

항만 도시 교통물류 안전 증진을 위한 실시간 기상 변화 및 항만 영향권 특성 별 사고 영향요인 분석에 관한 연구

박누리¹, 박준영^{2*}

¹한양대학교 ERICA 스마트시티공학과, ²한양대학교 ERICA 교통물류공학과 및 스마트시티공학과

Advanced computer learning-based accident severity analysis considering weather changes and port influence areas: Focused on South Korea Cases

Nuri Park¹, Juneyoung Park^{2*}

¹Department of Smart City Engineering, Hanyang University ERICA

²Department of Transportation and Logistics Engineering/Smart City Engineering,
Hanyang University ERICA

Port safety management should consider a variety of cargo shifting within trucks and containers, occurring at and near port areas. In particular, it is crucial for port safety management to consider not only incidents directly 'at-port' but also those in the surrounding 'near-port' areas, including the port influence area. This is significant because of the potential for high crash severity at near port areas, given the substantial truck traffic that could lead to large-scale crashes. Therefore, developing management strategies for port city safety requires identifying key risk factors that influence crash severity in each port area. During this process, because the key factors influencing crash severity may vary as one gets closer to the port center, it is essential to take into account the size of the port influence area. This study collected and matched both crash and weather data to consider various variables. Additionally, this study developed four machine learning-based crash severity models, including Naive Bayes Classification, Support Vector Machine, Extreme Gradient Boosting, and Light Gradient-Boosting Machine. Furthermore, the identification of key factors influencing high crash severity is determined through the application of an eXplainable Artificial Intelligence technique. It is expected that findings derived from this study can contribute to policy-making efforts aimed at enhancing traffic safety in the port area.

Keywords: Port safety, Crash severity model, Weather data, Traffic safety, Machine learning

논문접수일 : 2023.10.20. 논문수정일 : 2024.03.10. 게재확정일 : 2024.03.21.

이 성과는 정부(과학기술정보통신부)의 재원으로 한국연구재단의 지원을 받아 수행된 연구임(2022R1A2C1093424).

1. 한양대학교 ERICA 스마트시티공학과 박사과정

2*. 한양대학교 ERICA 교통물류공학과 및 스마트시티공학과 교수, Corresponding Author: juneyoung@hanyang.ac.kr

1. 서론

1.1 연구배경 및 목적

항만은 해상과 내륙을 통하는 물류의 중계지로, 항만 인근 도로에는 화물 운송을 위한 화물차의 교통량이 많다. 화물차의 경우 큰 차체로 인하여 일반 승용차에 비해 대형 사고를 초래할 가능성이 크다. 실제로 도로교통공단 사고분석시스템(Traffic Accident Analysis System, TAAS)에 따르면 2022년 교통사고 통계 결과 승용차 사고의 경우 전체 사고 중 사망자 또는 중상자 수가 사고 100건당 약 28명으로 나타났으나, 화물차 사고의 경우 사고 100건당 약 32명의 사망자 및 중상자가 발생한 것으로 나타났다. 이처럼 심각도 높은 사고를 초래할 수 있는 화물차 교통에 대한 안전 관리가 필요하며, 다량의 화물차 교통량의 영향을 받는 항만 인근 도로에서의 교통안전에 대한 각별한 주의가 필요하다.

하지만 현재 항만 도시에 초점을 맞춘 교통안전 연구는 대부분 항만 내에서 발생하는 작업자와 장비, 차량과의 상충으로 인해 발생하는 사고를 다루고 있으며, 항만 영향권 내 인근 도로를 고려한 연구는 미비한 수준이다 (Fabiano et al., 2010; Lee and Ha, 2021; Budiyanoto and Fernanda, 2020). 항만 근접 지역에서는 화물차 등 대형차량의 이동이 빈번하여 사고 발생 시 심각한 사고로 이어질 가능성이 크다. 실제로 한국의 주요 5대 항만(부산항, 울산항, 평택항, 인천항, 광양항)이 위치한 도시에서의 5년간의 사고 통계를 분석한 결과, 항만 근접 지역에서는 면적 당 화물차 사고 수가 31건으로 항만 근접 지역 외 도로 구간의 면적당 화물차 사고 수인 3건보다 높게 나타났으며, 중상 이상의 심각한 사고 발생 비율이 약 2% 높게 나타났다. 따라서 항만이 위치한 도시의 교통안전을 위해 사고 심각도를 낮추기 위한 각별한 노력이 필요하다. 또한 바다와 인접해 있는 항만 도시의 경우 내륙과는 다르게 해무와 같은 기상 요인의 영향을 받을 수 있으므로 교통사고 심각도 분석에 기상 조건 또한 고려되어야 한다. 이를 위해 항만 도시에서 항만 근접 영향권 별 사고 심각도 예측 모형을 개발하고, 다양한 기상 관련 요인을 포함하여 심각한 사고에 영향을 미치는 요인을 도출할 필요가 있으며, 특히 항만 중심지(본 연구에서는 항만 터미널이 위치한 곳을 항만 중심지로 정함)에 가까운 도로에서 발생한 사고의 심각도에 미치는 영향요인을 살펴볼 필요가 있다.

본 연구에서는 항만의 화물 운송의 영향을 받는 항만 영향권을 설정하고, 항만 영향권별로 사고 심각도를 예측하는 모형을 개발하였다. 예측 모형의 정확도 향상을 위해서는 다음과 같은 4가지 머신러닝 방법론을 활용하였다; Naive Bayes Classification (NBC), Support Vector Machine (SVM), Extreme Gradient Boosting (XGBoost), Light Gradient-Boosting Machine (LightGBM). 머신러닝 모형은 기존 사고 심각도 예측에 사용되던 다양한 통계 모형에 비해 높은 정확도를 보인다는 장점이 있으나, 모형 구축에 사용된 독립변수와 종속변수 간의 관계를 설명할 수 있는 통계 모형과는 달리 설명 불가능한 블랙박스 모형이라는 단점이 있었다. 하지만 최근에는 설명할 수 있는 인공지능 기법(explainable Artificial Intelligence, XAI)이 등장함에 따라 머신러닝 모형의 예측성능에 주요한 영향을 미친 요인들과 양의, 음의 상관성을 분석할 수 있다. 본 연구에서는 정확도와 신뢰도가 높은 사고 심각도 분류모형을 구축하기 위해 다양한 머신러닝 기법을 활용하였으며, 사망과 중상이 포함된 심각도 높은 사고와 나머지 상대적으로 심각도가 낮은 사고를 분류하는 분류모형을 구축하였다. 그리고 네 가지 머신러닝 모형 중 가장 예측성능이 뛰어난 모형을 대상으로 영향권별 주요 영향요인들을 도출하였다. 마지막으로 도출된 주요 영향요인에 대하여 설명 가능한 인공지능 기법 중 SHapley Additive exPlanations (SHAP) 방법론을 사용해 모형을 해석하였다.

본 연구에서는 먼저, 항만 도시의 항만 영향권을 설정하고 머신러닝 기법을 활용하여 항만 영향권별 예측 정확도가 높은 심각도 예측 모형을 개발하였으며, 이 과정에서 기상청에서 수집하는 기상 빅데이터를 활용해 항만 지역에서 사고 심각도에 영향을 끼치는 주요 기상 요인을 도출할 수 있도록 하였다. 그 후 설명할 수 있는 머신러닝 기법을 사용해 사고 심각도에 영향을 미치는 주요 요인을 도출하였으며, 항만 영향권에 따른 사고 심각도 영향요인과 해당 요인을 개선 또는 제거할 수 있는 교통안전 관리 전략을 제시하였

다. 이를 통해 항만 도시에서, 특히 기상 변화와 항만 영향권 특성을 고려하여 도로 교통안전을 개선 시키기 위한 근거와 기반을 마련하는 것이 본 연구의 목적이다.

1.2 기존 문헌 고찰

기존 연구에서는 주로 항만과 같이 특수한 용도의 지역을 특정하는 것이 아닌 도로의 종류 또는 위계별로 도로에서 발생한 사고를 대상으로 심각도 모형을 개발하고, 도출된 심각도 모형을 통해 사고 심각도에 유의한 영향을 미치는 주요 요인을 도출하였다. 그리고 그 과정에서 일부 연구들은 모형에 지역적 특성 또는 시공간적 이질성을 반영하기 위해 다단계 모형 등 다양한 방법론을 활용하였다 (Xiong et al., 2014; Xu and Huang, 2017; Gong and Fan, 2017; Park et al., 2017; Claros et al., 2020; Li et al., 2021; Barman et al., 2023). 하지만 특수 지역과 근접한 정도에 따라 영향권을 설정하여 사고 심각도 분석을 수행한 연구는 아직 미비한 실정이다.

