관계를 가지고 있음을 뜻한다(그림 3, 4). 본 연구에서 활용된 기계학습 모형인 Gradient Boost 모형을 기반으로 모형을 학습시키고, 플롯을 구성하는 점들은 각각 하나의 설문조사 응답 데이터 및 독립변수 데이터, 즉 입력 데이터 샘플 값이며, 여러 개의 점들이 모여 오른쪽이나 왼쪽으로 만드는 긴 꼬리 모양은 특정

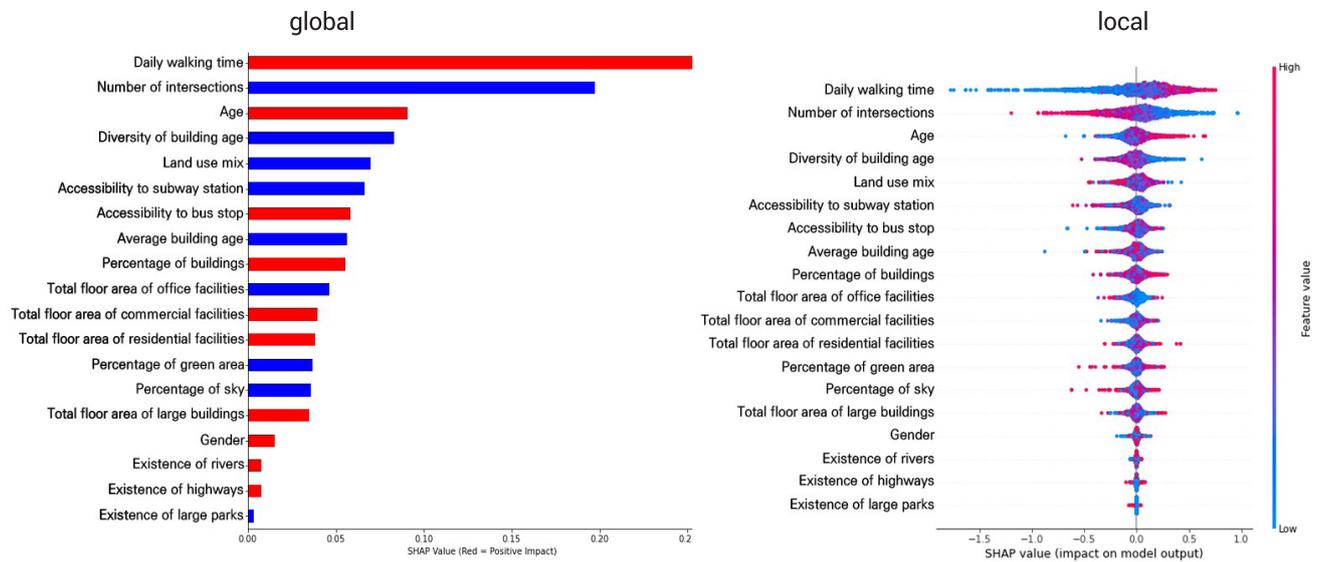


그림 3. 일상보행만족도에 대해 'SHAP'로 도출한 각 변수의 중요도
 Figure 3. Importance of each variable derived by 'SHAP' on the daily walking satisfaction

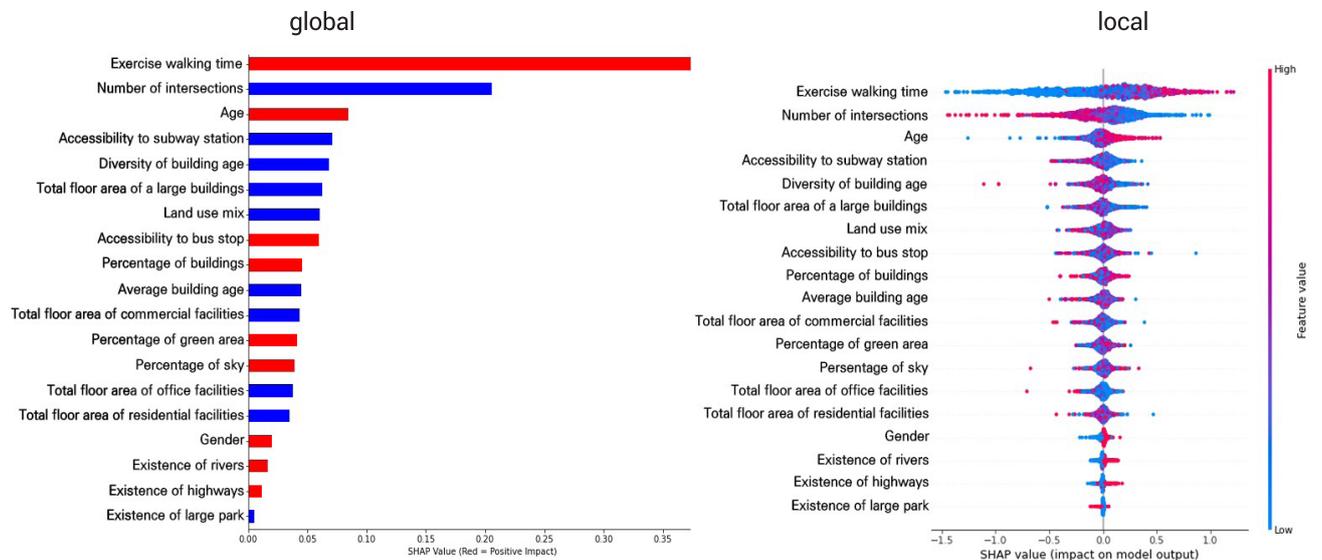


그림 4. 운동보행만족도에 대해 'SHAP'로 도출한 각 변수의 중요도
 Figure 4. Importance of each variable derived by 'SHAP' on exercise walking satisfaction

개인에게 극단적인 측정값이 중요할 수 있음을 뜻한다(Lundberg et al., 2018; 이지운, 2021).

