

Research Paper

## 기계학습 기반의 로드킬 발생 예측과 영향 요인 탐색에 대한 연구

허소진 · 김지영

경북대학교

### A Study on Machine Learning-Based Estimation of Roadkill Incidents and Exploration of Influencing Factors

Sojin Heo · Jeeyoung Kim

Kyungpook National University

**요약:** 본 연구에서는 충청남도를 중심으로 로드킬 발생을 예측하고 영향을 미치는 요인을 탐구하여 로드킬 예방 대책 수립에 이바지하고자 하였다. 날씨, 도로 및 환경 정보를 종합적으로 고려하여 기계학습을 기반으로 로드킬 발생을 예측하고 각 변수의 중요성을 분석하여 주요 영향 요인을 도출하였다. 가장 우수한 성능을 보인 Gradient Boosting Machine(GBM)은 정확도 92.0%, 재현율 84.6%, F1-score 89.2%, AUC 0.907을 기록했다. 로드킬에 영향을 미치는 주요 요인은 평균 지역 기압(hPa), 평균 지면 온도(°C), 월, 평균 이슬점 온도(°C), 중앙 분리대 존재 여부, 평균 풍속(m/s)이었다. 이러한 결과는 로드킬 예방 및 교통안전에 이바지할 것으로 기대되며, 생태계와 도로 개발 간의 균형 유지에 중요한 역할을 할 것으로 예상된다.

**주요어:** 로드킬, 기계학습, 영향요인분석

**Abstract:** This study aims to estimate roadkill occurrences and investigate influential factors in Chungcheongnam-do, contributing to the establishment of roadkill prevention measures. By comprehensively considering weather, road, and environmental information, machine learning was utilized to estimate roadkill incidents and analyze the importance of each variable, deriving primary influencing factors. The Gradient Boosting Machine (GBM) exhibited the best performance, achieving an accuracy of 92.0%, a recall of 84.6%, an F1-score of 89.2%, and an AUC of 0.907. The key factors affecting roadkill included average local atmospheric pressure (hPa), average ground temperature (°C), month, average dew point temperature (°C), presence of median barriers, and average wind speed (m/s). These findings are anticipated to contribute to roadkill prevention strategies and enhance traffic safety, playing a crucial role in maintaining a balance between ecosystems and road development.

**Keywords:** Roadkill, Machine Learning, Influential Factors

## I. 서론

로드킬(Roadkill)은 야생동물이 생태계 파편화로 인하여 서식지를 이동하면서 도로를 횡단하는 과정에서 발생하는 ‘동물 찾길 사고’로 정의되며 국제적으로는 Wildlife Vehicle Collision(이하 WVC)으로도 알려져 있다. 국토교통부에 따르면 2022년 12월 31일 기준 전국 도로 연장은 114,314km로서 전년(113,405km) 대비 909km 증가하였다(Road Status Report 2022). 이와 더불어 철도, 수로 등의 선형개발까지 포함한다면 생태 축 단절은 점차 심화하고 있다고 볼 수 있다. 생태 축 단절로 인해 일정 공간 내에서 먹이가 부족하여 야생동물의 이동이 증가하고, 그로 인해 로드킬이 점점 빈번하게 일어나고 있는 추세이다. 또한, 상대적으로 넓은 서식지가 필요한 고라니와 멧돼지 같은 야생동물의 개체 수가 감소하는 등 생태계가 불안정하여 결과적으로 인간의 편익을 저하하는 악순환으로 연결될 수 있다(Lee et al. 2014).

생태계 파편화와 로드킬을 완화하고자 국토교통부와 환경부는 2020년 6월에 처음으로 로드킬 다발 상위 50구간을 선정하여 로드킬 저감 대책을 추진한 바 있다(Song et al. 2022). 하지만 국립생태원에서 공개한 6년간(2017년~2022년) 로드킬 현황을 보면 로드킬이 점점 증가하는 경향을 보인다(Figure 1). 특히, 2020년 이후 로드킬 건수가 많이 증가하였는데, 이는 로드킬에 관한 관심과 데이터 수집의 향상을 가리키며, 동시에 연구의 중요성도 상승하고 있음을 시사한다. 국내·외 로드킬 관련 연구는 현황분석 연구, 특정 종을 대상으로 한 로드킬 방지 연구, 핫스팟 및 시공간 패턴 분석 등 다양하게 이루어진 반면에 예측 모델링을 목표로 하는 연구는 소수에 불과하다.

Song et al. (2019)는 로드킬 다발 구간 선정을 위한 공간 군집지역 탐색 방법을 검토하기 위해 시범적으로 4개 시군(영주시, 문경시, 안동시, 청송군) 내 로드킬 접수 건을 바탕으로 로드킬 핫스팟 분석을 수행하여 국내 상황에 적합한 로드킬 다발 구간 분석 방법을 제시하였다.