교통안전에 주의가 필요한 여러 특수 지역 중에서도 항만의 경우 항만 내외를 오가는 화물차의 통행이 많고, 고속의 연속류가 이동하는 고속도로와 다르게 도심 도로 등 생활권 내의 도로가 분포되어 있어 사고 발생 시 심각도가 높은 대형 사고로 이어질 수 있다. 이에 따라 항만 도시에서의 교통사고 심각도 영향요인 도출에 관한 연구들도 등장하고 있다. Abbassi et al. (2021)은 대형차량의 통행이 잦은 미국 항구도시에 교통사고 심각도 영향요인을 도출하기 위해 모형에 도로 요인뿐만 아니라 다양한 시간적, 환경적 요인을 고려하였으며, CART (Classification and Regression Trees) 방법론을 사용하여 사고 심각도에 영향을 미치는 주요 요인을 도출하였다. 모형 도출 결과 모형의 예측 정확도는 약 62%로 도출되었으며, 주간, 기상, 사고 발생 시각 등이 사고 심각도에 주요한 영향을 미치는 것으로 나타났다 (Abbassi et al., 2021). 이렇듯 항만 도시에서 사고 심각도 영향요인을 도출하는 연구의 필요성은 등장하였으나, 항만이 영향을 주는 인접 도로를 영향권별로 설정하여 영향권에 따른 사고 심각도 주요 영향요인과 심각도를 줄이는 방안에 관해 고찰한 연구는 많지 않다. 본 연구에서는 항만 도시의 교통안전 관리를 위해 사고 심각도에 영향을 미치는 요인을 도출하고자 하였으며, 영향권을 설정하여 항만에 가까운 정도에 따른 심각도 영향요인을 분석하고, 심각도를 낮추기 위한 교통안전 관리 전략을 제안한다는 점에서 기존 연구와 차별성이 있다. 또한, 주로 사고자료에 포함된 도로, 교통, 환경 정보를 사용한 기존 연구와 다르게 별도의 실시간성이 반영된 구체적인 기상 데이터를 수집하여 사고 심각도에 주요한 영향을 미치는 환경적 요인을 분석하고자 하였다.

본 연구에서는 교통사고 심각도 예측 모형과 관련된 기존 연구뿐만 아니라 항만 영향권 설정 기준에 관한 기존 문헌을 고찰하였다 (Table 1.). 기존 문헌 고찰 결과 선행 연구들은 다양한 기준으로 항만 배후도로 등 항만 인근지역을 정의하고 분석을 수행하였으나, 명확하게 항만 영향권을 설정하지 않고 행정구역을 기준으로 항만 인접 지역을 설정하거나 환경 보호청 기준을 따라 영향권을 설정하였다 (Jung, 1999; Lee et al., 2019; Ziemska, 2021). 본 연구에서는 기존 연구 중 교통영향평가 지침을 참고한 논문을 참고하여 항만의 최근접 영향권을 항만 중심지로부터 반경 2km 이내로 설정하였으며, 영향권 크기에 따른 사고 심각도와 심각도 주요 영향요인을 도출하기 위해 영향권을 추가로 4km, 8km로 설정하여 분석을 수행하였다. 또한, 도시 수준(City level)의 항만 영향권에서 벗어난 지역에서 발생하는 교통사고 심각도 영향요인과의 차이를 살펴보기 위해 City level도 별도로 모형을 구축하여 분석을 수행하였다.

Table 1. Previous studies on the port influence area

Authors	Contents	Port influence area
Jung (1999)	• Review of impact on at-grade intersections on major arterial roads around Busan Port	The roads where container movement occurs frequently
Lee and Kim (2009)	• Analysis of the operating conditions of cargo vehicles in Incheon City through surveys and qualitative and quantitative analysis of the resulting damage	Area adjacent to port based on administrative district
Lee et al. (2019)	• Present ideas for estimating environmental impacts • Establishing policies due to increased cargo transport activities between ports and hinterlands	Area of influence within 0.5km from the port gate
Ziemska. (2021)	• Analysis of the environmental impact of large freight vehicles in port cities using traffic flow simulation VISSIM	Road network including 5 intersections

2. 분석 데이터

2.1 데이터 수집 및 전처리

항만 인근 도로의 사고 심각도 모형을 개발하기 위해 도로교통공단 사고분석시스템에서 GIS 분석을 통해 2018년부터 2022년까지의 교통사고 데이터를 수집하여 활용하였으며, 같은 기간으로 종관기상관측(Automated Synoptic Observing System, ASOS) 기상 데이터를 수집하여 사고데이터와 매칭을 수행하였다. 데이터 수집 범위의 경우 한국 5대 항만(부산항, 울산항, 평택항, 인천항, 광양항)을 대상으로 하였으며, 기존 문헌에서 고찰한 대로 항만 영향권 범위를 2km, 4km, 8km로 설정하여 데이터를 수집하였다. 또한, 항만 영향권 외 도시 수준에서 교통사고 심각도 영향요인을 도출했을 때 차이점을 분석하기 위해 각 항만이 위치한 도시 수준의 데이터도 수집하여 활용하였다.

교통사고 데이터의 경우 사고 발생 일시와 가해자, 피해자의 인적 정보, 차종, 위반정보, 사고 발생 당시의 노면 상태 및 기상 상황, 사고유형 등의 정보를 포함하고 있다. 하지만 사고 발생 당시의 기상의 경우 맑음, 흐림, 비, 눈 정도로만 구분되어 있어 구체적으로 날씨가 사고 심각도에 미치는 영향을 포착하기 어려울 수 있다. 항만 영향권의 경우 내륙 지역에 비해 기온의 일교차가 적은 특징이 있으나, 해안과 내륙과의 온도 차이로 인해 바람의 세기 및 방향에 영향을 줄 수 있고, 안개 등이 발생할 수 있기 때문에 시정거리와 같은 세부적인 날씨 정보를 같이 고려할 필요가 있다. 따라서 본 연구에서는 사고 심각도 예측 모형 개발 및 사고 심각도 주요 영향요인 도출에 교통사고 데이터뿐만 아니라 기상청에서 제공하는 기상 데이터를 함께 활용하였다. 교통사고 데이터의 경우 시간 단위로 기록되기 때문에 기상청에서 제공하는 기상 데이터 중 시간 단위로 넓은 범위의 기상을 관측하는 ASOS 자료를 사용하였으며, ASOS의 경우 시정거리(10m) 정보를 포함하고 있다. ASOS는 2021년 기준으로 전국 98개의 관측 지점이 있으며, 98개의 관측 지점 중 한국 5대 항만이 위치한 행정구역 또는 가장 인접해 있는 행정구역에 위치한 관측 지점의 기상 정보를 수집하여 활용하였다 (Figure 1). 5개의 항만 인접 지역에서 수집된 기상 데이터는 총 219,053건으로, 사고가 발생한 시간 및 위치와 매칭하여 분석을 수행하였다.

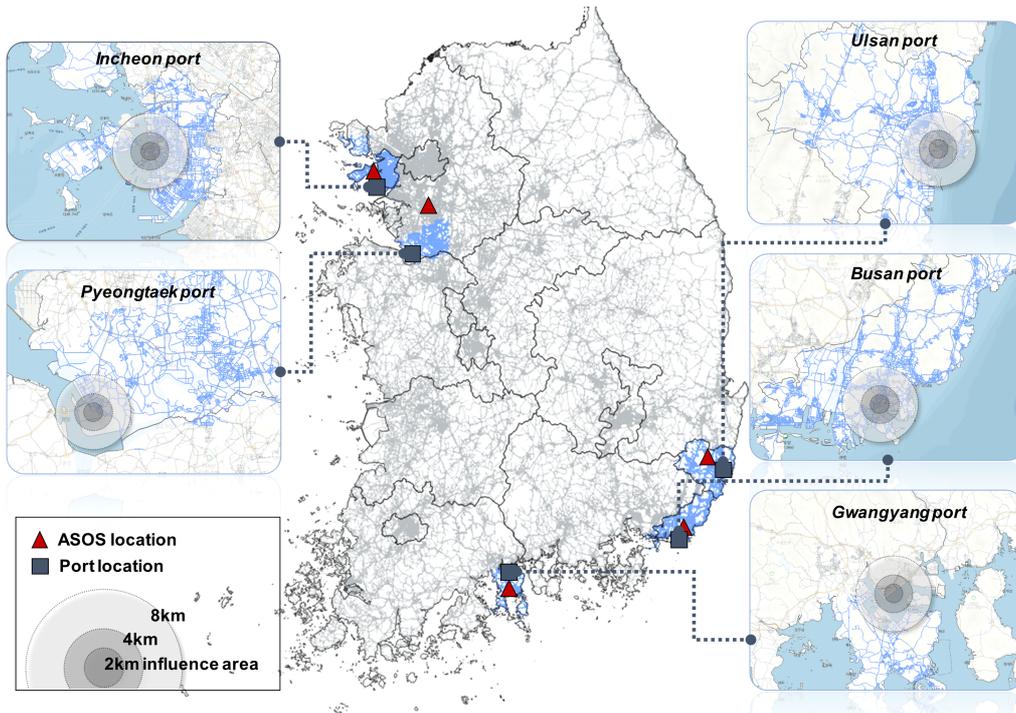


Figure 1. Data collection spatial range

수집된 교통사고 데이터는 심각도를 KABC로 분류하며, K는 사망사고(Fatal injury), A는 중상 사고(Serious injury), B는 경상 사고(Slightly injury), C는 부상 사고(Possible injury)에 해당한다. 본 연구에서는 사망사고와 중상사고를 묶어 심각도 높은 사고(KA)로, 경상 사고와 부상 사고를 묶어 상대적으로 심각도가 낮은 사고(BC)로 분류하였으며, 두 가지 심각도의 사고를 분류할 수 있는 이진 분류 모형을 개발하였다. 데이터 수집 기간 동안 총 25,377건의 KA 사고를 포함한 120,338건의 사고를 수집하였으며, 항만 영향권에 따라서는 2km, 4km, 8km, 지역 전체 영향권에서 1km² 면적 당 약 2.0건, 2.9건, 2.3건, 0.2건의 사망사고가 발생하였다. 이는 항만 영향권 외 전체 도시 지역과 비교할 때 항만 영향권에서 면적 당 심각도 높은 사고의 발생이 많아지는 것을 의미하며, 영향권 별로, 특히 4km의 항만 근접 영향권에서 사고 심각도를 줄이기 위한 노력이 필요함을 보여준다.

