

자율주행 차량 사고 패턴 분석 및 차량 손상 예측 모형 구축[†]

심기영¹ · 이인규² · 양정아³ · 연구필⁴

^{1,2,3}호서대학교 데이터사이언스학과 · ⁴호서대학교 빅데이터AI학부
접수 2024년 10월 8일, 수정 2024년 11월 12일, 게재확정 2024년 11월 20일

요약

스마트 시대가 도래함에 따라 자율주행 차량 기술이 비약적으로 발전되어 가고 있다. 그러나 자율주행 기술의 고도화에도 불구하고 자율주행 차량의 사고 발생도 지속적으로 나타나고 있다. 자율주행 차량 사고는 기술적 결함이나 교통 환경에 대한 낮은 대처 또는 상대 차량의 가해로 인한 피할 수 없는 상황 등 여러 가지 이유로 발생할 수 있다. 자율주행 차량이 상용화됨에 따라 늘어나고 있는 자율주행 사고를 예방하고 피해를 최소화하기 위해 사고유형과 피해정도 등을 면밀히 분석할 필요가 있다. 이에 본 연구에서는 자율주행 차량 사고 시의 사고 유형을 살펴보고 여러 가지 사고와 관련된 환경변수가 피해 정도에 미치는 영향을 혼합효과 로지스틱회귀모형을 통해 분석하였다. 분석에 활용된 자료는 미국 캘리포니아 교통관리국에서 제공한 자율주행 차량 충돌 보고서이다. 주요 분석 결과로는, 자율주행 차량 사고시 피해(차량손상) 정도가 더 크게 나타나게 되는 환경은 수동모드 운행, 차대차 사고, 어두운 환경, 부상자 발생, 자율주행차가 가해차량인 경우 등의 요인으로 파악되었다. 본 연구 결과는 자율주행 차량의 사고를 줄이고 사고 시 피해를 최소화할 수 있는 방향으로의 기술개발 및 교통 환경 개선을 유도하는데 활용될 수 있을 것이다.

주요용어: 자율주행, 자율주행차 사고, 캘리포니아 교통관리국, 혼합효과 로지스틱회귀모형.

1. 서론

스마트 시대가 도래함에 따라 미래 모빌리티는 자율주행 기술, 빅데이터, IoT, 전기차 기술 등의 발전으로 인해 우리의 삶에 큰 변화가 다가올 것으로 생각된다. 미국 도로교통안전국 (national highway traffic safety administration, NHTSA)에서 교통 자동화는 안전, 교통체증, 이동성 등의 문제에서 근본적인 해결책으로 다가올 수 있다고 하였다. 특히 인적 오류를 줄일 수 있다는 점에서 사고를 획기적으로 줄일 것으로 예상된다.

국제자동차기술자협회 (society of automotive engineers, SAE)에서는 자율주행 단계를 기술 수준에 따라 6단계로 구분하고 있다 (Table 1.1). 현재 우리나라에서는 차간거리 제어, 차선유지 지원 등 특정한 자동화 시스템이 적용 (레벨1)되어 있으며, 2개 이상의 자동화 시스템이 통합되는 자동차 (레벨2) 역시 많은 자율주행차 업체들이 양산 중에 있고, 제한된 조건에서 눈을 땔 수 있는 레벨3 자율주행차는 테스트가 활발히 진행되고 있다.

[†] 이 논문은 정부(과학기술정보통신부)의 재원으로 한국연구재단의 지원을 받아 수행된 연구임(No. 2020R1F1A1A01073456). 본 과제(결과물)는 2024년도 교육부의 재원으로 한국연구재단의 지원을 받아 수행된 지자체-대학 협력기반 지역혁신 사업의 결과임(2021RIS-004).

¹ (31499) 충청남도 아산시 배방읍 호서로 79번길 20, 호서대학교 일반대학원 데이터사이언스학과, 석사과정.

² (31499) 충청남도 아산시 배방읍 호서로 79번길 20, 호서대학교 일반대학원 데이터사이언스학과, 석사과정.

³ (31499) 충청남도 아산시 배방읍 호서로 79번길 20, 호서대학교 일반대학원 데이터사이언스학과, 석사과정.