Son et al. (2016)는 특정 도로구간을 대상으로 로지스틱 회귀분석을 실시하여 로드킬이 발생한 지역의 공

간적 특성을 분석하고, 분석 결과의 적합도 및 예측도를 평가하여 로드킬 발생 근본 원인을 파악하고자 하였다. 추정된 모형의 정확도(Accuracy)는 72.2%이며, 모든 변수의 회귀계수는 유의하였다( $p < 0.01$ ). 펜스·옹벽 등이 설치된 구간일수록 로드킬 발생 가능성이 낮으며 가장 큰 영향력을 보였다. 주거지역·건물 등의 시설과 하천으로부터의 거리가 가까울수록, 식생 영급이 커질수록 로드킬 발생 확률이 높아질 수 있다고 하였다.

Choi & Park (2006)은 26개월간 현장에서 조사된 로드킬 자료를 이용한 GIS 분석을 통하여 로드킬 발생 밀도와 주변 토지이용 유형의 특성을 밝히고자 하였다. 초지와 인접한 도로에서 매우 높은 로드킬 발생 밀도(179.2건/km<sup>2</sup>)를 보였으며, 다음으로 주거지(75.3건/km<sup>2</sup>), 밭(63.2건/km<sup>2</sup>), 하천(42.3건/km<sup>2</sup>), 산림(38.0건/km<sup>2</sup>), 논(35.5건/km<sup>2</sup>)으로 나타났다.

Pagany et al. (2020)는 Gaussian Naive Bayes, Stochastic Gradient Descent, 그리고 Random Forest 세 가지 머신러닝 모델을 사용하여 WVC를 예측하였다. Random Forest가 86.7%의 정확도로 가장 높은 성능을 보였다. 불순도(Impurity) 기반 변수 중요도의 경우 도로 등급(Road Rank)이, 순열(Permutation) 기반 변수 중요도의 경우 태양 고도(Solar altitude)가 가장 높게 나왔다. 해당 연구에서는 시간, 기온과 같은 동적 변수와 도로 인프라와 관련한 정적 변수 모두가 WVC 예측에 중요하다는 것을 강조하였다.

Ha & Shilling (2018)은 북캘리포니아 지역 고속도로

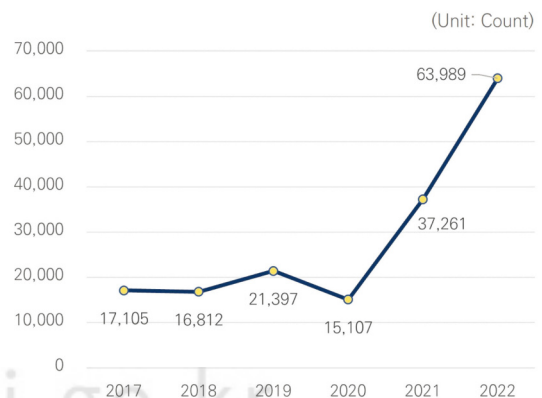


Figure 1. Roadkill by year from 2017 to 2022

에서 수집된 WVC를 최대 엔트로피 접근법(Maximum entropy approach)을 바탕으로 하는 Maxent 모형에 적용하였다. WVC를 네 가지 분류군(유제류, 조류, 중형 포유류, 소형 포유류)으로 나누어 환경 및 인구 특성 간의 관계를 분석하였다. WVC와 환경 및 인구 특성 관계는 비선형적이고 불연속적인 함수인 경우가 많으므로 Maxent와 같은 머신러닝 알고리즘이 이를 식별하는데 매우 적합하다고 언급하였다. 결과로는 유제류 WVC는 산림 밀도가 높고, 도로 밀도가 낮은 지역에서 발생하는 반면, 조류 및 중형 포유류 WVC는 도로 밀도가 높은 지역에 집중되어 있음을 보였다.

본 연구에서는 로드킬 예방과 교통안전 강화를 목적으로 다양한 변수와 기계학습을 활용하여 로드킬 발생을 예측하고, 로드킬 영향 요인을 탐색하고자 하였다. 본 연구의 기여점은 다음 3가지이다. 첫째, 국내 지역적 공간 특성을 반영한 로드킬 데이터셋을 구축하고, 지역의 특성을 반영하여 로드킬을 예측하고 요인을 도출하였다는 점이다. 둘째, 단순히 변수의 중요도가 아닌 변수 간 상호작용을 고려하여 개별 변수가 로드킬 발생 예측에 미치는 영향을 설명하였다. 셋째, 다른 지역도 적용 가능한 모델이라는 점에서 범용성을 가진다.

## II. 연구방법

Figure 2는 지도학습(Supervised Learning)을 이용한 로드킬 발생 예측을 위한 과정을 나타낸다. 본 연구에서는 총 7가지 데이터베이스(기상, 도로망, 로드킬, 생

태자연도, 생태통로, 토지피복도, 표준노드링크)를 기반으로 로드킬 발생 예측을 위한 데이터셋을 구축하였다. 분석 단위 형태는 점(Point) 사상의 공간 데이터를 사용하며, 로드킬이 발생한 지점은 국립생태원 로드킬 데이터베이스의 위도, 경도 속성을 이용하였다. 로드킬이 발생하지 않은 지점은 도로망 위 500m 간격으로 점을 생성하였고, 이후, 라벨링과 결측치 및 이상치 처리를 위한 전처리 단계를 거치고, 지도학습 기반의 분류 모델을 통하여 로드킬 발생 예측을 수행한다. 4개의 머신러닝 분류 모델을 실험하고, 성능이 가장 우수한 모델을 대상으로 변수 중요도를 검토하여 로드킬에 영향을 미치는 요인을 탐색하고자 한다. 본 장에서는 연구대상지 선정과 데이터셋 구축을 포함하여 실험에 사용한 분류 기법에 이르기까지, 본 논문에서 제안하는 방법 전반으로 구성된 과정에 대하여 설명한다.