2.2 변수 설정

사고데이터와 시간대별 기상 데이터를 결합하여 통합 데이터를 구축한 후 구축된 통합 데이터를 기반으로 인적 요인(Driver characteristics), 차량 요인(Vehicle characteristics), 사고 관련 요인(Crash characteristics), 도로 요인(Road characteristics), 기상 및 환경 요인(Environmental characteristics)을 포함한 다양한 변수를 구축하고, 이를 사고 심각도 모형 개발에 활용하였다. 변수 설정과 사고 심각도별 기초 통계에 관한 표는 Table 2와 같다. Table 2에서 Percentage는 전체 사고 중 각 요인의 비율을 의미하며, Frequency는 각 요인 별, 항만 영향권 별 1km² 면적 당 사고 빈도수를 의미한다. 전반적으로 항만 영향권에서 면적 당 사고 빈도가 높게 나타났다. 본 연구에서는 Table 2에 제시된 81개의 변수 중 상관분석 결과 상관성이 높게 도출된 변수를 제외하고 사고 심각도 예측 모형을 구축하였다.

Table 2. Variable set

Category	Variable	Description	Percentage	Frequency (per 1km ²)				
				2k m	4k m	8k m	Cit y	
Driver characteristics (4)	Offender	sex_offender	1 if sex of offender is male; 0 others	78.67%	214	243	198	15
		age_offender	1 if age of offender is youth (<19)	1.74%	4	4	5	0
			2 if age of offender is young (19-34)	22.37%	55	62	57	4
			3 if age of offender is middle age (35-64)	61.86%	162	186	153	12
	4 if age of offender is older (>64)	14.03%	45	47	36	3		
	Victim	sex_victim	1 if sex of victim is male; 0 others	71.75%	197	220	179	13
		age_victim	1 if age of victim is youth (<19)	3.85%	7	8	10	1
			2 if age of victim is young (19-34)	24.61%	58	70	63	5
3 if age of victim is middle age (35-64)			58.58%	159	175	143	11	
4 if age of victim is older (>64)	12.95%	41	46	35	2			
Vehicle characteristics (15)	Vehicle type	veh_offender_pc	1 if passenger car	68.25%	183	200	170	13
		veh_offender_van	1 if van	4.82%	11	13	11	1
		veh_offender_freight	1 if freight car	12.32%	33	37	29	2
		veh_offender_two	1 if motorcycle	9.04%	19	29	28	1
		veh_offender_pm	1 if bicycle or personal mobility	2.06%	3	5	5	0
		veh_offender_special	1 if special vehicles (agricultural, etc.)	3.45%	15	14	8	1
		veh_offender_unclassified	1 if unclassified	0.07%	0	0	0	0
		veh_victim_pc	1 if passenger car	50.50%	142	150	118	10
		veh_victim_van	1 if van	3.63%	10	11	8	1
		veh_victim_freight	1 if freight car	5.73%	16	16	12	1
		veh_victim_two	1 if motorcycle	12.02%	24	37	35	2
		veh_victim_pm	1 if bicycle or personal mobility	3.31%	4	8	8	1
		veh_victim_special	1 if special vehicles (agricultural, etc.)	2.42%	9	10	6	0
		veh_victim_pedestrian	1 if pedestrian	22.32%	60	67	63	4
veh_victim_unclassified	1 if unclassified	0.07%	0	0	0	0		

Category	Variable	Description	Percentage	Frequency (per 1km ²)				
				2km	4km	8km	City	
Crash characteristics (21)	Crash type	vtov_rearend	Vehicle to vehicle (1 if rear-end crash)	16.97%	43	44	40	3
		vtov_headon	Vehicle to vehicle (1 if head-on crash)	3.55%	7	11	8	1
		vtov_reverse	Vehicle to vehicle (1 if driving in reverse)	1.49%	4	4	4	0
		vtov_side	Vehicle to vehicle (1 if side crash)	35.70%	102	122	89	7
		vtov_others	Vehicle to vehicle (1 if other crash type)	19.96%	49	51	46	4
		vtop_roadside_area	Vehicle to pedestrian (1 if vehicle to pedestrian crash on the roadside area)	1.55%	3	6	5	0
		vtop_sidewalk	Vehicle to pedestrian (1 if vehicle to pedestrian crash on the sidewalk)	1.24%	4	3	4	0
		vtop_road	Vehicle to pedestrian (1 if vehicle to pedestrian crash on the road)	2.40%	11	8	7	0
		vtop_crosswalk	Vehicle to pedestrian (1 if vehicle to pedestrian crash in a crossing)	8.76%	21	27	25	2
		vtop_others	Vehicle to pedestrian (1 if other crash type)	8.37%	21	23	23	2
	Violation	vi_speeding	1 if speeding	0.32%	1	1	1	0
		vi_intersection	1 if violation of driving method in intersection	6.79%	32	23	15	1
		vi_pedestrian	1 if violation of pedestrian protection obligations	4.20%	12	12	11	1
		vi_urn	1 if illegal U-turn	0.56%	1	2	1	0
		vi_signal	1 if signal violation	13.36%	33	41	33	3
		vi_safe_distance	1 if violation of to maintain safe distance	10.21%	33	35	24	2
		vi_unsafe	1 if unsafe driving	53.24%	125	151	137	10
		vi_center_line	1 if violation over the center line	3.82%	9	12	9	1
		vi_rightturn	1 if crash with right-turn veh and straight veh	2.33%	4	7	6	0
		vi_lane_compliance	1 if violation of lane compliance obligations	2.23%	9	7	6	0
		vi_others	1 if unclassified	2.94%	6	8	7	1

Category	Variable	Description	Percentage	Frequency (per 1km ²)				
				2k m	4k m	8k m	Cit y	
Road characteristic (14)	Road surface condition	dry	1 if road surface is dry	90.64%	240	270	227	17
		wet	1 if road surface is wet	8.64%	24	28	22	2
		snowy/icy	1 if road surface is snowy or icy	0.28%	0	1	0	0
		weather_others	1 if others	0.44%	0	1	1	0
	Road geometry	intersect_near	1 if crash occurred near the intersection	15.78%	36	49	40	3
		intersect_inside	1 if crash occurred inside the intersection	31.36%	79	93	78	6
		intersect_crosswalk	1 if crash occurred in the intersection crosswalk	3.37%	7	8	10	1
		seg_overpass	1 if crash occurred in the overpass segment	0.44%	5	2	1	0
		seg_bridge	1 if crash occurred on the bridge	0.76%	2	3	1	0
		seg_underpass	1 if crash occurred in the underpass segment	1.64%	3	3	5	0
		seg_tunnel	1 if crash occurred in the tunnel	0.40%	2	1	1	0
		seg_others	1 if crash occurred in the other segment	41.52%	121	127	103	8
		parking_lot	1 if crash occurred in the parking lot	0.45%	2	1	1	0
		road_others	1 if others; 0 otherwise	4.27%	8	11	10	1
Environmental characteristic (27)	Region	pyeongtaek	Port located region (1 if pyeongtaek)	13.19%	13	11	7	4
		busan	Port located region (1 if busan)	46.72%	182	189	122	8
		ulsan	Port located region (1 if ulsan)	15.28%	8	47	66	2
		incheon	Port located region (1 if incheon)	22.33%	9	33	52	5
		gwangyang	Port located region (1 if gwangyang)	2.48%	54	19	3	0
	Season	spring	1 if spring; 0 others	24.84%	63	75	62	5
		summer	1 if summer; 0 others	25.22%	66	74	63	5
		fall	1 if fall; 0 others	26.32%	71	77	67	5
		winter	1 if winter; 0 others	23.61%	65	73	59	4
	Week	week	1 if weekday; 0 others	74.81%	199	227	188	14