⁴ 교신저자:(31499) 충청남도 아산시 배방읍 호서로 79번길 20, 호서대학교 빅데이터AI학부, 교수.

E-mail: kpyeon1@hoseo.edu

Table 1.1 Definitions for six levels of driving automation

Level 0	Level 1	Level 2	Level 3	Level 4	Level 5
No driving automation	Driver assistance	Partial driving automation	Conditional driving automation	High driving automation	Full driving automation

자율주행 차량의 발전은 교통안전을 크게 향상시킬 수 있는 좋은 기회임에도 불구하고 여전히 혼잡한 교통 환경 속에서 시스템 오류나 감지 성능의 문제로 인한 사고의 위험성도 상존한다 (Khattak 등, 2019). 따라서 안전에 관한 기술적 안정성 및 정확성에 대한 연구가 자율주행 산업 연구에서 유망한 연구 분야로 대두되고 있고 (Jin 등, 2023), 이와 더불어 자율주행 차량 제어권 전환에 관한 연구 (Ko와 Lee, 2023) 및 교통안전을 위한 교통심리학에 대한 연구도 활발히 진행되고 있다 (Chung과 Choi, 2024).

2023년 말 경찰청에서 발표한 「완전 자율주행 시대에 대비한 도로교통안전 추진전략」에 따르면, 경찰청은 도로교통안전 주요 과제를 ‘평가검증체계, 운행안전관리, 기반 조성’의 3개 분야로 구분하여 체계적으로 추진하고자 하는 계획을 담고 있다. 특히, 자율주행 상용화 시기를 3단계로 구분하여 각 단계에 필요한 총 28개 추진과제를 제시함으로써, 자율주행 차량 사고 시 사고의 주체를 명확히 하고 사고 원인 조사를 위한 자율주행 차량 운행 데이터 활용 근거를 마련하는 등 자율주행 상용화에 대비하고 있다.

본 연구에서는 자율주행 차량 사고 데이터를 분석하여 사고 유형과 특징을 파악하고 특정 사고 상황과 환경이 ‘차량 손상 정도’에 미치는 영향을 탐색하고 해석하고자 하였다. 탐색적 자료 분석 및 예측모형 구축을 통해 주요 설명 변수들이 자율주행 차량 사고 시 차량 손상 정도에 미치는 영향력을 알아보았다. 논문의 구성은 다음과 같다. 다음 장에서는 자율주행 차량 사고 분석과 관련한 선행연구를 살펴보고, 3장에서는 연구에 사용한 데이터 수집 및 탐색적 자료분석 결과를 보여주며, 4장에서는 혼합효과로 지스틱회귀모형 결과와 해석을 제시한다. 끝으로 5장에서 결론을 서술한다.

2. 선행연구

캘리포니아 교통 관리국 (California department of motor vehicles, CA DMV)의 자율주행 차량 충돌 보고서를 이용한 기존 연구들은 다음과 같다. Das 등 (2020)은 베이지안 클러스터링을 통해 자율주행 차량의 사고 유형 및 환경을 6개의 특징적인 클러스터로 분류하였다. Boggs 등 (2020)은 자율주행 시스템이 해제되는 주체를 언제, 어디서, 무엇이, 왜의 4가지 상황을 로지스틱회귀모형으로 비교하여 시스템이 해제되는 것에 어떠한 영향을 미치는지 알아봄으로써 자율주행 시스템이 다양한 요인에 의해 해제될 수 있다는 가능성을 제시하였다. Kim 등 (2024)은 다양한 자율주행 차량 사고 관련 기존 문헌, 보고서, 차량 영상, 국제 표준 등을 연구하여 자율주행 차량 사고에 대한 조사 항목을 도출하여 보다 정확한 사고 인과관계를 파악할 수 있는 발판을 마련하였다. Wang 등 (2024)은 24개의 자율주행 차량 충돌 전 시나리오를 자율주행 차량의 수동/자동/전환 3개의 모드별 운전 패턴 차이를 비교·분석하고 시나리오별 차량 손상 정도, 인명피해, 충돌의 특성에 미치는 영향분석 및 위험 수준을 계산함으로써 복잡한 환경에서 차량 제어권을 자동에서 수동으로 전환하면 사고의 위험도를 줄일 수 있다고 하였다. 또한 Kuo 등 (2024)은 자율주행 차량 충돌 심각도 분석에 있어 CA DMV 자율주행 차량 충돌 보고서의 문제 중 하나인 데이터 불균형 문제를 해결하려고 시도하여, 보고서를 그냥 사용하는 것보다 오버샘플링을 적용하여 분석을 진행하는 것이 모델 성능향상에 도움을 주었다고 하였다. Zhou 등 (2024)은 충돌 사진, 영상, 물리적 증거, 목격자 진술, 전문가 조사 등을 활용하여 자율주행 차량 사고를 반사실적 시뮬레이션으로 재구성한 후 분석하여 사고의 시나리오를 불가피한 사고 또는 충돌을 피한 사고 2가지 상황에 따라 사람 운전자와 자동 운전 간 안전 성능을 평가하여 자율주행 차량이 사람 운전자에 비해 주행 속도 등에서 더 높은 안정성을 보임을 확인하였다. Lee 등 (2023)은 자율주행 차량 사고 발생 시 다양한 상