### 1. 연구대상

2019년에서 2021년까지 전국 17개 시도의 로드킬 발생을 분석한 결과, 충청남도에서 발생한 로드킬의 3개년 총합계가 7,331건으로 가장 많이 발생하였다. 2004년부터 2019년까지의 고속도로상의 로드킬 분석 결과, 타 시도에 비해 충청남도의 로드킬 건수가 많이 증가하는 추세이며, 이로 인한 대책 마련이 시급하다고 분석하였다(Kim et al. 2021). 야생 동물이, 대체로 받을 포함하는 형태의 토지이용구간에서 도로 접근이 많아 토지이용을 고려하여 산림 위주의 토지이용 구간보다는 밭이 인접한 도로를 중심으로 야생동물의 도로 진

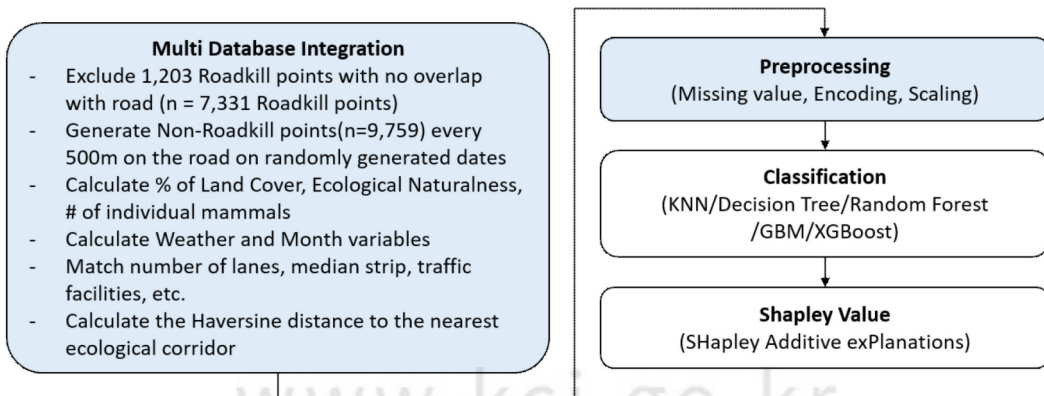


Figure 2. Process for estimation analysis of roadkill incidents

입 대책이 필요하다(Min & Han 2010). 충청남도에는 농경지, 초지, 습지 등 로드킬 빈도가 높은 동물의 서식지가 많고, 서식지 인근에 도로가 발달 되어있어 로드킬 증가에 따른 대책 역시 필요하다. 따라서, 본 연구에

서는 전국 시도 중 충청남도를 우선적인 연구 대상 지역으로 선정하였다.

다음으로 로드킬 발생 지점의 도로 등급(Road Rank)을 파악하기 위하여 로드킬 데이터와 국가 교통 DB(ktdb.go.kr)의 도로망 데이터를 공간 조인(Spatial Join)하였다. 점 형식의 로드킬 데이터와 선(Line) 형식의 도로망 데이터를 공간 조인하기 위해 도로망 데이터에 100m 반경의 버퍼(Buffer)를 생성하고, 교차 영역이 가장 작은 버퍼의 도로망 정보와 결합하였다. 이러한 로드킬 발생 지점의 도로 등급 분석 결과, 고속도로와 일반국도가 전체 로드킬 발생 지점의 83.6%로 대부분의 로드킬이 고속도로와 일반국도에서 발생함을 확인하였다.

따라서, 본 연구에서는 로드킬 발생이 상대적으로 많은 충청남도의 고속도로와 일반국도를 우선 연구 대상으로 선정하였다. Figure 3은 충청남도의 3개년(2019-2021) 로드킬 발생지점을 나타낸다.

## 2. 데이터셋 구축

본 연구에서는 지도학습을 위해 로드킬이 발생한 지점과 발생하지 않은 지점으로 나누어 데이터셋을 구축하였고, 변수 구성은 Table 1과 같다.

로드킬이 발생한 지점은 국립생태원 로드킬 정보시스템에서 수집된 로드킬 발생 지점 정보(위도, 경도, 접수 일자)를 사용하여 도로 외 지점에서 발생한 데이터의 경우 이상치로 간주하고 도로망 50m 반경을 벗어나는 데이터는 분석에서 제외하였다. 로드킬이 발생하지 않은 지점의 경우 고속도로와 일반국도 위에 500m 간격의 점(Point)을 생성하였다. 로드킬이 발생하지 않은 지점의 경우 기상정보 매칭을 위해 해당 지점의 기준 날짜를 로드킬 데이터의 연도별 비율과 동일하게 무작위 추출하여 생성하였다. 이러한 과정을 거친 최종 데이터셋은 15,887개의 샘플로 구성되어 있으며 로드킬이 발생한 지점은 6,128개, 로드킬이 발생하지 않은 지점은 9,759개이다.