분석 결과 일상보행만족도와 운동보행만족도 두 경우 모두에서 보행시간 및 도시의 물리적 환경이 모형의 예측력에 큰 영향을 미치고 있음을 알 수 있었다. 이후, 물리적 환경의 경우에는 일상보행만족도의 경우 교차로 개수, 건물 연령 다양성, 토지이용 혼합도 등의 순으로 영향을 미치고 있었으며, 운동보행만족도의 경우에는 교차로 개수, 지하철 접근성, 건물 연령 다양성 등의 순으로 모형의 설명력에 영향을 미치고 있었다.

특히 교차로 개수의 경우에는 Local Shapley Value 값을 보았을 때 높은 값을 가지는 붉은색의 점들이 음(-)의 방향으로 길게 나타났다. 이는 교차로의 개수가 많을수록 즉, 블록의 크기가 작을수록 보행만족도에 매우 부정적인 영향을 미친다는 것을 알 수 있다. 이는 가로활력과 관련하여 Jacobs(1961)가 언급했던 작은 크기의 블록과는 반대의 결과이며, 블록의 크기가 작을수록 교차로의 횡단 등으로 인한 보행의 연속성을 저해하기 때문이라고 판단된다. 또한, 건물 연령 평균과 건물 연령 다양성 변수도 일상보행만족도와 운동보행만족도 두 경우 모두에서 음(-)의 관계가 나타났다. 건물 연령의 다양성이 낮을수록, 건물 연령 평균이 낮을수록 보행만족도가 높아지며, 건물 연령의 영향력이 보행만족도에 미치는 정도가 크게 나타났다.

토지이용 혼합도를 살펴보면 Global Shapley Value의 경우에 일상보행만족도와 운동보행만족도 두 가지 경우 모두에서 음(-)의 관계가 있는 것으로 도출되었다. 그러나 Local Shapley Value의 경우에는 붉은색 점과 푸른색 점이 혼합된 정도가 높은 것으로 보아 비선형 관계가 있음을 시사한다. 이와 관련하여 상업시설 연면적과 업무시설 연면적의 경우에도 운동보행만족도와 일상보행만족도 모두에 음(-)의 관계를 가지는 것으로 나타났다. 특히, 상업시설 연면적의 경우에도 Local Shapley Value의 붉은색 점과 푸른색 점이 혼합되어 나타나는 것으로 보아 비선형 관계가 있음을 알 수 있다.

경계 공백 지대 변수의 경우에는 대부분 변수가 보행만족도에 미치는 영향이 적은 것으로 나타났다. 먼저 일상보행만족도의 경우 대규모 건물 연면적, 고속도로 유무, 하천 유무, 대규모 공원 유무를 제외한 대규모 건물 연면적, 고속도로 유무, 하천 유무의 경우에는 일상보행만족도와 양(+)의 관계를 가지는 것으로 나타났다. 이는 고속도로, 대규모 공원, 하천이 있을 경우 보행만족도가 높다는 것을 뜻한다. 이러한 결과는, 가로활력과 관련된 Jacobs의 주장과 상충하는 부분이 있지만, 그래프에서 붉은색 점과 푸른색 점의 혼합도가 높은 것을 보아 비선형 관계를 가지고 있는 것으로 판단된다.

운동보행만족도의 경우에는 대규모 건물 연면적, 하천 유무, 고속도로 유무, 대규모 공원 유무 순으로 변수의 영향력이 높은

것으로 도출되었다. 하천 유무와 고속도로 유무의 경우에는 양(+)의 관계를 가지는 것으로 나타났으며, 대규모 건물 연면적과 대규모 공원 유무의 경우에는 음(-)의 관계를 가지는 것으로 나타났다. 하천과 고속도로가 있을 경우 보행만족도가 높아지며, 대규모 건물 연면적이 작을수록, 대규모 공원이 없을 경우 보행만족도가 높아짐을 의미한다. 대규모 건물 연면적의 경우 운동보행만족도에서 변수 기여도 상위를 차지하고 있어 우선적인 제어가 필요할 것으로 판단된다. 하지만, 나머지 경계 공백 지대 변수의 경우에도 일상보행만족도와 마찬가지로 변수의 중요도가 낮아 향후 추가적인 연구가 필요할 것으로 예상된다.

4. PDP를 활용한 비선형 관계 분석

〈그림 5〉는 Gradient Boost 모형을 기반으로 변수의 Partial Dependence Plot(PDP) 그래프를 생성한 결과이다. X축은 해당 독립변수의 값이며, Y축은 보행만족도를 나타낸다. 전체적으로 대부분의 물리적 환경변수와 보행만족도 사이에는 비선형적인 연관성이 있는 것으로 도출되었다. 먼저, 교차로 개수와 토지이용 혼합도의 경우에는 일상보행만족도와 운동보행만족도에서 비슷한 형태의 결과가 도출되었다. PDP 그래프는 앞서 분석한 Local Shapley Value에서 붉은 점과 푸른 점의 혼합도가 높아 비선형적 관계가 있을 것으로 판단되는 변수를 중심으로 결과를 도출하였다.

앞선 Shapley Value 해석에서 언급했던 것처럼 교차로의 개수가 많을수록 보행만족도는 낮아지는 것으로 나타났다. 토지이용 혼합도의 경우에는 일정한 수준을 유지하다가, 약 0.4~0.6 정도의 값에서 최댓값을 나타낸 후 점점 보행만족도가 감소하는 비선형 관계가 나타났다. 이러한 패턴은 일상보행만족도와 운동보행만족도에서 유사하게 나타났다. 이는 적당한 수준의 토지이용 혼합도가 일상 및 운동보행만족도에 긍정적인 영향을 미치고 있음을 시사한다. 건물 연령 다양성의 경우에는 다양성이 높을수록 보행만족도는 감소하는 결과가 도출되어 노후 건물의 재정비 및 외관 개선은 도시민의 보행만족도에 긍정적인 효과가 있을 것으로 판단된다.