Category	Variable	Description	Percentage	Frequency (per 1km ²)				
				2km	4km	8km	City	
Environmental characteristic (27)	Time	nonpeak_morning	1 if crash in the early morning indicator (0-6)	11.77%	28	37	31	2
		peak_morning	1 if crash in the morning peak (7-8)	8.80%	24	27	21	2
		nonpeak_daytime	1 if crash in the daytime off-peak indicator (9-16)	41.56%	121	129	103	8
		peak_afternoon	1 if crash in the afternoon peak indicator (17-18)	14.68%	38	42	36	3
		nonpeak_night	1 if crash in the nighttime off-peak indicator (19-23)	23.19%	54	64	59	4
	Weather	sunny	1 if sunny; 0 others	89.81%	237	266	224	17
		rainy	1 if rainy; 0 others	6.82%	20	23	17	1
		cloudy	1 if cloudy; 0 others	2.56%	7	8	6	0
		snowy	1 if snowy; 0 others	0.19%	0	0	0	0
		others	1 if others; 0 others	0.62%	1	2	2	0
		temperature	hourly temperature (°C)	-	-	-	-	-
		rain	hourly precipitation (mm)	-	-	-	-	-
		wind_speed	hourly wind speed (m/s)	-	-	-	-	-
		humidity	hourly humidity (%)	-	-	-	-	-
		snow	hourly snow (cm)	-	-	-	-	-
visibility	hourly visibility (10m)	-	-	-	-	-		
ground_temperature	hourly ground temperature (°C)	-	-	-	-	-		

* veh-vehicle; offender-offender vehicle; victim-victim vehicle; pc-passenger car; two-two wheel vehicle; pm-personal mobility; vtov-vehicle to vehicle crash; rearend-rear end crash; headon-head on crash; vtop-vehicle to pedestrian crash; vi-violation; intersect-intersection; seg-segment

3. 분석 방법론

본 연구는 머신러닝 기법을 사용하여 항만 도시에서 항만 중심지로부터의 영향권에 따른 도로 교통 사고 심각도 모형을 개발하고, 가장 예측성능이 뛰어난 모형을 선정하였으며, 최적 모형을 대상으로 다양한 도로, 교통, 환경, 인적 요인이 항만 영향권 도로의 사고 심각도에 미친 영향을 분석하였다. 또한, 분석된 영향요인을 바탕으로 항만 지역에서의 교통안전 관리를 위한 정책 마련의 필요성을 제시한다. 본 연구에서는 빅데이터 분석 및 머신러닝 모형 개발을 위한 도구로 Python을 사용하였으며, 전체적인 연구 흐름도는 Figure 2와 같다.

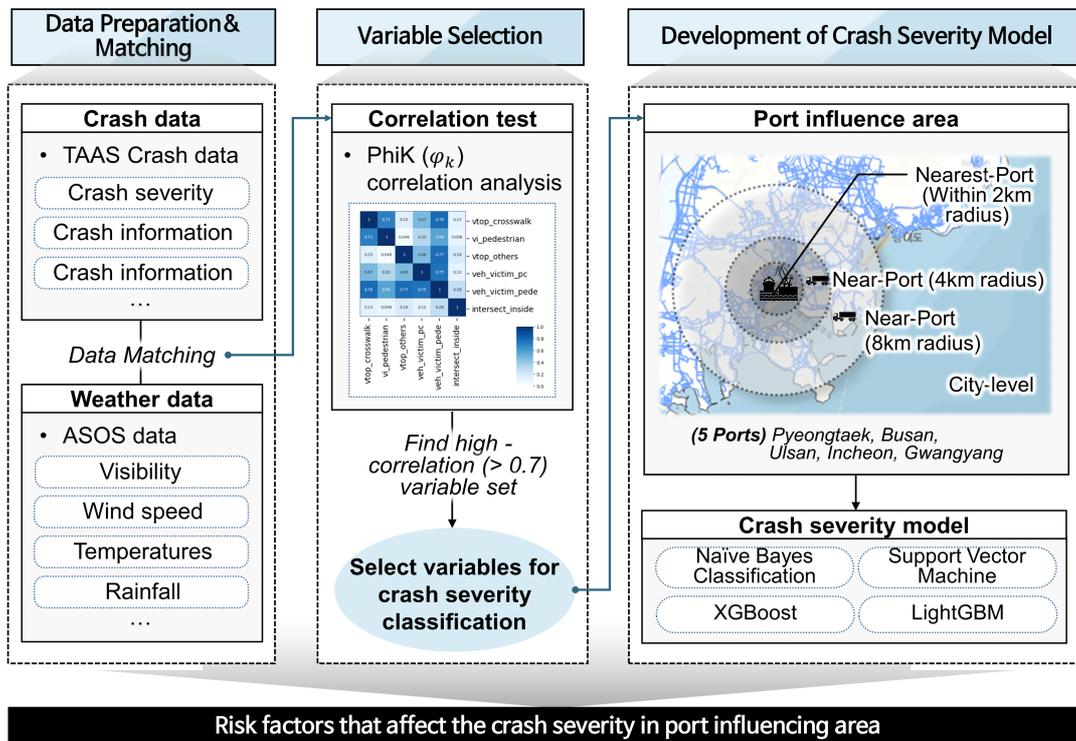


Figure 2. Research framework

3.1 변수 선정

본 연구에서는 사고 심각도 예측 모형에 활용할 변수 선정을 위해 PhiK(ϕ_k) 상관분석을 사용하였다. PhiK 상관분석은 Baak et al. (2020)이 개발한 상관분석 기법으로, 연속형 변수 간의 상관성을 검정하는 Pearson 상관분석과 범주형 변수 간 상관성을 분석하는 Phi 상관분석과 다르게 데이터가 범주형과 연속형 변수로 혼합되어 있을 때 동시에 변수 간 상관성을 분석할 수 있도록 한다 (Baak et al., 2020). PhiK 상관분석의 경우 Pearson 상관 계수와 다르게 0~1의 값을 가지며, 값이 1에 가까울수록 두 변수 간 상관관계가 높음을 의미한다. 본 연구에서는 Rovai et al. (2013)이 제시한 기준에 따라 0.7 이상의 높은 값을 가질 때 상관성이 높다고 판단하였으며, 0.7 이상의 높은 상관관계를 가지는 변수 쌍을 도출하였다.

PhiK 상관분석 결과 같은 카테고리에 속하는 변수를 제외하고 노면 상태 변수(Road surface condition)와 기상 관련 변수 (Weather), 계절 변수(Season)와 기상 관련 변수(Weather)의 상관관계가 높은 것으로 도출되었다. 따라서 기상 관련 변수만을 남기고 기상 변수와 상관성이 높게 도출된 노면 상태 변수와 계절 변수는 분석에서 제외하였다. 0.7 이상의 높은 상관관계를 보인 변수 쌍들을 시각화한 그림은 Figure 3과 같다.