황에서 베이지안 접근 방식을 이용하여 충돌 전 조건, 자율주행 차량의 운전 모드, 충돌 유형 및 충돌 결과 간의 상호 관계를 조사하여 자율주행 차량의 기술적 보완점 등을 제시하였다. Leilabadi와 Schmidt (2019)는 사고 유형과 심각도 간 영향을 보기 위해 의사결정나무와 베이지안 네트워크 기법을 통해 사고의 주요 원인을 제시하였다. Kutela 등 (2021)은 취약 도로 이용자 (vulnerable road users, VRU)를 중심으로 자율주행 차량 충돌 패턴을 분석하고 VRU 사고와 non-VRU 사고를 비교하여 자율주행 차량과 VRU 관련 충돌의 주요 변수를 알아내었다. Chen 등 (2020)은 사고 지점 (point of interest)을 고려한 자율주행 차량 사고 심각도를 예측하는 XGBoost 모형을 개발함으로써 사고 지점에 따른 차량 충돌 심각도를 제시하였다.

3. 데이터 수집 및 탐색적 자료분석

3.1. 데이터 수집 및 전처리

자율주행 차량 사고 발생 시 차량 손상에 미치는 영향을 분석하기 위해 CA DMV에서 제공하는 자율주행 차량 충돌 보고서를 활용하였다. CA DMV 자율주행 차량 충돌 보고서는 자율주행 차량을 테스트하는 제조업체가 자사 차량과 관련된 모든 충돌 사고를 보고하며, 이로 인해 재산 피해, 신체 상해 또는 사망이 발생하였을 때 보고된다. 2014년부터 2024년 6월까지 총 722건의 사고 신고를 접수하여 충돌 보고서가 만들어졌는데, 그 중 2018년부터 2024년까지 총 7년간의 데이터 총 637건의 보고서를 수집하였다. PDF 형식으로 되어있는 보고서에서 관련 수치를 수작업으로 발췌하여 데이터 파일을 만들고 분석을 진행하였다. 수집한 데이터는 ‘자율주행 차량 업체’, ‘사고 발생 날짜’, ‘사고 발생 시간’, ‘자율주행 차량 제조 년도’, ‘자율주행 차량 제조업체’, ‘자율주행 차량 모델’, ‘사고 주소’, ‘사고 도시’, ‘자치 주’, ‘사고관계자’, ‘관련 차량 수’, ‘자율주행 차량 손상 정도’, ‘자율주행 차량 손상 부위’, ‘상대 차량 제조 년도’, ‘상대 차량 모델’, ‘주행 모드’, ‘사고 세부 기록’, 사고 발생 시에 ‘날씨’, ‘조명’, ‘도로 표면’, ‘도로 상태’, 차량1 (autonomous vehicle, AV)과 차량2 (other vehicle, OV)의 ‘충돌 직전 움직임’, 차량 1과 차량2의 ‘충돌 유형’, ‘기타 사고가 발생한 다른 관련 요인’ 등이다.