대규모 건물 연면적의 경우에는 일상보행만족도에 유의한 영향을 미치지 않았지만, 운동보행만족도의 경우에는 500미터 반경 안에 대규모 건물 연면적이 0.6km² 이하일 때 보행만족도가 가장 높게 나타났다. 주거시설 연면적의 경우 보행목적에 따라 차이가 있는 것으로 나타났다. 주거시설 연면적이 약 0.6km² 이상일 때 일상보행만족도는 높게 나타났지만 운동보행만족도는 꾸준히 감소하는 형태로 나타났다. 이는 운동 목적의 보행은 주거시설의 연면적이 적을수록 보행만족도가 높다는 것을 의미한다. 상업시설 연면적의 경우에는 일상보행만족도와 운동보행만족도 사이에 반대의 결과가 도출되었다. 먼저 일상보행만족도의 경우

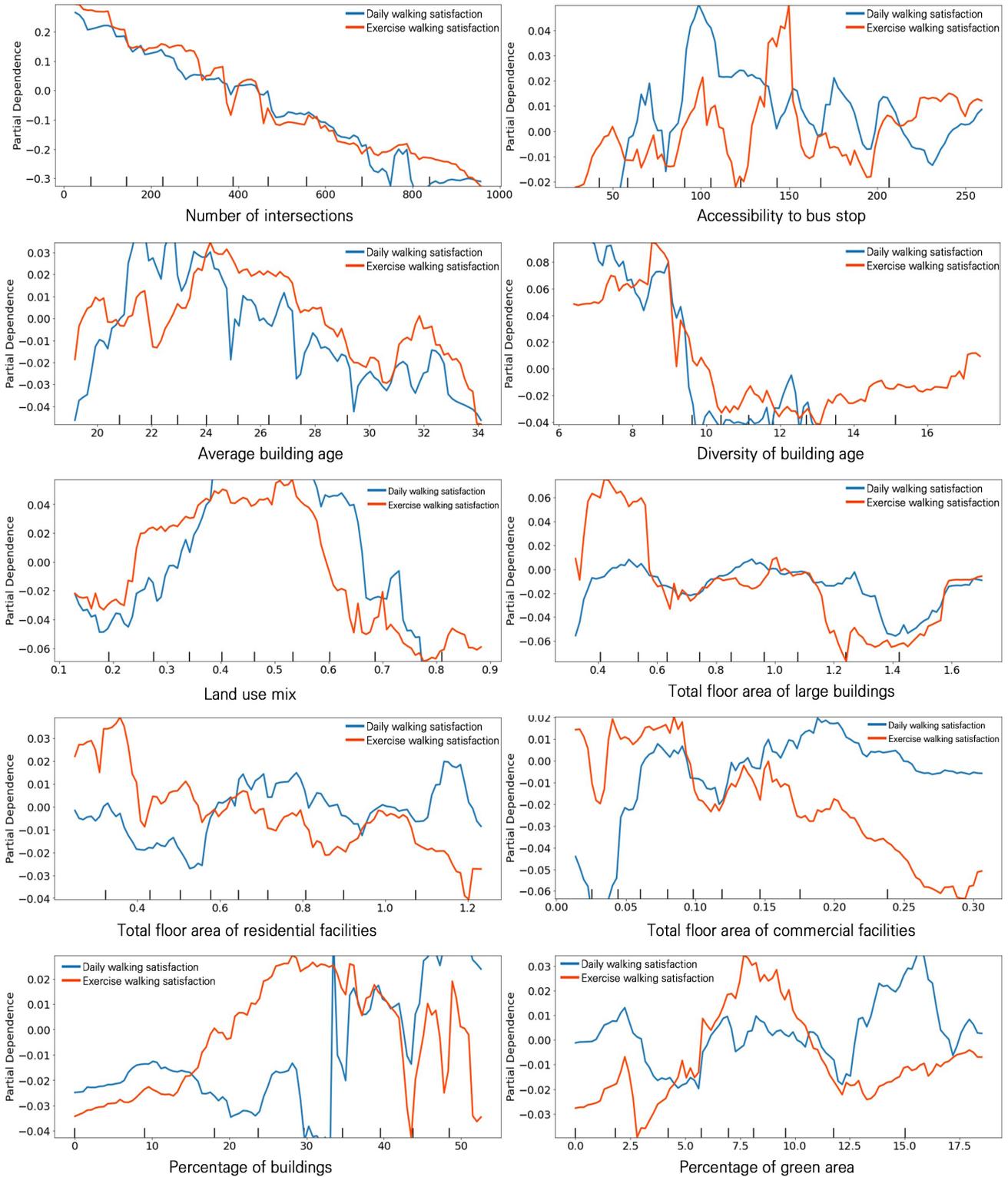


그림 5. 보행만족도 Partial Dependence Plot (PDP) 분석 결과
 Figure 5. Partial Dependence Plot (PDP) analysis results of walking satisfaction

에는 상업시설 연면적이 높을수록 보행만족도가 높아지는 형태를 나타내다 약 0.2km²를 기점으로 서서히 낮아지는 결과가 도출되었다. 반면, 운동보행만족도의 경우에는 상업시설 연면적이 높아질수록 보행만족도는 점점 감소하는 형태를 보여 운동 시에는 상업시설의 연면적이 보행만족도에 부정적인 영향을 미친다

는 점을 시사한다.