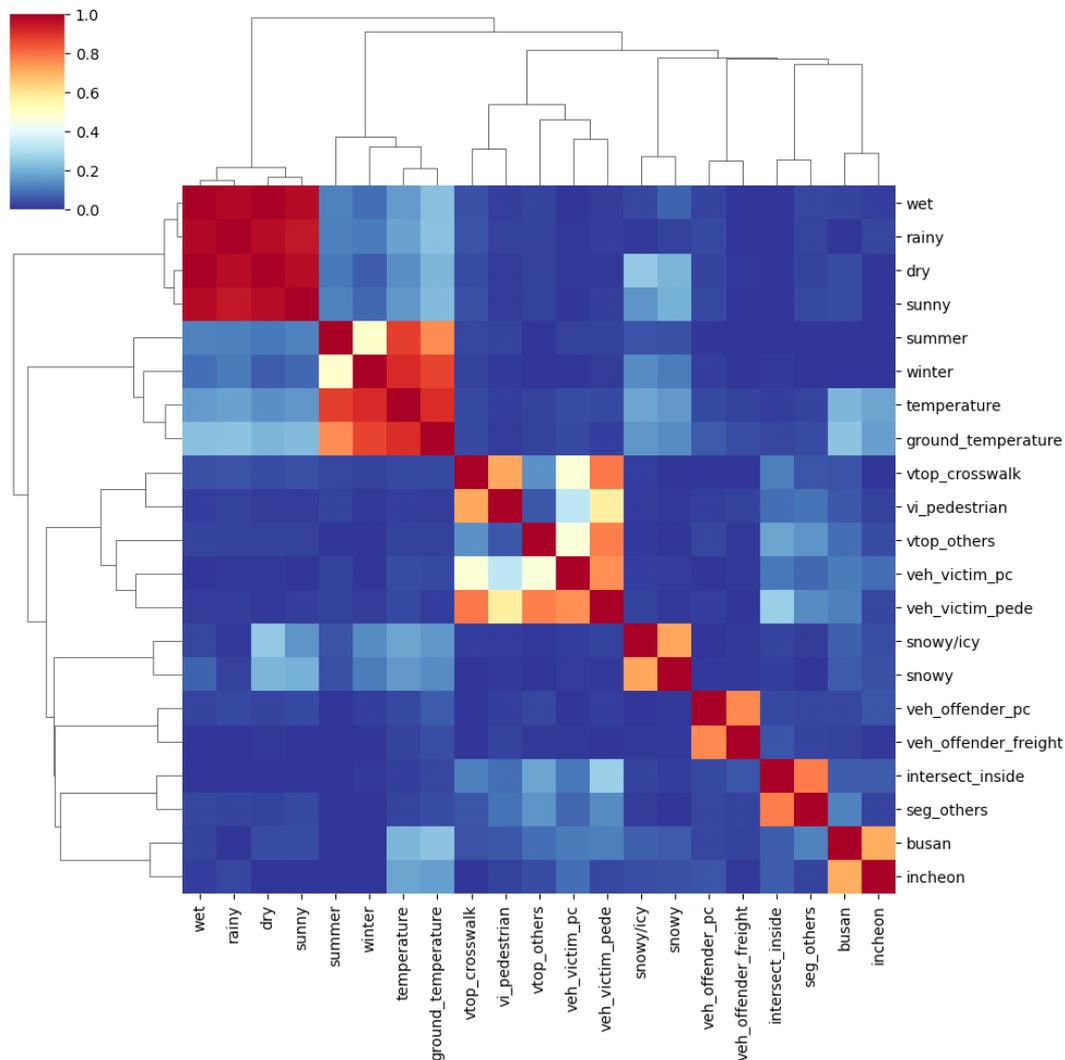


Figure 3. PhiK correlation analysis result

3.2 사고 심각도 예측 모형

1) Naive Bayes Classification (NBC)

Naive Bayes Classification (NBC)은 이진 분류에 많이 사용되는 가장 기본적인 머신러닝 지도학습 방법론의 하나로, 각 특성이 독립적이라고 가정하고 베이저안 방법론에 기반하여 사후 확률을 최대화하는 방향으로 데이터를 분류한다. 베이즈 정리 (Equation 1.)는 사전 정보라고도 불리는 사전 분포를 바탕으로 조건부 확률을 추정하는 것으로, A라는 사건이 발생하기 전 B 사건의 확률분포를 사전 분포(Prior distribution)라고 하고, 사건 A에 대한 정보를 얻은 후의 B 사건의 확률분포를 사후 분포(Posterior distribution)라고 부른다. 여기에서 베이즈 추정이란 사건 A가 주어졌을 때 B 사건의 확률분포를 추정하는 것으로, MCMC(Markov Chain Monte Carlo) 방법을 활용하여 베이저안 추정을 수행할 수 있다 (Besag et al., 1995).

$$P(B|A) = \frac{P(A|B)P(B)}{P(A)} \quad (1)$$

본 연구에서 사용한 NBC 기법의 경우 특성 변수가 너무 많은 경우 이를 단순화하여 빠른 분석이 가능하다는 특징이 있다.

2) Support Vector Machine (SVM)

서포트벡터머신(Support Vector Machine, SVM)은 머신러닝 지도학습 기법의 하나로, 분류 또는 회귀 분석에 사용할 수 있는 지도학습 모형이다 (Boser et al., 1992). SVM에서는 Margin을 이용하여 Hyper plane을 생성해 데이터를 분류하며, Margin을 최대화하는 Hyper plane을 채택하여 서로 다른 카테고리의 데이터를 분류한다. 여기서 Margin은 Hyper plane과 Support vector 사이의 거리를 의미한다. Hyper plane은 데이터를 분류하는 평면이고, Support vector는 이 Hyper plane과 가장 가까운 포인트를 의미한다 (Ukil, A. 2007).

본 연구에서는 SVM의 하이퍼 파라미터 중 저차원 데이터를 고차원으로 매핑할 때 연결해주는 함수인 kernel, 오류 허용 정도를 의미하는 C, Decision boundary의 정밀한 정도를 조정할 수 있는 gamma값을 조정하였으며, 모든 영향권 모형에 대하여 과적합을 방지하기 위해 gamma값을 0.01, C값을 100으로 고정하여 모형을 개발하였다.

3) Extreme Gradient Boosting (XGBoost)

XGBoost는 앙상블 기법의 하나로, 배경 기법인 Random Forest와 달리 여러 개의 결정 트리를 묶어 약한 분류기를 결합해 강한 분류기를 만들어내는 Boosting 기법 기반의 방법론이다 (Tianqi Chen, 2016). XGBoost의 경우 Gradient Boost의 일종으로 병렬 처리를 통해 학습 및 추정을 진행하기 때문에 속도가 빠르며, 예측 신뢰도가 높다는 장점이 있으나, 샘플 수가 적으면 과적합이 일어날 수 있고, 매개변수 설정에 민감하다. 따라서 XGBoost의 하이퍼 파라미터 설정에 주의를 기울여야 하며, 본 연구에서는 XGBoost의 다수의 하이퍼파라미터 중 트리의 개수를 의미하는 n_estimators를 100~750으로, 트리의 깊이를 의미하는 Maximum depth를 3으로 설정하였다.

4) Light Gradient-Boosting Machine (LightGBM)

LightGBM은 XGBoost와 같이 Boosting 기법 기반의 앙상블 방법론이다. LightGBM은 학습 속도가 느리다는 XGBoost의 단점을 보완하게 등장한 방법론으로, XGBoost와 같이 병렬 처리를 통해 빠른 속도를 보여준다. LightGBM 역시 마찬가지로 샘플 수가 적을 경우 과적합이 일어날 수 있기 때문에 하이퍼 파라미터 조정에 유의해야 하며, 본 연구에서는 Random search 방법을 통해 LightGBM 분류 모형 구축을 위한 최적의 하이퍼 파라미터를 도출해 사용하였다. LightGBM에서는 트리의 개수를 의미하는 *n_estimators*를 1000으로, 트리의 깊이를 의미하는 *max_depth*를 146~165로, 전체 트리의 잎의 수를 의미하는 *num_leaves*를 92~196으로 조정하였다.

3.3 예측 성능 평가 및 모형 해석

본 연구에서는 머신러닝 기반의 사고 심각도 예측 모형 개발을 위해 훈련 셋(Train set)과 검증 셋(Test set)을 무작위로 7:3으로 분류하였으며, 검증 셋을 대상으로 모형의 예측성능을 평가하였다. 개발된 사고 심각도 예측 모형은 높은 심각도의 사고와 낮은 심각도의 사고를 분류하는 이진 분류 모형이므로 모형의 예측성능을 평가하기 위해 오차 행렬(Confusion matrix) 기반의 정확도와 재현도(Recall), 정밀도(Precision)를 산출하였다. 여기서 정확도는 전체 데이터 중 옳게 분류한 비율을 의미하며, 재현도의 경우 실제 높은 심각도의 사고 중 옳게 예측한 비율을, 정밀도의 경우 예측된 높은 심각도의 사고 중 옳게 예측된 비율을 의미한다. 오차 행렬 기반의 세 가지 지표뿐만 아니라, 재현도와 정밀도의 조합으로 산출할 수 있는 F1-score와 Youden index를 도출하여 각 사고 심각도 모형의 예측성능을 비교하였으며, 가장 예측성능이 뛰어난 모형에 대하여 설명가능한 인공지능 기반의 해석을 수행하였다. F1-score와 Youden index를 수식으로 나타내면 (Equation 2.)와 (Equation 3.)과 같다.

$$F1 - score = \frac{2 \times (Precision \times Recall)}{Precision + Recall} \quad (2)$$

$$Youden\ index = Recall + Specificity - 1 \quad (3)$$

머신러닝 기법의 경우 Black-box 모형으로 알고리즘 내에서 어떤 변수가 영향을 미쳤는지 확인할 수 없기 때문에 이를 해결하기 위한 방법들이 고안되어왔다. 앙상블 기법의 경우 노드 생성 시의 불순도를 계산할 수 있기 때문에 변수 중요도를 도출할 수 있지만, 변수의 중요도뿐만 아니라 양의 영향을 미쳤는지 음의 영향을 미쳤는지를 판단할 수 없다. 따라서 본 연구에서는 머신러닝 모형에 대한 해석으로 XAI를 사용하였다. XAI의 종류에는 LIME (Local Interpretable Model-agnostic Explanation), SHAP (Shapley Additive exPlanations), Permutation Feature Importance 등 다양한 방법론이 있으며, 본 연구에서는 SHAP 기법을 사용하였다. SHAP 알고리즘에서는 변수의 존재 여부에 따른 Shapley value의 변화를 측정하며, Shapley value를 통해 변수의 기여도를 확인하고 중요도와 영향의 정도, 양과 음의 영향 정도를 확인할 수 있다.