각 지점별 주변 환경에 대한 설명변수 생성을 위해 지점을 기준으로 500m 반경을 생성하였다. 또한, 주변 공간 특성을 파악하기 위하여 토지피복도(중분류)를 활용하였다. 토지피복도는 지구 표면 지형지물의 형태

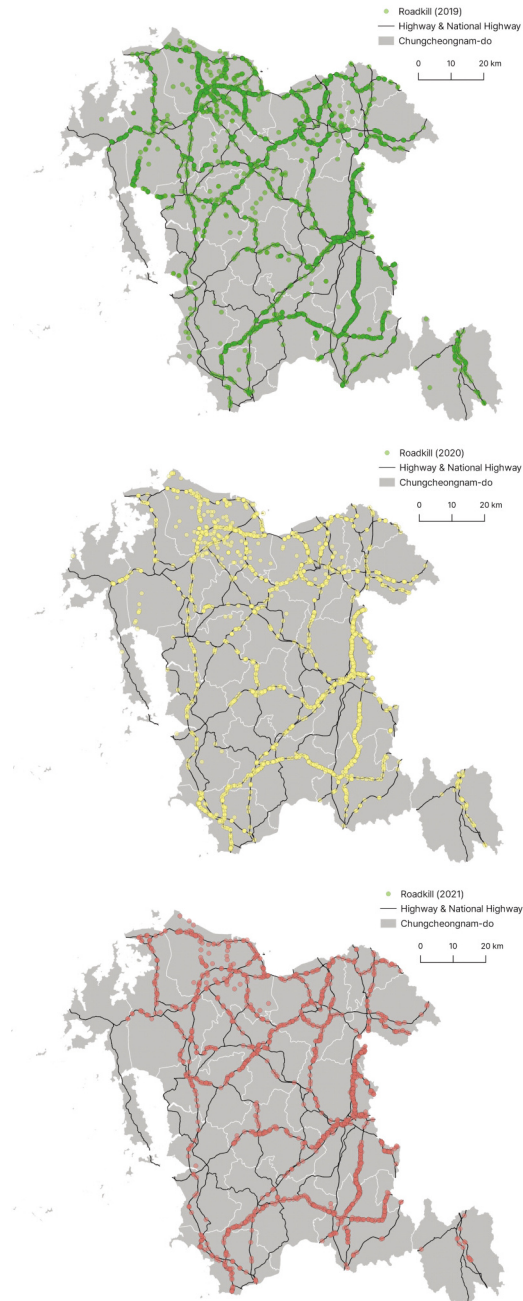


Figure 3. Roadkill locations in Chungcheongnam-do (2019-2021)

Table 1. Configure the roadkill occurrence estimation dataset variables

Category		Variable
Dependent variable		Roadkill occurrence
Description Variables	Weather	Average temperature (°C)
		Average dew point temperature (°C)
		Average Ground-surface temperature (°C)
		Daily precipitation (mm)
		Average wind velocity (m/s)
		Average relative humidity (%)
		Average Local pressure (hPa)
		Average cloud cover (1/10)*
	Time	Month
	Road Infrastructure	Number of lanes
		Median strip
		Traffic Facilities
		Speed limit
	Environment	Number of mammal individuals within a 500m radius
		Percentage of ecological naturalness (1st grade) within a 500m radius
		Haversine distance (meters) to the nearest ecological wildlife crossing structure
Percentage of land cover (subcategory) within a 500m radius, categorized by type		

를 일정한 과학적 기준에 따라 분류한 공간정보 DB로 중분류 데이터셋의 경우 22개 범주로 구성 되어있다. 500m 반경과 토지피복도를 중첩하여 토지피복도의 각 범주가 반경 내 차지하는 면적의 비율을 산출하였다.

다음으로 자연환경 특성을 나타내기 위하여 생태자연도를 토지피복도와 같은 방식으로 생태자연도에서 1등급 지역만 추출하여 500m 반경 내 해당 면적의 비율을 산출하였다. 야생동물의 서식지가 도로와 인접할 수록 로드킬 핫스팟과 유사한 특성을 가진다는 기존 연구 사례가 있다(Song et al. 2011). 따라서, 해당 지점이 야생동물이 많이 서식하는 지역인지 파악하기 위하여 포유류 개체 수를 설명변수로 활용하고자 하였다. 각 지점을 기준으로 주변에 해당 개체들의 분포를 확인하고자 '전국 자연환경 조사'의 점 형식의 포유류 데이터를 사용하였다. 분석에 활용한 포유류 종으로는 로드킬 상위 6개 종(고라니, 너구리, 멧돼지, 삿, 오소리, 족제비) 중에서 자연환경 조사 데이터셋에 존재하지 않는 '삿'을 제외한 5개 종을 대상으로 하였다. 이와 같은 방식으로 지점별 500m 반경 내 포유류 개체 수를 집계하였다.