가로 건물 비율의 경우에는 일상보행만족도의 경우 일정한 수준을 보이다 약 30%를 넘어서면서부터 보행만족도가 급격하게 증가하는 형태를 보이고 있다. 하지만 운동보행만족도의 경우에는 건물의 비율이 약 30% 정도일 때 보행만족도가 가장 높다는

결과가 나타났으며, 적당한 정도의 건물 비율이 운동 시 보행만족도에 긍정적인 영향을 미칠 수 있음을 시사한다. 마지막으로 녹지 비율의 경우에는 일상보행만족도의 경우 어느 정도 일정한 수준을 유지하다가 녹지 비율이 약 15%일 때 보행만족도가 가장 높음을 알 수 있다. 하지만 운동보행만족도의 경우에는 녹지 비율이 높을수록 보행만족도는 높아지는 경향을 보이다가, 녹지 비율이 약 7.5%를 넘어서면서부터는 보행만족도가 급격하게 감소하고 있음을 알 수 있다. 따라서 이는 일상생활의 경우에는 녹지의 비율이 높을수록 보행만족도는 높아지지만, 운동 시에는 약 7.5% 정도의 적당한 녹지 비율이 보행만족도에 긍정적인 영향을 미치고 있음을 시사한다. 이는 건물 비율이 높을수록 보행만족도가 감소하며, 녹지 비율이 높을수록 보행만족도는 높아진다는 박근덕 외(2021)의 연구와는 달리 건물과 녹지의 비율은 보행의 목적에 따라 보행만족도와 비선형적인 관계가 있음을 시사한다.

5. 상호작용 효과 분석

다음 <그림 6>은 독립변수의 상호작용 효과에 대한 보행만족도와의 관계를 그래프로 나타낸 결과이다. 상호작용 효과는 토지이용 혼합도와 상업시설 연면적, 대규모 건물 연면적과 주거시설 연면적, 가로 녹지 비율과 가로 건물 비율의 세 가지 조합에 집중하여 분석하였다. 우선 토지이용 혼합도와 상업시설 연면적 조합의 경우 상업시설의 연면적에 따른 토지이용 혼합도가 일상 및 운동보행만족도에 미치는 영향을 분석하기 위해서 선정하였다. 두 번째, 대규모 건물 연면적과 주거시설 연면적 조합의 경우 주거 연면적 수준에 따른 대규모 건물 연면적이 경계 공백의 역할을 하여 일상 및 운동보행만족도에 미치는 영향을 분석하기 위해 선정하였다. 마지막으로 가로 건물 비율과 녹지 비율의 상호작용은 일상보행만족도와 운동보행만족도에 서로 다른 영향을 미칠 것으로 판단되어 선정하였다.

먼저, <그림 6>의 (a)와 (a')는 토지이용 혼합도와 상업시설 연면적과의 관계가 보행만족도에 미치는 영향을 나타낸 결과이다. 토지이용 혼합도의 경우에는 앞선 PDP 결과와 같이 일상보행만족도와 운동보행만족도 모두 약 0.4~0.6 범위의 토지이용혼합도 수준일 때 보행만족도가 높게 나타났다. 하지만 비슷한 수준의 토지이용혼합도에서도 상업시설 연면적이 클수록 일상보행만족도는 높아지는 형태였으며, 이와 반대로 운동보행만족도는 상업시설 연면적이 작을수록 높아지는 형태를 나타내고 있다. 이는 일상보행 시에는 상업시설 연면적이 높은 수준의 적당한 토지이용 혼합도에서 보행만족도가 높으며, 운동 시에는 그 반대의 경우일 때 보행만족도가 높아 같은 토지이용 혼합도라도 보행목적에 따라 혼합되는 비율을 다르게 해야 함을 시사한다.

두 번째, <그림 6>의 (b)와 (b')는 주거시설 연면적과 대규모 건물 연면적이 보행만족도에 미치는 영향에 대한 그래프이다. 먼

저, 일상보행만족도의 경우에는 대규모 건물과 주거시설 연면적이 높을수록 보행만족도는 높다는 결과가 도출되었다. 이와는 반대로, 운동보행만족도에서는 대규모 건물과 주거시설 연면적 모두 작을수록 보행만족도는 높다는 결과가 도출되었다. 이는 일상생활에서는 대규모의 주거시설이 많이 분포할수록 보행만족도는 높아지며, 반대로 운동 시에는 주거시설 연면적이 높은 지역일수록 보행만족도가 감소하는 것을 의미한다.

마지막으로, <그림 6>의 (c)와 (c')는 가로 녹지 비율과 가로 건물 비율이 보행만족도에 미치는 영향에 대한 그래프이다. 먼저 일상보행만족도의 경우에는 가로녹지 비율이 약 15%, 가로 건물 비율이 높을수록 보행만족도는 높다는 결과가 도출되었다. 이는 일상보행 시 적절한 수준의 녹지와 보행목적의 다양성을 높일 수 있는 건물의 비율이 높을수록 만족도가 높음을 시사한다. 운동보행만족도의 경우에는 가로녹지 비율이 약 7.5%, 가로 건물 비율이 약 30%의 적당한 비율일 때 보행만족도가 가장 높다는 결과가 도출되었다. 이는 무조건적으로 많은 양의 녹지 또는 건물의 비율을 높이는 것은 보행의 목적에 따라 오히려 부정적인 역할을 할 수 있으며, 적정 수준의 비율이 보행만족도에 긍정적인 영향을 미친다는 것을 시사한다.

V. 결론

본 연구에서는 서울시 거주민을 대상으로 실시한 2016년 설문조사 자료를 바탕으로 일상생활 보행만족도와 운동보행만족도 영향요인을 분석하였다. 독립변수로는 설문조사지 내의 개인 특성변수와 딥러닝 의미론적 분할 기법을 활용한 가로환경 특성변수, 토지이용 변수, 경계 공백 지대 변수로 유형을 구분하여 구축하였다. 나아가, 분석 방법론으로는 해석 가능한 기계학습 방법론을 활용하여 보행만족도에 영향을 미치는 변수의 기여도를 분석하고, 변수가 보행만족도와 가지는 비선형 관계와 상호작용 효과를 분석하였다.