4. 분석 결과

4.1 사고 심각도 예측 모형 구축 결과

네 가지 머신러닝 기법을 사용하여 사고 심각도 예측 모형을 구축한 결과, 높을수록 모형의 정확도와 신뢰성이 높음을 나타내는 F1-score와 Youden index를 확인해보았을 때 (Table 3.)과 같이 2km의 항만 영향권 모형을 제외하고 모든 모형에서 NBC 모형의 예측성능이 높은 것으로 나타났다.

Table 3. Crash severity model performance results

Influence area	Model	Accuracy	Precision	Recall	F1-score	Youden index
2km	NBC	0.321	0.290	0.972	0.446	0.037
	SVM	0.686	0.393	0.396	0.394	0.183
	XGBoost	0.717	0.497	0.305	0.378	0.184
	LightGBM	0.703	0.462	0.326	0.383	0.177
4km	NBC	0.682	0.475	0.505	0.489	0.263
	SVM	0.682	0.470	0.410	0.438	0.210
	XGBoost	0.721	0.565	0.326	0.413	0.217
	LightGBM	0.701	0.508	0.326	0.397	0.189
8km	NBC	0.638	0.428	0.570	0.489	0.237
	SVM	0.676	0.460	0.380	0.416	0.185
	XGBoost	0.720	0.574	0.298	0.392	0.202
	LightGBM	0.703	0.517	0.338	0.409	0.201
City level	NBC	0.678	0.439	0.496	0.466	0.246
	SVM	0.695	0.445	0.319	0.371	0.162
	XGBoost	0.740	0.594	0.263	0.365	0.192
	LightGBM	0.687	0.473	0.327	0.387	0.169

4.2 영향권 별 사고 심각도 영향요인 도출 결과

예측성능이 높게 도출된 항만 영향권별 NBC 모형에 대하여 SHAP 방법론을 통해 사고 심각도 예측에 주요한 영향을 미친 변수들을 각 20개씩 도출하였으며, 도출한 결과는 (Figure 4.)와 같다. SHAP 그래프에서 상위에 존재할수록 중요도가 높은 변수를 의미하며, 붉은색은 각 변수의 값이 클수록, 0을 기준으로 음수 쪽으로 그래프가 치우칠 때 예측에 부정적인 영향을 끼쳤음을 의미하고, 양수 쪽 (오른쪽)으로 치우칠 경우 예측에 긍정적인 영향을 끼쳤음을 의미한다. 예를 들어, 4km 영향권 모델에서 차대 보행자 사고 (보행자 횡단 중) 변수 (vtop_crosswalk)의 경우 붉은색이 양수 쪽에 치우쳐 있으므로 해당 변수의 값이 클수록, 즉, 사고가 차와 횡단 중이던 보행자 간의 사고일 때 높은 사고 심각도와 양의 상관성을 가짐을 의미한다. 도출된 SHAP 그래프를 토대로 영향권별 주요 변수와 높은 사고 심각도 예측에 끼친 영향을 분석한 결과는 (Table 4.)와 같다. Table 4는 각 영향권 사고 심각도 예측의 영향을 주는 변수를 상위 20개씩 정리한 것으로, 중요도가 높은 순서부터 나열하였다. SHAP 그래프 해석 결과를 바탕으로 Table 4에서 각 변수의 영향 (Impact)를 정리하였으며, (-)는 음의 영향을, (+)는 양의 영향을 끼쳤음을 의미한다.

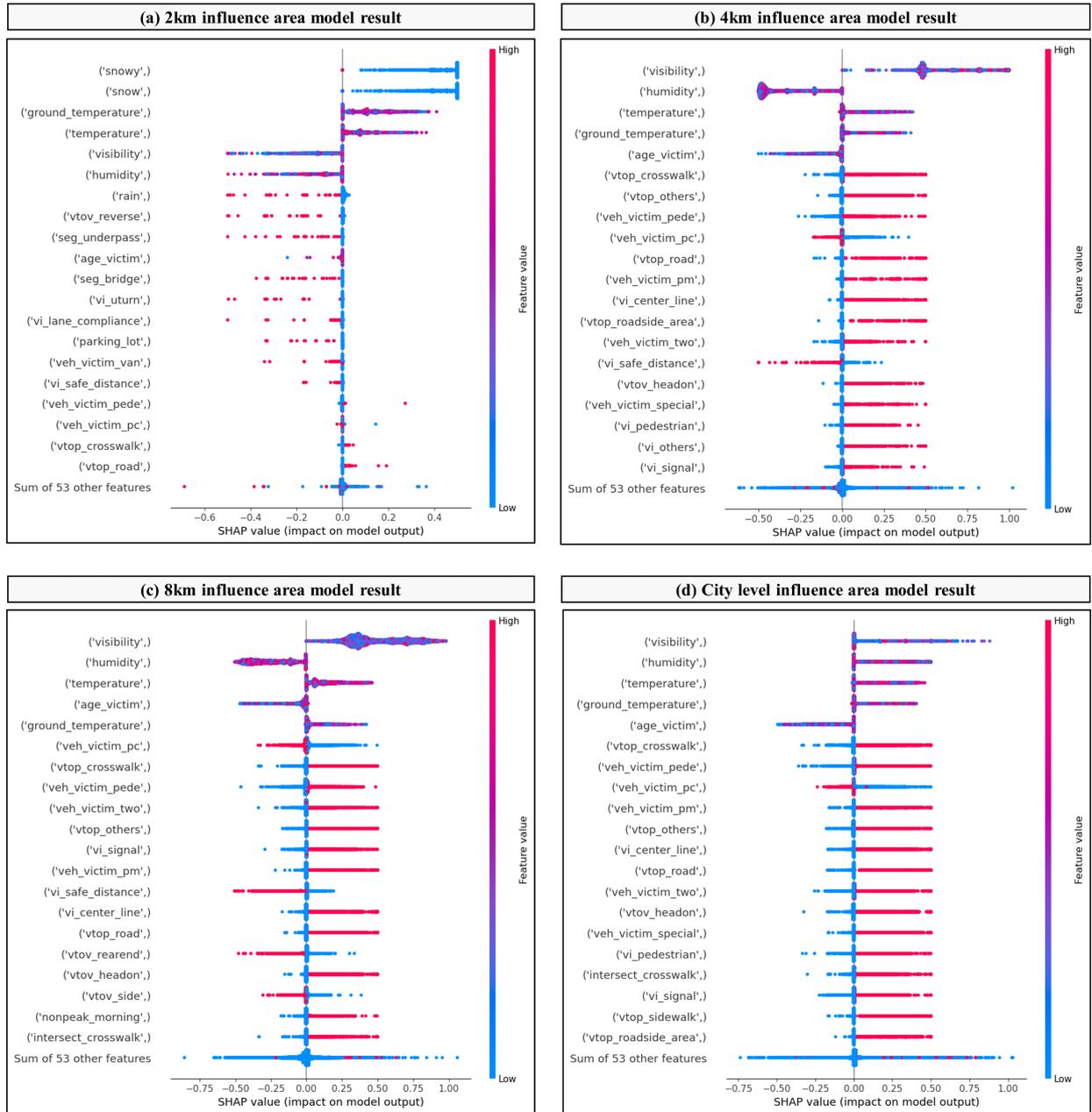


Figure 4. SHAP results (a) 2km influence area model result; (b) 4km influence area model result; (c) 8km influence area model result; (d) City level influence area model result

Table 4. Major variables affecting high crash severity

no	2km influence area model		4km influence area model		8km influence area model		City level model	
	Variable	Impact	Variable	Impact	Variable	Impact	Variable	Impact
1	snowy	(-)	visibility	.	visibility	.	visibility	.
2	snow	(-)	humidity	.	humidity	.	humidity	.
3	ground_temper	(+)	temperature	.	temperature	.	temperature	.
4	temperature	(+)	ground_temper	.	age_victim	.	ground_temper	.
5	visibility	.	age_victim	.	ground_temper	.	age_victim	.
6	humidity	.	vtop_crosswalk	(+)	victim_pc	(-)	vtop_crosswalk	(+)
7	rain	(-)	vtop_others	(+)	vtop_crosswalk	(+)	victim_pede	(+)
8	vtov_reverse	.	victim_pede	(+)	victim_pede	(+)	victim_pc	(-)
9	seg_underpass	.	victim_pc	(-)	victim_two	(+)	victim_pm	(+)
10	age_victim	.	vtop_road	(+)	vtop_others	(+)	vtop_others	(+)
11	seg_bridge	.	victim_pm	(+)	vi_signal	(+)	center_line	(+)
12	uturn	.	center_line	(+)	victim_pm	(+)	vtop_road	(+)
13	lane_compliance	.	vtop_roadside	(+)	safe_distance	(-)	victim_two	(+)
14	parking_lot	.	victim_two	(+)	center_line	(+)	vtov_headon	(+)
15	victim_van	.	safe_distance	(-)	vtop_road	(+)	victim_special	(+)
16	safe_distance	.	vtov_headon	(+)	vtov_rearend	(-)	vi_pedestrian	(+)
17	victim_pede	(+)	victim_special	(+)	vtov_headon	(+)	crosswalk	(+)
18	victim_pc	(-)	vi_pedestrian	(+)	vtov_side	(-)	vi_signal	(+)
19	vtop_crosswalk	(+)	vi_others	(+)	nonpeak_morning	(+)	vtop_sidewalk	(+)
20	vtop_road	(+)	vi_signal	(+)	crosswalk	(+)	vtop_roadside	(+)