그 다음, 앞서 언급한 로드킬 접수날짜와 무작위 추출한 날짜를 기준으로 해당 지점의 기상정보를 결합하였다. 기상 데이터의 경우 기상자료개방 포털에서 제공하는 종관기상관측(ASOS)자료를 사용하였다. 충청남도에는 총 6개의 기상관측소가 있으며, 본 연구에서는 지점별로 가장 가까운 관측소의 기상정보를 사용하였다. 또한, 계절의 변화에 따른 야생동물의 이동패턴을 고려하고자 기준 날짜로부터 월(Month)을 추출하여 설명변수로 활용하였다.

충돌 방지턱과 같은 도로 시설물이 WVC의 위험을 절반으로 줄인다(Pagany & Dorner 2019). 중앙분리대가 없는 도로보다 중앙분리대가 있는 도로에서 로드킬 발생률이 2배 이상 높았다(Kweon et al. 2008). 따라서, 본 연구에서도 도로 인프라와 관련된 변수를 도출하고자 하였다. 먼저, 선(Line) 데이터 형식의 도로망에 50m 반경을 설정한 후, 각 지점과 교차하여 최대 중첩 영역을 가진 도로 정보를 결합하였다. 이 데이터는 제한속도, 중앙분리대 종류, 교통시설물 종류, 차선 수로 구성되어 있다. 제한속도는 국가교통정보센터에서 제공하는 표준노드링크(2022. 11) 데이터베이스를 활용

하였으며, 중앙분리대 종류, 교통시설물 종류, 차선 수는 국가 교통 DB에서 제공하는 도로망 데이터(2021. 12)를 활용하였다.

생태통로(Wildlife Crossing Structure)는 단절된 생태계의 연결 및 야생동물의 이동을 위한 인공구조물을 말하며 생태통로까지의 거리가 가까울수록 로드킬 발생 가능성이 적어질 것이라는 가설하에 이를 변수로 활용하였다. 이를 위해 각 지점에서 가장 근접한 생태통로까지의 거리를 산출하였는데, 이때의 거리 산출 공식으로는 유클리디안(Euclidean) 거리가 아닌 지구의 곡률을 고려한 하버사인(Haversine) 거리 공식을 사용하였다. 하버사인 거리( $d$ )의 수식은 (1)과 같다.

$$d = 2r \times \arcsin(\sqrt{\text{hav}(\theta_2 - \theta_1) + \cos(\theta_1)\cos(\theta_2)\text{hav}(\lambda_2 - \lambda_1)}) \quad (1)$$

### 3. 분류모델과 SHAP

지도학습 기반의 분류를 수행하기 위해 가장 널리 통용되어 쓰이는 모델들을 사용하였는데, K-최근접 이웃 알고리즘(K-Nearest Neighbor, KNN), 의사결정트리(Decision Tree), 랜덤포레스트(Random Forest), GBM(Gradient Boosting Machine), XGBoost(Extreme Gradient Boosting) 등 총 5개 모델이다.

KNN은 입력 데이터와 그 주변 이웃 데이터들의 거리를 기준으로 예측을 수행하는 지도학습 알고리즘이다. 새로운 입력 데이터를 분류할 때 입력값과 가장 가까운 K개의 이웃들의 레이블을 살펴보고 다수결, 가중치 등의 방법을 사용하여 분류를 수행한다. 거리 측정 방법은 유클리드 거리, 맨하탄(Manhattan) 거리 등을 활용할 수 있고 단순하고 직관적인 알고리즘으로 알려져 있다.

의사결정트리는 데이터에 내재하여 있는 패턴을 특성들의 조합으로 나타내는 분류 모델을 나무 형태로 만드는 알고리즘이다. 데이터를 두 개 이상의 부분집합으로 나누는데 이때 데이터가 균일해지도록 나눈다. 랜덤포레스트는 앙상블 학습(Ensemble Learning)의 하나로 여러 개의 의사결정트리를 결합하여 최종 예측을 수행하는 모델이다. 각각의 의사결정트리는 특정 특성을 기준으로 데이터를 분할하여 예측한다. 의사결정트리를

학습시킬 때, 기존 데이터셋과 같은 크기의 샘플을 복원추출을 통해 샘플링하는 부트스트랩(Bootstrap) 기법을 이용하여, 각 의사결정트리가 서로 다른 데이터셋을 사용하여 학습하게 한다.

GBM은 부스팅 알고리즘의 대표적인 모델이다. 부스팅 알고리즘은 여러 개의 약한 학습기를 순차적으로 학습하여 예측하며, 잘못 예측한 데이터에 가중치를 부여하여 통해 오류를 개선해 나가면서 학습하는 방식이다. XGBoost는 GBM과는 달리 경사하강법을 사용하여 예측 오차를 최소화하는 새로운 트리를 구성한다. 각 트리는 이전 트리의 예측 오차를 보완하도록 학습되어 모델을 개선하고, 예측 시에 모든 트리의 결과를 결합하여 최종 예측을 수행한다.