분석 결과를 종합해보면, 도시의 물리적 환경요인과 보행만족도 사이에는 비선형적인 관계가 강하게 나타나는 것을 알 수 있다. 또한, 이러한 비선형적 관계는 보행목적별로 다르게 나타났으며, 물리적 환경변수 사이의 상호작용 효과도 보행만족도에 영향을 미치는 것으로 나타났다. 이러한 분석 결과는 보행목적에 따른 보행만족도 개선을 위해 차별적인 정책적 접근이 필요함을 시사한다. 구체적인 분석 결과와 정책적 시사점은 다음과 같다.

첫째, 해석 가능한 기계학습을 활용하여 전통적인 회귀분석 방법론의 모형보다 월등히 높은 설명력을 도출하였다. 일상보행만족도의 경우에는 전통 회귀모형의 설명력인 10%에서 기계학습 모형의 경우 설명력이 43%로 향상되었다. 이와 유사하게 운동보행만족도의 경우 전통 회귀모형의 설명력인 15%에서 기계학습 모형의 설명력이 52%로 매우 높게 나타났다. 따라서 기계학습 방법론

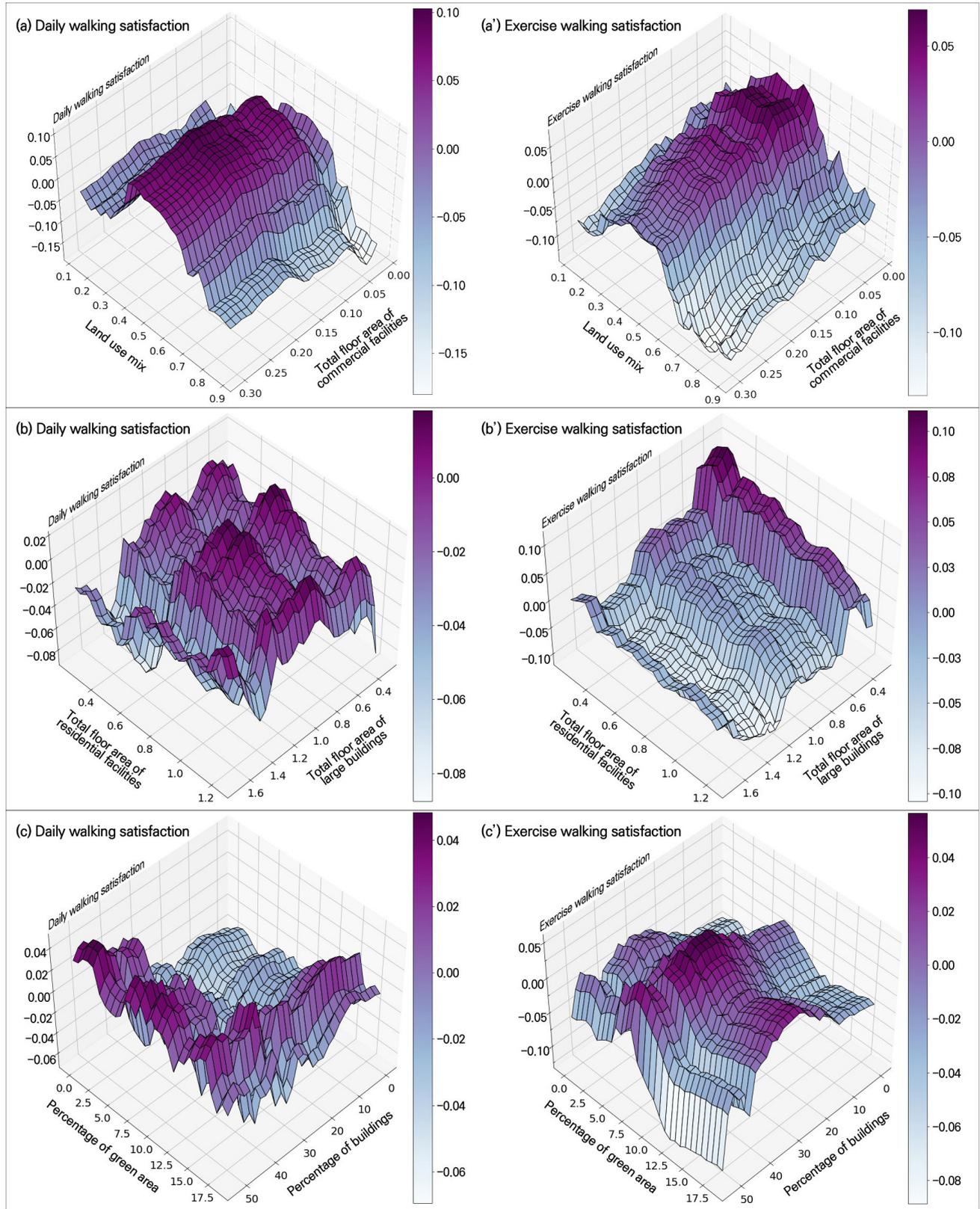


그림 6. 보행목적에 따른 보행만족도 상호작용효과 분석

Figure 6. Interaction effect of built environmental variables on walking satisfaction by purpose

의 활용은 기존의 전통적인 회귀분석 방법론과 비교하여 모형의 설명력 향상이 매우 높은 것을 알 수 있다. 이러한 모형 설명력의 향상은 향후 가로의 보행목적별 보행만족도 예측 모형의 개발에

크게 기여할 수 있다. 이러한 보행만족도 예측모형의 개발은 보행만족도 기반의 보행 경로 등을 제공하여 서비스 이용자에게 적절한 경로를 제시할 수 있다는 장점이 있다(이계승·이현희, 2019).

둘째, 일상보행만족도와 운동보행만족도 두 경우 모두에서 교차로의 개수가 적을수록 보행만족도는 높아지는 결과가 도출되었다. 이러한 결과는 블록의 규모가 작아질 경우 보행 빈도 및 다양성은 증가할 수 있겠으나, 잦은 교차로 횡단 등으로 보행 활동의 연속성을 저해시키기 때문에 보행만족도에는 부정적인 영향을 미치는 것으로 판단된다. 따라서 도시민의 보행만족도를 높이기 위해서는 보행 활동이 끊기지 않도록 보행의 연속성을 강화할 수 있는 접근이 보행만족도 향상에 매우 중요함을 의미한다.