* Increase model performance (.); Negative impact (-); Positive impact (+)

모든 영향권 모형에서 전체적으로 지면 온도(ground_temper), 기상(temperature), 습도(humidity), 시정거리(visibility)와 같은 기후 변화 관련 변수의 중요도가 높게 도출되었고, 운전자 특성과 관련해서는 피해자 연령(age_victim), 보행자 피해자(victim_pede), 피해자 차종 승용차(victim_pc) 변수가 높은 사고 심각도 예측에 주요한 영향을 미친 것으로 나타났다. 또한, 사고 특성의 경우 차대 보행자 횡단보도 사고(vtop_crosswalk), 차대 보행자 도로 위 사고(vtop_road) 변수가 중요도가 높은 것으로 나타났다. 전체적으로 항만 도시에서 사고 심각도에 영향을 미치는 변수들을 종합해 보았을 때, 기상 변수는 영향의 방향성이 모호하게 나타났으나 전반적으로 높은 사고 심각도와 양의 상관성을 가지는 것으로 나타났다. 승용차의 경우 낮은 사고 심각도와 상관성을 가지고, 이와 반대로 피해자가 보행자일 때 높은 사고 심각도와 양의 상관성을 가지는 것으로 나타났다. 특히 차대 보행자 사고 중 횡단보도와 도로 위에서 발생한 사고가 높은 심각도의 사고와 관련 있는 것으로 나타나 항만 도시에서는 모든 영향권에서 전반적으로 보행자 보호를 위한 특별한 주의와 안전 관리 대책 마련이 필요한 것으로 도출되었다.

City level을 제외한 항만 영향권에서는 안전거리 미준수(vi_safe_distance) 변수가 높은 사고 심각도와 음의 상관성을 가지는 것으로 나타났다. 즉, 안전거리 미준수 변수의 경우 모두 값이 낮을수록 높은 사고 심각도 예측에 긍정적인 영향을 미치는 것으로 나타나 규칙 위반으로 인한 사고 중에서도 낮은 심각도의 사고로 이어지는 것으로 나타났다. 따라서 안전거리 미준수 외 보행자 보호법 위반, 안전 운전 불이행 등 다른 법규 위반에 초점을 맞추어 이를 단속하려는 노력이 필요한 것으로 도출되었다.

반경 2km에서 4km까지의 항만 근접 영향권의 경우 비교적 차체가 큰 차종이 높은 사고 심각도와 연관

되는 것으로 나타났으며, 특히 4km 영향권의 경우 특수차량이 피해자 차종일 때 (victim_special) 높은 심각도의 사고로 이어지는 것으로 나타났다. 항만 인접 영향권의 경우 상대적으로 화물차와 특수차량의 교통량이 많고, 도시부 도로와 항만으로 이어지는 도로가 고속도로와 비교할 때 다양한 요소에 의하여 방해받기 때문에 차체가 큰 차량이 사고에 개입되면 높은 사고 심각도를 초래할 수 있다. 항만 영향권 내 사고 심각도 모형 중 최근접 영향권(반경 2km)에서의 사고 심각도 모형은 나머지 영향권 모형과는 다른 양상을 보였는데, 강수량(rain), 날씨 눈(snowy), 적설량(snow)과 같은 기상 변수가 높은 사고 심각도에 예측에 주요한 영향을 미치는 것으로 도출되었고, 교량(seg_bridge), 지하도로(seg_underpass), 주차장(parking_lot)과 같은 사고 발생 위치 변수 또한 예측성능에 큰 영향을 미치는 변수로 나타났다. 기상의 경우 눈이 오거나 강수량이 많을 때는 상대적으로 사고 심각도에 음의 영향을 주는 것으로 나타났으나 시정거리의 경우 값이 낮을수록 전반적으로 낮은 심각도의 사고로 이어지는 것으로 도출되었다. 즉, 비, 눈과 같은 악천후 보다는 안개 등 운전자를 방해하면서 운전자에게 부주의를 일으킬 수 있는 기상 상황이 사고 발생 시 높은 심각도의 사고로 이어지게 하는 것으로 나타났다. 따라서 항만 최근접 영향권에서는 심한 악천후 상황에 비해 상대적으로 심하지 않은 악천후 상황일 때 시정거리를 줄일 수 있는 안개 등에 유의해야 하며, 특히 항만 지역의 경우 해무가 자주 발생할 수 있기 때문에 이에 대한 주의 정보를 제공할 필요가 있다. 또한, 심하지 않은 악천후 상황에서 운전자의 부주의를 방지해야 하는 안전 전략을 세울 필요가 있다. 이 외에도 항만 최근접 영향권에서는 후진 중 충돌(vtov_reverse), 차선 준수 위반(lane_compliance) 변수가 사고 심각도 예측 모형의 성능에 주요한 영향을 미치는 것으로 도출되어 항만 내로 직접적으로 유·출입하거나 인접 도로를 통행하는 운수 종사자들에게 운전 부주의와 교통안전에 대한 교육을 제공할 필요가 있는 것으로 나타났다.

이 외에 항만 영향권에서 피해자가 개인형 이동 수단(Personal mobility, PM)을 이용할 때, 피해자 차종이 이륜차일 때 사고 심각도가 높아지는 것으로 나타나 해당 차종의 교통안전 관리에 더욱 주의를 기울여야 하는 것으로 나타났다.

5. 결론

본 연구에서는 항만 도시에서 교통사고 심각도에 영향을 미치는 주요 요인을 항만 중심지로부터의 거리에 따라 살펴보기 위해 항만 지역을 3가지 영향권(반경 2km, 4km, 8km)으로 구분하였으며, 각 영향권 안에 속하는 도로에서 발생한 사고 심각도를 예측하는 모형을 개발하였다. 예측 모형은 이진 분류에 뛰어난 성능을 보이는 머신러닝 기법 중 4가지 모형(NBC, SVM, XGBoost, LightGBM)을 사용하였으며, 사고 심각도 모형 구축 결과 NBC 모형의 성능이 가장 뛰어난 것으로 도출되어 SHAP 방법론을 통해 해당 방법론으로부터 도출된 사고 심각도 모형을 해석하였다.

- 항만 영향권의 경우 안개 등 운전 부주의를 유발할 수 있는 악천후 상황에서 심각한 사고가 발생할 수 있기 때문에 조명시설, 안전표지 등의 개선을 통하여 항만 인근 도로에서 운전자들이 시거를 확보할 수 있도록 해야 하며, 특히 항만 지역의 새벽 시간대의 경우 해안지역에서 발생하는 안개인 해무의 영향을 받을 수 있으므로 기상 정보를 바탕으로 운전자에게 주의를 환기하는 등의 교통안전 관리 전략을 구축할 필요가 있다.
- 반경 2km의 항만 최근접 영향권의 경우 영향권 내에 도로가 적게 존재하지만, 다른 영향권 범위에 비하여 날씨의 영향을 크게 받으므로 미끄럼 방지를 위한 도로 설비, 운전자에게 안전 운전 경고 정보 제공 등의 노력이 필요하다.
- 항만 영향권을 유·출입하는 운수 종사자를 대상으로 한 교통안전 교육을 통해 차선 준수 위반, 불법 유턴, 후진 중 충돌과 같이 운전자 요인으로 인해 발생하는 사고의 발생을 방지하고, 이를 통해 항만 영향

권 도로에서 사고의 심각도를 낮출 수 있도록 해야 한다. 특히 다양한 화물차량들이 집결하는 항만 최근접 영향권에서는 차량이 회피행동을 취하기 위한 충분한 공간을 확보할 수 있도록 주변에 통행을 방해하는 시설물의 유무를 점검할 필요가 있으며, 도로확장 등의 대안을 검토할 필요가 있다.

- 항만 최근접 영향권의 경우 ‘주차장’ 변수가 주요 변수로 도출되었으며, 따라서 항만 인근지역에서의 교통안전을 위해 터미널을 출입하는 화물차를 위한 충분한 주차면 수가 확보되어야 하며, 터미널로 향하는 통로에서의 교통류와 주차장에서 유출입하는 교통류가 상충하지 않도록 주차장 위치의 적정성을 검토해야 한다.