본 연구에서는 위에서 언급한 다섯 가지 모델을 사용하여 로드킬 발생 여부를 예측하는 이진 분류를 수행한다. 다양한 모델 선택을 한 이유는 데이터셋의 특성을 다각도로 이해하고, 각 모델의 장단점을 고려하여 최적의 예측 결과를 도출하기 위한 것이다. 이러한 방법론은 로드킬 발생 여부를 예측하는데 있어서 데이터의 복잡성을 탐색하고, 효과적인 예측 모델을 구축하는 데 이바지할 것으로 기대된다.

SHAP(SHapley Additive exPlanations) 값은 게임 이론에서 파생된 개념으로 공정한 대가 분배를 결정하는데 사용되는 방법론이다(Lundberg et al. 2018). 각 변수의 조합에 따른 기댓값의 차이를 고려하여 변수의 상대적인 영향력을 계산하는 방법으로, 변수의 중요도를 정량적으로 파악하는 데에 유용하다. 각 변수의 SHAP 값은 해당 변수가 없을 때와 있을 때의 모델 예측값의 차이에 기초한다. 변수가 추가될 때마다 이 차이를 계산하고, 모델 예측에 미치는 영향을 평가하기 위해 가능한 모든 변수 조합에 대해 반복한다. 이때, 각 변수 조합의 크기에 따라 가중평균을 취하여 변수가 모델 예측에 미치는 영향을 추정한다. SHAP 값은 각 변수가 예측에 미치는 영향을 공정하게 평가하는 방법으로 변수 간 상호작용을 고려하여 모델의 예측에 대한 기여도를 정량화한다. 이러한 접근 방식은 변수의 상대적인 중요도와 모델의 예측을 해석하는 데 유용하다.

Table 2. Comparison of the performance of classification models for classifying roadkill occurrence

Model		KNN	Decision Tree	Random Forest	GBM	XGBoost
Accuracy (%)	Train (Mean k-fold cv)	79.5	81.2	87.2	91.3	89.6
	Test	79.6	81.2	87.5	92.0	89.8
Recall (%)		79.0	65.8	75.2	84.6	80.9
F1-score (%)		75.0	73.0	82.4	89.2	86.0
AUC		0.795	0.784	0.853	0.907	0.882

Table 3. Hyperparameters by classification model

Model	Hyperparameters
KNN	metric='manhattan', n_neighbors=5, weights='uniform'
Decision Tree	nim_samples_split=2, min_samples_leaf=1, max_depth=9
Random Forest	n_estimators=150, max_depth=25, min_samples_leaf=1, min_samples_split=2
GBM	learning_rate=0.1, max_depth=9, min_samples_leaf=3, min_samples_split=8, n_estimators=150
XGBoost	objective='binary:logistic', colsample_bytree=0.9, learning_rate=0.1, max_depth=5, subsample=0.9, n_estimators=250

### III. 연구결과

학습과 실험 데이터셋의 비율은 8:2로 설정하였으며 과적합(Overfitting) 방지를 위해 Stratified K-Fold Cross Validation(k=5)을 학습 과정에 활용하였다. 모델별 성능(Table 2)과 하이퍼파라미터(Table 3)는 아래와 같다. 학습 데이터셋과 테스트 데이터셋의 정확도를 비교하였을 때 큰 차이가 없는 것으로 보아 일반화가 성공적으로 수행되었음을 알 수 있다. 정확도 92.0%, 재현율 84.6%, F1-score 89.2%, AUC 0.907로 4개 성능 평가 지표에서 GBM이 가장 높은 성능을 보였다. 정확도와 재현율, F1-score는 각각 가장 낮은 지표 대비 12.4%p, 18.8%p, 16.2%p 더 높은 성능을 보였으며, AUC는 가장 낮은 지표 대비 0.123 더 높았다.

가장 높은 성능을 보인 GBM 모델을 기준으로 SHAP 기법을 이용하여 변수 중요도를 분석한 결과, 평균 현지기압(hPa), 평균 지면온도(°C), 월, 평균 이슬점온도(°C), 중앙분리대 순으로 예측에 큰 영향을 미쳤다(Figure 4). 하나의 점은 하나의 데이터를 나타내며, 색깔은 변수값의 크고 작은 정도를 나타낸다. X축은 SHAP 값을 나타내고, 이 값이 양수이면 예측값에 긍정적인 영향을 미친다고 해석할 수 있다.