분석에 활용하기 위해 보고서에 있는 내용의 자료값들을 코딩하여 전처리 후 재범주화 하였다. 결측값이 많거나 사고 세부 기록을 보고 사고 원인 차량을 파악하기 어려운 8개의 데이터를 분석에 부적합하다고 판단하여 제외하였다. 수집된 데이터 중 10개의 변수만을 분석에 활용하였으며, 종속변수로 ‘차량 손상정도’를 기준에 없음, 경미, 보통, 중대 4개의 범주에서 없음과 경미를 합하여 ‘경미함’으로 만들었으며, 보통과 중대를 합쳐 ‘중대함’으로 만들어 2개의 범주로 재범주화를 진행하였다. ‘자율주행모드 여부’는 보고서의 Section 5를 확인하여 자율주행모드 및 수동주행모드 여부를 수집하였다. ‘사고 관계자’는 차량, 보행자, 자전거 이용자, 오토바이, 사물 (도로 연석, 타인이 던진 물건) 등이 수집되었는데, ‘차량vs차량 (오토바이)’, ‘차량vs기타 (보행자, 자전거, 사물)’ 2가지로 재범주화 하였다. ‘자율주행 차량 제조업체’는 빈도수를 기준으로 범주화 하였는데, ‘chevrolet’, ‘chrysler’, ‘cruise’, ‘jaguar’, ‘toyota’를 범주화하고, 나머지 빈도가 적은 것은 ‘others’로 구분하였다. ‘조명’은 ‘밝음’, ‘어두움’으로, ‘도로 표면’은 ‘건조함’과 ‘미끄러움’으로, ‘차량1과 차량2의 충돌 직전 움직임’은 ‘자율주행 차량의 충돌 직전의 움직임’을 분석에 사용하였다. 기존 18가지의 범주에서 정지한 상태이면 ‘정지’, 주행 중인 상황이라고 판단되는 직진 주행, 도로를 벗어남, 다른 차량 추월, 잘못된 길로 주행을 ‘주행’으로, 주차나 후진하는 상황을 저속 주행하는 상태로 확인하여 주차, 후진, 감속/정지, 주차 조작, 교통 진입 등을 ‘주차 및 후진’으로, 차선변경이나 방향 전환하는 상황인 우회전, 좌회전, 유턴, 기타 안전하지 않은 방향 전환, 차선변경, 반대 차선으로 끼어들기, 병합 등을 ‘차선 및 방향 전환’으로 범주화하였고, ‘차량1과 차량2의 충돌 유형’은 ‘상대 차량의 충돌 유형’을 분석에 사용하였다. 기존 8가지 범주 (정면충돌, 측면을 쓸면서 충돌, 후방 충돌, 측면 충돌, 사물과 충돌, 차량 전복, 차량 및 보행자 사고, 기타 충돌)를 ‘차량 간 충돌이

아닌 경우’, ‘진방 충돌’, ‘측면 스칩’, ‘후방 충돌’, ‘측면 충돌’로 재범주화 하였다. 또한, Figure 3.1과 같이 보고서 내에 서술된 ‘사고 세부 기록’을 활용하여 ‘사고 원인 차량’과 ‘부상자 여부’라는 변수를 만들었다. ‘사고 원인’은 사고 세부 기록을 번역하고 내용을 살펴 만들어졌으며, 상대방 (차량, 자전거, 보행자)이 가해한 경우와 자율주행 차량이 가해한 경우를 연구자가 사고 손상 경위를 확인하고 분류를 진행하였다. 부상자 여부 또한 사고 세부 기록을 보고 부상자가 보고된 내용이 있는 경우 ‘부상자 있음’, 없는 경우엔 ‘부상자 없음’, 부상자 내용이 기록되지 않았을 때는 ‘부상자 모름’으로 기재하였다. 이러한 전처리 과정을 통해 629건의 데이터가 만들어졌고 분석에 활용되었다.