셋째, 토지이용 혼합도의 경우에는 역 U자 모양의 포물선으로 전형적인 비선형 관계를 가지는 것으로 나타났다. 토지이용 혼합도 지수가 높아질수록 보행만족도가 증가하지만, 토지이용 혼합도가 약 0.4~0.6 이상일 때 보행만족도는 빠르게 감소하는 것으로 나타났다. 또한, 상호작용 효과를 추가적으로 보았을 때, 비슷한 수준의 토지이용혼합도 중에서도 운동보행만족도의 경우에는 상업시설 연면적이 작을수록 만족도는 높게 나타났다. 그리고 이러한 분석 결과는 특정 토지이용 밀도와 토지이용 혼합도가 보행 목적에 따라 보행만족도에 미치는 영향이 다를 수 있음을 시사한다.

넷째, 건물 연령의 경우 건물 연령 평균값이 약 22년일 때 일상보행만족도가 가장 높은 것으로 나타났다. 운동보행만족도의 경우에는 24~26년 정도 수준일 때 보행만족도가 높은 것으로 나타났다. 건물 연령 다양성의 경우에는 일상보행만족도와 운동보행만족도 두 가지 모두에서 낮을수록 보행만족도는 높다는 결과가 도출되었다. 이런 분석 결과를 종합하여 봤을 때 건물 연령의 평균이 20~24년 정도의 동시대에 지어진 건물일수록 보행만족도가 높다는 결론이 도출되었다. 따라서 도시민의 보행만족도 증진을 위해서는 오래된 노후 건축물에 대한 꾸준한 정비가 필요함을 시사한다.

다섯째, 운동보행만족도의 경우 가로 파노라마 이미지를 기준으로 가로의 녹지 비율은 약 7.5%, 건물 비율은 약 30% 수준일 때 보행만족도가 가장 높은 것으로 나타났다. 또한, 상호작용 분석 결과를 같이 고려할 때, 과도한 녹지 비율의 증가는 오히려 보행만족도에 부정적인 영향을 미치는 것으로 나타났다. 이러한 분석 결과는 보행목적에 따라 가로환경에서의 건물 및 녹지와의 조화가 이루어질 때 보행만족도는 증가할 수 있음을 시사한다.

마지막으로 경계 공백 지대 변수의 경우에는 대규모 건물 연면적을 제외한 하천, 고속도로, 대규모 공원 등 변수는 보행 만족도에 미치는 중요도가 크지 않은 것으로 나타났다. 이들 경계 공백 지대 변수는 더미 변수로 구축되었기 때문에 세부적인 영향력을 보기 위해서는 추가적인 연구가 필요할 것으로 판단된다. 반면 대규모 건물 연면적의 경우에는 주거시설 연면적과의 상호작용 효과를 보았을 때 일상보행만족도의 경우 상당히 복잡한 비선형적 관계가 나타난 반면, 운동보행만족도에서 대규모 주거시설 연면적이 클수록 보행만족도는 감소하는 경향이 뚜렷하게 나타났다. 따라서 대규모 아파트 단지과 같은 주거시설의 증가는 경계

공백으로 작용하여 보행만족도에 부정적인 영향을 미칠 수 있음을 시사한다.

한편, 본 연구는 다음과 같이 몇 가지의 한계점을 가진다. 우선, 본 연구에서 사용한 NSV 이미지의 경우 자동차를 기준으로 촬영된 이미지기 때문에 보행자가 인식하는 가로경관과는 차이가 있을 수 있다. 또한, 설문조사 자료에서 구축한 개인 특성변수의 경우 보행만족도에 대한 영향력이 크어도 불구하고 이에 대한 해석이 부족하다는 한계점이 존재한다. 따라서 향후 최근의 도시환경을 반영하여 개인적인 특성들이 보행만족도에 어떠한 영향을 미치는지 살펴볼 필요가 있을 것으로 판단된다.

그럼에도 불구하고 본 연구는 해석 가능한 기계학습 방법론을 활용하여 도시민의 보행만족도에 영향을 미치는 도시 물리적 환경요인을 분석했다는 점에서 의의가 있다. 또한, 선행연구들에서 한계점으로 언급되었던 변수 간의 비선형 관계를 확인하였고, 보행목적별로 다른 정책적 시사점을 제공했다는 점에서 의의가 클 것으로 판단된다. 도시민의 보행만족도 증진 목적의 도시환경에 대한 정확한 이해는 보행 친화적인 도시를 조성하고 이에 따른 사용자의 만족감을 높이는 데에 중요한 자료로서 활용될 수 있음을 시사한다.

인용문헌 References

- 권철민, 2019. 「파이썬 머신러닝 완벽 가이드」, 과주: 위키북스.
Kwon, C.M., 2019. *Python Machine Learning Perfect Guide*, Paju: Wikibooks.
- 기동환·김선재·이수기, 2021. “Google Street View와 딥러닝을 활용한 서울시 녹지형평성 분석: NDVI와 가로이미지 기반 녹지 산출방법과의 비교를 중심으로”, 「국토계획」, 56(4): 194-211.
Ki, D.H., Kim, S.J., and Lee, S., 2021. “Analysis of the Green Equity Using Google Street View and Deep Learning in Seoul, Korea: Focused on the Comparison between NDVI and Street Image-Based Green Calculation Method”, *Journal of Korea Planning Association*, 56(4): 194-211.
- 김창국, 2015. “물리적 가로환경 구성요소가 보행자의 지점방문 목적별 보행만족도에 미치는 영향 분석”, 한양대학교 대학원 석사학위 논문.
Kim, C.G., 2015. “Impact of Street’s Physical Environment on the Satisfaction of Pedestrian Regarding with the Different Walking Purposes in Seoul”, Master’s Dissertation, Hanyang University.
- 김창국·임하나·최창규, 2016. “보행 목적별 보행자 만족도에 영향을 미치는 근린 건조 환경 구성요소 특성 분석”, 「국토계획」, 51(4): 145-159.
Kim, C.K., Im, H.N., and Choi, C.G., 2016. “Built Environment, Walking Trip for Different Purposes, and Pedestrian Satisfaction”, *Journal of Korea Planning Association*, 51(4): 145-159.