평균 현지기압(hPa), 평균 지면온도(°C), 평균 이슬

점온도(°C), 평균 풍속(m/s), 평균 상대습도(%), 중앙분리대 유무, 차선 수의 경우 예측값에 양의 상관성으로 영향을 미쳤다. 또한, 상반기(1~6월)일수록, 중앙분리대가 설치되어 있을수록, 생태통로까지의 거리가 가까울수록 로드킬이 발생했다고 예측했다. 토지피복도 범주 내에서 살펴보면 활엽수림, 공업지역, 상업지역, 과수원, 교통, 논, 침엽수림, 기타채배, 인공초지, 혼효림

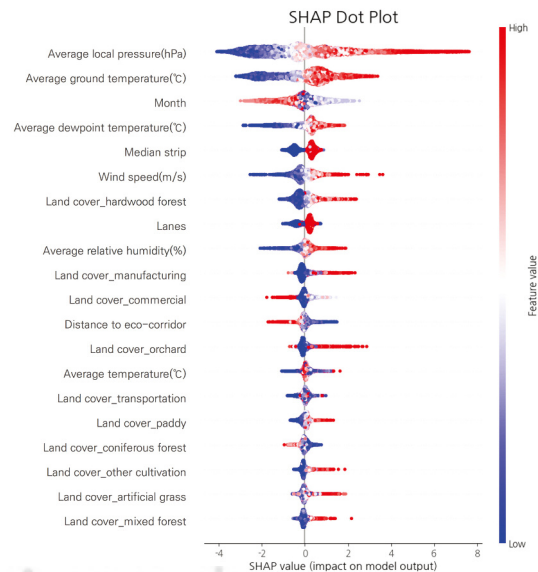


Figure 4. Predicted impact by SHAP variable based on Gradient Boosting Machine

순으로 영향을 크게 미쳤다. 500m 반경 내 활엽수림, 논, 기타채배, 혼효림 면적의 비율은 예측값에 양의 상관성, 침엽수림 면적의 비율은 예측값에 음의 상관성을 미쳤다. 500m 반경 내 공업, 상업, 과수원, 교통, 인공초지 면적의 비율과 평균기온(°C)은 예측값과 비선형 관계에 있음을 알 수 있다.

#### IV. 결론

로드킬은 발생 시점과 위치를 예측하기 어려운 사고이다. 비교적 짧은 기간 동안 급격한 도시화를 겪은 한국에서 로드킬은 매년 증가하는 추세이다. 한국에서 로드킬 관련 연구는 1998년 이후 꾸준히 증가해 왔지만, 대부분의 연구는 특정 요소와의 관계를 분석하는 데에 초점을 맞추었으며, 다양한 요인과의 관계에 관한 연구는 제한적이었다.

따라서, 본 연구에서는 기계학습을 활용하여 로드킬 발생을 예측하고 로드킬에 영향을 미치는 여러 요인들을 탐색하고자 하였다. 머신러닝 모델은 KNN, 의사결정트리, 랜덤포레스트, GBM, XGBoost를 사용하였다. 가장 높은 성능을 보인 모델은 GBM으로 정확도 92.0%, 재현율 84.6%, F1-Score 89.2%, AUC는 0.907이었다.

GBM 모델을 기준으로 SHAP 변수중요도 분석 결과, 평균 현지기압(hPa), 평균 지면온도(°C), 월, 평균 이슬점온도(°C), 중앙분리대, 평균 풍속(m/s) 순으로 예측에 영향을 많이 미치는 요인으로 나타났다. 기압이 높은 곳에서 로드킬이 발생하는 것은 동물 종의 서식지 특성과 먹이자원이 존재하는 고도대가 다르기 때문이다(Do & Yoo, 2014). Table 4은 평균 현지기압(hPa)을 다섯 개의 구간으로 나누고, 기압 구간별 로드킬 발생 여부에 따른 건수의 비율을 나타낸다. 평균 현지 기압이 1023.66hPa~1035.2hPa 구간에서만 유일하게 로드킬 발생 비율이 로드킬이 발생하지 않은 비율보다 높음을 확인할 수 있다. 이는 비교적 높은 기압에서 로드킬 건수가 많음을 나타내고, 기압에 따라 동물의 행동 패턴 변화가 일어나 로드킬 건수가 증가했다고 추측할 수 있다.

월 변수의 경우 월이 변화함에 따라 야생동물의 번

Table 4. Roadkill occurrence ratio by 5 intervals of average local atmospheric pressure (hPa)

Average Local Pressure (hPa)	Roadkill	
	Not occurred (0)	Occurred (1)
(977.442, 989.04]	61 (0.54)	49 (0.46)
(989.04, 1000.58]	2,334 (0.73)	875 (0.27)
(1000.58, 1012.12]	4,600 (0.60)	3,029 (0.40)
(1012.12, 1023.66]	2,627 (0.60)	1,806 (0.40)
(1023.66, 1035.2]	117 (0.24)	361 (0.76)

식기가 종별로 다르고, 교통량 또한 월별로 다르다. 이는 로드킬이 주로 발생하는 기간에 집중적으로 방지할 수 있는 정책 마련이 필요함을 시사한다. 도로 내 중앙분리대는 교통안전 증진을 목적으로 설치되었지만 야생동물이 도로를 횡단하는 경우 머무르는 시간을 증가시켜 차량과 충돌할 가능성을 높일 수 있다. 따라서, 로드킬 예방을 위해 중앙분리대 설치 시 교통량, 도로 여건 뿐 아니라 로드킬 발생 빈도를 종합적으로 고려하여야 한다.