본 연구에서는 실시간 기상 데이터와 사고 이력 데이터를 수집하여 항만 영향권과 항만 도시에서 발생하는 교통사고에 영향을 미치는 주요 요인을 도출하는 연구를 수행하였으며, 다음과 같은 한계점이 존재한다. 첫 번째로, 도로, 교통, 환경, 기상 등 다양한 요인이 사고 심각도에 끼치는 영향을 분석하기 위해 강력한 변수 선정이 아닌 상관분석 기반의 단순한 변수 선정 과정을 거쳤으나, 머신러닝 모형 개발 시 너무 많은 변수의 경우 과적합을 유발할 수 있기 때문에 적절한 변수 선정 과정을 거칠 필요가 있다. 예를 들어, 선행 연구에서는 사고 빈도 모형 또는 사고 위험 예측 모형 개발을 위한 변수 선정을 위해 Random Forest 학습 시 계산되는 Mean Decrease Gini (MDG) 등의 변수를 활용하여 사전에 모형에 중요한 영향을 미치는 변수를 선별하였다 (Lin et al., 2015; Shi and Abdel-Aty, 2015; Wang et al., 2015; You., 2017; Basso et al., 2018; Lei et al., 2021). 이처럼 변수 선정 단계를 다단계화 하여 모형을 단순화 시킴으로써 예측성능을 높일 수 있으며, 보다 정확한 사고 심각도 영향요인 분석이 가능할 것으로 예상된다. 또한, 본 연구에서는 사용한 방법론 이외에 머신러닝 성능에 준하는 통계 모형 기반의 사고 심각도 예측 모형 개발 및 영향 요인 분석 또는 구조방정식 기반의 사고 심각도 영향요인 해석 등을 수행하여 사고 심각도와 영향 요인 간의 인과관계를 밝히기 위한 시도가 필요하다. 다음으로, 본 연구에서는 항만 영향권을 교통영향평가의 항만 배후도로 기준 등을 참고하여 설정하였으나, 반경 2km의 경우 항만 인근 도로가 충분히 포함되지 않았으며, 특정 항만의 경우 항만 인근 도로에 도시부 도로가 포함되지 않을 수 있어 적절한 영향권 크기와 경계를 찾는 연구가 수행되어야 한다. 마지막으로 본 연구에서는 항만 인근 영향권 지역(2~8km)과 도시 수준(City level)을 구분하여 분석을 수행하였으나, 더욱 정확한 비교 분석을 위해 항만이 존재하지 않는 해안 도시, 화물차의 유출입이 많지만 해안에 인접해 있지 않은 도시 등 다른 지역에서의 사고 심각도 영향요인 분석 결과와 비교할 필요가 있다.

참고문헌

- Abbasi, E., Li, Y., Wu, X., Craig, B.(2021), Using Classification and Regression Trees (CART) to Identify Factors Contributing to Vehicle Crash Severity in a Port City, *International Journal of Transportation Systems*, 6(2021), 29-38
- Baak, M., Koopman, R., Snoek, H., Klous, S.(2020), A new correlation coefficient between categorical, ordinal and interval variables with Pearson characteristics, *Computational Statistics & Data Analysis*, 152(2020), 107043
- Barman, S., Bandyopadhyaya, R.(2023), Analyzing Urban Midblock Crash Severity Outcomes Using Proposed Three-Step Pattern Clustering, *Journal of Transportation Engineering, Part A: Systems*, 149(7), 05023003
- Basso, F., Pezoa, R., Varas, M., Villalobos, M.(2021), A deep learning approach for real-time crash prediction using vehicle-by-vehicle data, *Accident Analysis & Prevention*, 162(2021), 106409
- Besag, J., Green, P., Higdon, D., Mengersen, K.(1995), Bayesian computation and stochastic systems, *Statistical science*, 10(1), 3-66
- Boser, B. E., Guyon, I. M., Vapnik, V. N.(1992), A training algorithm for optimal margin classifiers, In *Proceedings of the fifth annual workshop on Computational learning theory*, 144-152
- Budiyanto, M. A., Fernanda, H.(2020), Risk assessment of work accident in container terminals using the fault tree analysis method, *Journal of Marine Science and Engineering*, 8(6), 466
- Claros, B., Chitturi, M., Vorhes, G., Bill, A., Noyce, D.(2020), Seasonal crash prediction model for urban signalized intersections: Wisconsin southeast region, *Traffic injury prevention*, 21(7), 447-452
- Fabiano, B., Currò, F., Reverberi, A. P., Pastorino, R.(2010), Port safety and the container revolution: A statistical study on human factor and occupational accidents over the long period, *Safety science*, 48(8), 980-990
- Gong, L., Fan, W. D.(2017), Modeling single-vehicle run-off-road crash severity in rural areas: Accounting for unobserved heterogeneity and age difference, *Accident Analysis & Prevention*, 101(2017), 124-134
- Jung, H. Y.(1999), A Study on the Influence of Heavy Vehicles at the Port Road, *KSCE Journal of Civil and Environmental Engineering Research*, 19(3-2), 225-234
- Lee, H., Ha, M.(2021), Identification of Risk Factors and Evaluation of the Importance of Risk Factors in Container Ports, In *Korean Logistics Research Association, Korea Logistics Research Association*, 31(6), 23-33
- Lee, H., Pham, H. T., Kim, C., Lee, K.(2019), A study on emissions from drayage trucks in the port city-focusing on the port of incheon, *Sustainability*, 11(19), 5358
- Lei, T., Peng, J., Liu, X., Luo, Q.(2021), Crash Prediction on Expressway Incorporating Traffic Flow Continuity Parameters Based on Machine Learning Approach, *Journal of Advanced Transportation*, 2021, 1-13
- Li, Y., L. Song, Fan, W. D.(2021), Day-of-the-week variations and temporal instability of factors

- influencing pedestrian injury severity in pedestrian-vehicle crashes: A random parameters logit approach with heterogeneity in means and variances, *Analytic methods in accident research*, 29(2021), 100152
- Lin, L., Wang, Q., Sadek, A. W.(2015), A novel variable selection method based on frequent pattern tree for real-time traffic accident risk prediction, *Transportation Research Part C: Emerging Technologies*, 55(2015), 444-459
- Myung-Shin Lee, Yong-Jin Kim.(2009), A study on Freight-only-lane and Freight-only-highway Policies Implemented in the Incheon City, *Journal of Korea Port Economic Association*, 25(3), 117-138
- Park, H. C., D. K. Kim, S. Y. Kho, P. Y. Park.(2017), Cross-classified multilevel models for severity of commercial motor vehicle crashes considering heterogeneity among companies and regions, *Accident Analysis & Prevention*, 106(2017), 305-314
- Rovai, A. P., Baker, J. D., Ponton, M. K.(2013), *Social science research design and statistics: A practitioner's guide to research methods and IBM SPSS*, Watertree Press LLC
- Shi, Q., Abdel-Aty, M.(2015), Big data applications in real-time traffic operation and safety monitoring and improvement on urban expressways, *Transportation Research Part C: Emerging Technologies*, 58, 380-394
- Traffic Accident Analysis System (TAAS).(2023), <https://taas.koroad.or.kr/>
- Ukil, A.(2007), Support vector machine, In *Intelligent Systems and Signal Processing in Power Engineering*, 161-226
- Wang, L., Abdel-Aty, M., Shi, Q., Park, J.(2015), Real-time crash prediction for expressway weaving segments, *Transportation Research Part C: Emerging Technologies*, 61(2015), 1-10
- Xiong, Y., Tobias, J. L., Mannering. F. L.(2014), The analysis of vehicle crash injury-severity data: A Markov switching approach with road-segment heterogeneity, *Transportation research part B: methodological*, 67(2014), 109-128
- Xu, P., Huang H.(2015), Modeling crash spatial heterogeneity: Random parameter versus geographically weighting, *Accident Analysis & Prevention*, 75(2015), 16-25
- You, J., Wang, J., Guo, J.(2017), Real-time crash prediction on freeways using data mining and emerging techniques, *Journal of modern transportation*, 25(2), 116-123
- Ziemska, M.(2021), Exhaust emissions and fuel consumption analysis on the example of an increasing number of hgvs in the port city, *Sustainability*, 13(13), 7428

요약문

항만과 항만 인근 항만의 영향을 받는 도로는 대형 사고를 초래할 수 있는 화물차의 이동이 많기 때문에 교통안전에 각별한 주의가 필요하다. 따라서 사고 심각도를 낮출 수 있는 안전 관리 대책을 마련하기 위해 항만 도시의 도로 구간을 항만의 영향을 받는 정도에 따라 구분하고, 각 항만 영향권에서 사고 심각도에 영향을 미치는 요인을 도출해 안전 관리 전략을 수립할 필요가 있다. 본 연구에서는 항만 영향권에서 사고 심각도에 영향을 미치는 요인을 네 가지 머신러닝 기법을 통해 도출하고자 하였다. 모형 개발 후에는 가장 예측성능이 뛰어난 모형에 대하여 설명가능한 인공지능 기법을 통해 높은 사고 심각도에 영향을 미치는 요인을 도출하였다. 본 연구에서 도출된 결과를 활용하여 항만 지역의 사고 심각도 감소를 위한 정책 수립의 기초자료로 활용될 수 있을 것으로 기대된다.

주제어: 항만 안전, 사고 심각도 모형, 기상 데이터, 교통안전, 머신러닝