5. 박근덕·기동환·이수기, 2021. “서울시 가로환경 요소의 시각적 특성이 보행만족도에 미치는 영향 분석: 구글 가로이미지와 딥러닝 의미론적 분할 기법을 활용하여”, 『도시설계』, 22(3): 55-72.
Park, K.D., Ki, D.H., and Lee, S., 2021. “Analysis of Visual Characteristics of Urban Street Elements on Walking Satisfaction in Seoul, Korea: Application of Google Street View and Deep Learning Technique of Semantic Segmentation”, *Journal of the Urban Design Institute of Korea Urban Design*, 22(3): 55-72.
6. 성현곤, 2014. “주거지 건조환경과 보행활동과의 비선형 연관성 진단: Jacobs의 삶에 기반한 물리적 환경의 보행목적별 차이를 중심으로”, 『국토계획』, 49(3): 159-174.
Sung, H.G., 2014. “Diagnosis on the Non-linear Association of Built Environment with Walking Activity in Residential Areas: Focused on the Difference of Walking Purposes for Physical Environment Based on Jacobs’ Life”, *Journal of Korea Planning Association*, 49(3): 159-174.
7. 성현곤·이수기·천상현, 2014. “보행활동에 영향을 미치는 커뮤니티 물리적 환경의 조절효과: 서울시 사례를 중심으로”, 『도시설계』, 15(2): 173-189.
Sung, H.G., Lee, S., and Cheon, S.H., 2014. “Moderation Effects of Community Physical Environment Factors on Walking Activity: With Case Study of Seoul, Korea”, *Journal of the Urban Design Institute of Korea Urban Design*, 15(2): 173-189.
8. 유진은, 2017. “기계학습을 통한 TIMSS 2011 중학생의 삶의 만족도에 영향을 미치는 변수탐색”, 『교원교육』, 33(1): 43-56.
Yoo, J.E., 2017. “TIMSS 2011 Predictors Relating to Korean 8th Graders’ Mathematics Achievement, Explored via Machine Learning”, *Korean Journal of Teacher Education*, 33(1): 43-56.
9. 윤준호·이수기, 2019. “기계학습과 LIME 방법론을 활용한 서울시 보행자 교통사고 심각도 요인 분석”, 한국지역학회 하반기 학술대회 논문집, pp.1-23.
Yoon, J. and Lee, S., 2019. “Analyzing the Determinant Factors of Pedestrian Accident Severity in Seoul Using Machine Learning and LIME”, *Proceeding of Fall Conference of Korea Regional Science Association*, pp.1-23.
10. 이경환·안건희, 2007. “커뮤니티의 물리적 환경이 지역 주민의 보행 시간에 미치는 영향: 서울시 40개 행정동을 대상으로”, 『국토계획』, 42(6): 105-118.
Lee, K.H., and Ahn, K.H., 2007. “The Correlation between Neighborhood Characteristics and Walking of Residents: A Case Study of 40 Areas in Seoul”, *Journal of Korea Planning Association*, 42(6): 105-118.
11. 이동엽·성현곤, 2020. “제인 제이콥스의 경계공백지대가 소매업 매출에 미치는 영향: 서울시를 중심으로”, 『서울도시연구』, 21(3): 167-183.
Lee, D.Y. and Sung, H.G., 2020. “How Do Jane Jacobs’ Border Vacuums Affect Retail Sale in the Great Seoul City?”, *Seoul Studies*, 21(3): 167-183.
12. 이수기·고준호·이기훈, 2016. “근린환경특성이 보행만족도에 미치는 영향 분석: 서울서베이 2013년 자료를 중심으로”, 『국토계획』, 51(1): 169-187.
Lee, S., Ko, J.H., and Lee, G.H., 2016. “An Analysis of Neighborhood Environment Affecting Walking Satisfaction: Focused on the ‘Seoul Survey’ 2013”, *Journal of Korea Planning Association*, 51(1): 169-187.
13. 이수기·이운성·이창관, 2014. “보행자 연령대별 보행만족도에 영향을 미치는 가로환경의 특성분석”, 『국토계획』, 49(8): 91-105.
Lee, S., Lee, Y.S., and Lee, C.K., 2014. “An Analysis of Street Environment Affecting Pedestrian Walking Satisfaction for Different Age Groups”, *Journal of Korea Planning Association*, 49(8): 91-105.
14. 이신해·정삼미, 2019. 「걷는 도시, 서울’ 정책효과와 향후 정책 방향」, 서울연구원 정책과제연구보고서, 서울
Lee, S.H. and Jeong, S.M., 2019. *Effects of ‘Walking City, Seoul’ Policy and Directions for Future Researches*, The Seoul Institute Policy Research Project Research Report, Seoul.
15. 이제승·이현희, 2019. “기계학습 알고리즘을 이용한 보행만족도 예측모형 개발”, 『국토계획』, 54(3): 106-118.
Lee, J.S. and Lee, H.H., 2019. “Developing a Pedestrian Satisfaction Prediction Model Based on Machine Learning Algorithms”, *Journal of Korea Planning Association*, 54(3): 106-118.
16. 이지윤, 2021. “보행만족도에 영향을 미치는 가로경관의 물리적 환경 특성 분석”, 한양대학교 대학원 석사학위논문.
Lee, J.Y., 2021. “A Study of the Physical Environment Characteristics of Streetscapes Affecting Pedestrian Satisfaction”, Master’s Dissertation, Hanyang University.
17. 조월·김신재·이수기, 2021. “근린환경이 도시활력에 미치는 영향 분석: 도시 빅데이터와 해석 가능한 기계학습을 활용하여”, 대한국토·도시계획학회 춘계산학협동대회 논문집, pp.1-23. 온라인학회.
Cao, Y., Kim, S.J., and Lee, S., 2021. “An Analysis of the Impact of Neighborhood Environment on Urban Vitality Using Big Data and Interpretable Machine Learning Model”, *Proceeding of Spring Congress of Korea Planning Association*, pp.1-23.
18. 조혜민·이수기, 2016. “보행목적별 보행활동시간에 영향을 미치는 근린환경 특성분석: 주관적 인지환경과 객관적 측정환경의 차이를 중심으로”, 『국토계획』, 51(4): 105-122.
Cho, H.M. and Lee, S., 2016. “Analysis of Neighborhood Environmental Characteristics Affecting Walking Activity Time: Focused on the Difference between Subjectively Measured and Objectively Measured Neighborhood Environment”, *Journal of Korea Planning Association*, 51(4): 105-122.
19. 조혜민·이수기, 2017. “근린환경특성이 사회적 자본의 수준에 미치는 영향 연구: 보행활동의 매개효과를 중심으로”, 『국토계획』, 52(4): 111-133.
Cho, H.M. and Lee, S., 2017. “A Study on the Effects of Neighborhood Environmental Characteristics on the Level of the Social Capital: Focused on the Mediating Effect of Walking Activity”, *Journal of Korea Planning Association*, 52(4): 111-133.
20. 최창규·성현곤·이수기·김태현·고준호·원보환, 2013. “지속가능 도시를 위한 보행활동 증진방안”, 『도시정보』, 373: 3-22.
Choi, C.G., Sung, H.G., Lee, S., Kim, T.H., Ko, J.H., and Won, B.W., 2013. “Promoting Walking Activities for Sustainable Cities”, *Urban Information Service*, 373: 3-22.
21. Chen, L., Yao, X., Liu, Y., Zhu, Y., Chen, W., Zhao, X., and Chi, T., 2020. “Measuring Impacts of Urban Environmental