본 연구의 결과를 토대로, 로드킬에 영향을 미치는 요인을 고려하여 생태통로 설치 및 로드킬 예방을 위한 울타리 설치와 같은 대책을 마련하는 데 기초 자료로 활용될 수 있을 것으로 기대하며, 로드킬 사고를 최소화하기 위한 전략을 개발하는 데 도움이 될 것으로 기대된다. 더불어, 국내의 지역 특성을 고려하고 직접 데이터셋을 구축한 점에서 특정 종이나 지역에 국한되지 않고 다양한 생태학적 맥락에 적용 가능한 예측 모델의 개선 등 향후 연구에 있어 높은 활용 가능성을 기대할 수 있다.

본 연구의 한계점으로는 로드킬이 도로 위에서 일어나는 사건이나 교통량을 특성으로 활용하지 못했다는 점이다. 연구 대상을 일반국도와 고속도로로 선정함에 있어 일반국도의 교통량 집계 방식이 본 연구에서 활용이 어렵다고 판단하여 사용하지 못하였다. 향후 일반국도의 평균 교통량으로 대체하거나 연구 대상을 고속도로만으로 특정함으로써 추가 연구가 가능할 것으로 보인다.

향후 연구로는 인구밀도, 동물 행동 패턴, 교통량 등 더 다양한 변수를 고려하여 모델을 더욱 정교하게 개선 예정이다. 또한, 현재 연구는 충청남도를 대상으로

분석했지만, 연구대상지를 확장할 예정이며, 특정 종에 국한된 로드킬 예측 모델을 일반화하는 연구를 수행하는 등 넓은 응용 가능성을 모색할 수 있다.

## 사사

본 연구는 정부(과학기술정보통신부)의 재원으로 한국연구재단의 지원을 받아 수행하였습니다(RS-2023-00242528).

## References

- Choi TY, Park JH. 2006. The Effects of Land Use on the Frequency of Mammal Roadkills in Korea. *Journal of the Korean Institute of Landscape Architecture*, 34(5): 52-58. [Korean Literature]
- Do MS, Yoo JC. 2014. Distribution Pattern According to Altitude and Habitat Type of the Red-tongue Viper Snake (*Gloydius Ussuriensis*) in the Cheonma Mountain. *Journal of Wetlands Research*, 16(2): 193-204. [Korean Literature]
- Ha H, Shilling F. 2018. Modelling potential wildlife-vehicle collisions (WVC) locations using environmental factors and human population density: A case-study from 3 state highways in Central California. *Ecological Informatics*, 43: 212-221.
- Kim M, Park H, Lee S. 2021. Analysis of Roadkill on the Korean Expressways from 2004 to 2019. *International Journal of Environmental Research and Public Health*, 18(19): 10252.
- Kweon HK, Choi YH, Kim MJ, Lee JW. 2008. Study on the Status and Cause of the Road Kill for Wildlife Killing Reduce – A Case Study of National Road in Daejeon~Seosan Section. *Journal of Forest and Environmental Science*, 24(2): 99-109. [Korean Literature]
- Lee GJ, Tak JH, Park SI. 2014. Spatial and Temporal Patterns on Wildlife Road-kills on Highway in Korea. *Journal of Veterinary Clinics*, 31(4): 282-287. [Korean Literature]
- Lundberg SM, Lee SI. 2017. A unified approach to interpreting model predictions. *Advances in neural information processing systems*, p. 30.
- Min JH, Han GS. 2010. A Study on the Characteristics of Road-kills in the Odaesan National Park. *Korean Journal of Environment and Ecology*, 24(1): 46-53. [Korean Literature]
- Ministry of Land, Infrastructure and Transport. 2022. Road Status Report 2022. General Statistics No. 116006. [Korean Literature]
- National Institute of Ecology. 2022. Intensive Survey on Road-kill Hotspots in South Korea. [Korean Literature]
- Son SW, Kil SH, Yoon YJ, Yoon JH, Jeon HJ, Son YH, Kim MS. 2016. Analysis of Influential Factors of Roadkill Occurrence – A Case Study of Seorak National Park. *Journal of the Korean Institute of Landscape Architecture*, 44(3): 1-12. [Korean Literature]
- Song EG, Seo HJ, Kim KM, Woo DG, Park TJ, Choi TY. 2019. Analysis of Roadkill Hotspot According to the Spatial Clustering Methods. *Journal of Environmental Impact Assessment*, 28(6): 580-591. [Korean Literature]
- Song JS, Lee KJ, Ki KS, Jun IY. 2011. The Efficiency and Improvement of the Highway Wild-Life Fences for Decrease of Mammals Road-kill – In Case of Manjong~Hongchun Section on Jungang Highway -. *Korean Journal of Environment and Ecology*, 25(5): 649-657. [Korean Literature]
- Pagany R, Dorner W. 2019. Do crash barriers and fences have an impact on wildlife-vehicle collisions? – An artificial intelligence and GIS-based analysis. *ISPRS International Journal of Geo-Information*, 8(2): 66.
- Pagany R, Valdes J, Dorner W. 2020. Risk Prediction

of Wildlife-vehicle Collisions Comparing  
Machine Learning Methods and Data Use. In  
2020 10th International Conference on Advanced

Computer Information Technologies (ACIT)  
436-440. IEEE.