- Elements on Housing Prices Based on Multisource Data—A Case Study of Shanghai, China”, *ISPRS International Journal of Geo-Information*, 9(2): 106.
22. Chen, L., Zhu, Y., Papandreou, G., Schroff, F., and Adam, H., 2018. “Encoder-decoder with Strous Separable Convolution for Semantic Image Segmentation”, *In Proceedings of the European Conference on Computer Vision (ECCV)*, 801-818.
 23. Chung, Y., 2013. “Factor Complexity of Crash Occurrence: An Empirical Demonstration Using Boosted Regression Trees”, *Accident Analysis & Prevention*, 61: 107-118.
 24. Ewing, R., Meakins, G., Hamidi, S., and Nelson, A.C., 2014. “Relationship between Urban Sprawl and Physical Activity, Obesity, and Morbidity—Update and Refinement”, *Health & Place*, 26: 118-126.
 25. Frank, L.D., Sallis, J.F., Saelens, B.E., Leary, L., Cain, K., Conway, T.L., and Hess, P.M., 2010. “The Development of a Walkability Index: Application to the Neighborhood Quality of Life Study”, *British Journal of Sports Medicine*, 44(13): 924-933.
 26. Friedman, J.H., 2001. “Greedy Function Approximation: A Gradient Boosting Machine”, *Annals of Statistics*, 29(5): 1189-1232.
 27. Jacobs, J., 1961. *The Death and Life of Great American Cities*, New York: Random House.
 28. Leslie, E., Cerin, E., and Kremer, P., 2010. “Perceived Neighborhood Environment and Park Use as Mediators of the Effect of Area Socio-economic Status on Walking Behaviors”, *Journal of Physical Activity and Health*, 7(6): 802-810.
 29. Lundberg, S.M., Erion, G.G., and Lee, S.I., 2018. “Consistent Individualized Feature Attribution for Tree Ensembles”, arXiv: 1802.03888.
 30. McCormack, G.R. and Shiell, A., 2011. “In Search of Causality: A Systematic Review of the Relationship between the Built Environment and Physical Activity among Adults”, *International Journal of Behavioral Nutrition and Physical Activity*, 8(1): 1-11.
 31. Parsa, A.B., Movahedi, A., Taghipour, H., Derrible, S., and Mohammadian, A.K., 2020. “Toward Safer Highways, Application of XGBoost and SHAP for Real-time Accident Detection and Feature Analysis”, *Accident Analysis & Prevention*, 136, 105405: 1-8.
 32. Riggs, W., 2014. “Steps Toward Validity in Active Living Research: Research Design that Limits Accusations of Physical Determinism”, *Health & Place*, 26: 7-13.
 33. Sallis, J.F., Saelens, B.E., Frank, L.D., Conway, T.L., Slymen, D.J., Cain, K.L., Chapman, J.E., and Kerr, J., 2009. “Neighborhood Built Environment and Income: Examining Multiple Health Outcomes”, *Social Science & Medicine*, 68(7): 1285-1293.
 34. Sung, H., Lee, S., and Cheon, S., 2015. “Operationalizing Jane Jacobs’s Urban Design Theory: Empirical Verification from the Great City of Seoul, Korea”, *Journal of Planning Education and Research*, 35(2): 117-130.
 35. Yang, J., Cao, J., and Zhou, Y., 2021. “Elaborating Non-linear Associations and Synergies of Subway Access and Land Uses with Urban Vitality in Shenzhen”, *Transportation Research Part A: Policy and Practice*, 144: 74-88.

Date Received 2021-11-18

Date Reviewed 2021-12-27

Date Accepted 2021-12-27

Date Revised 2022-02-07

Final Received 2022-02-07