Table 3.1 Pre-processing of variables

Original variable	Original attributes	Derived variable	Derived Attributes
Damage level	None, Minor, Mod, Major	Damage level	Minor(0), Major(1)
Description	-	Accident cause	Others(0), Autonomous vehicle(1)
Description	-	Injury	Not reported(0), Reported(1), Unknown(2)
Driving mode	Autonomous mode, Conventional mode	Driving mode	Autonomous mode(0), Conventional mode(1)
Lighting	Daylight, Dust-Dawn, Dark-Streetlight, Dark-No streetlight, Dark-Streetlight not functioning	Lighting	Daylight(0), Darkness(1)
Involved in the accident	Vehicle, Pedestrian, Object	Involved in the accident	Vehicle(0), Pedestrian & Object(1)
Vehicle 1,2 movement proceeding collision	Stopped, Proceeding straight, Ran off road, Making right turn, Making left turn, Making U turn, Backing, Slowing/Stopping, Passing other vehicle, Changing lanes, Parking maneuver, Entering traffic, other unsafe turning, xing into opposing lane, Parked, Merging, Traveling wrong way, Other	AV_MPC	Stopped(0), Driving(1), Parked and backing(2), Changing lanes and direction(3)
Make	Chevrolet, Chrysler, Cruise, Ford, Hyundai, Jaguar, Lexus, Lincoln, Nissan, Mercedes-Benz, Toyota	AV_Maker	Chevrolet(0), Chrysler(1), Cruise(2), Jaguar(3), Toyota(4), Others(5)
Vehicle 1,2 type of collision	Head-on, Side swipe, Rear end, Broadside, Hit object, Overturned, Vehicle/Pedestrian, Other	OV_Type of collision	Not between vehicles(0), Head-on(1), Side swipe(2), Rear end(3), Broadside(4)
Roadway surface	Dry, Wet, Snowy/Icy, Slippery	Roadway surface	Dry(0), Slippery(1)

SECTION 5 — ACCIDENT DETAILS - DESCRIPTION

Autonomous Mode Conventional Mode

A Zoox vehicle (Vehicle 1) in autonomy was traveling westbound on Chess Drive in Foster City when a parallel-parked passenger car (Vehicle 2) abruptly entered the Zoox vehicle's travel lane from a parking space in the right parking lane without yielding or activating a turn signal. Vehicle 2 made contact with the right front side and wheel of the Zoox vehicle, and the Zoox vehicle made contact with the driver side door and left front tire/wheel well of the passenger car. Both vehicles sustained moderate damage and the Zoox vehicle was towed from the scene. There were no injuries and police were not called.

Figure 3.1 An example of accident details for the new variables 'Accident cause' & 'Injury'

3.2. 탐색적 자료분석

전처리된 629건의 데이터를 바탕으로 탐색적 자료 분석 및 모델링을 수행하였다. 기본적인 탐색적 분석 결과는 Table 3.2와 Table 3.3에 제시하였고, 이에 따른 주요 시사점을 간단히 요약하면 다음과 같다.

Table 3.2 Number of accidents according to damage level and driving mode

Driving mode	Damage level	
	Major	Minor
Autonomous mode	69	337
Conventional mode	40	183

Table 3.3 Exploratory data analysis

		Autonomous mode		Conventional mode	
		Major(%)	Minor(%)	Major(%)	Minor(%)
Lighting	Darkness	31(22)	109(78)	15(28)	38(72)
	Daylight	38(14)	228(86)	25(15)	145(85)
Involved in the accident	Vehicle	64(18)	295(82)	35(19)	150(81)
	Other	5(11)	42(89)	5(13)	33(87)
Roadway surface	Dry	62(17)	304(83)	38(19)	161(81)
	Slippery	7(18)	33(82)	8(27)	22(73)
Injury	No	38(17)	192(83)	23(17)	115(83)
	Yes	9(18)	42(82)	8(36)	14(64)
	Unknown	22(18)	103(82)	9(14)	54(86)
Accident cause	Others	61(17)	294(83)	28(19)	116(81)
	Autonomous vehicle	8(16)	43(84)	12(15)	67(85)
Vehicle 1,2 movement proceeding collision	Stopped	25(13)	163(87)	13(19)	56(81)
	Driving	26(21)	97(79)	17(30)	39(70)
	Parked and backing	5(11)	41(89)	6(10)	56(90)
	Changing lanes and turning	13(27)	36(73)	5(14)	32(86)
OV_Type of collision	Not between vehicles	5(11)	42(89)	5(13)	33(87)
	Head on	45(19)	190(81)	15(19)	65(81)
	Side swipe	12(15)	66(85)	11(18)	49(82)
	Rear end	4(12)	30(88)	1(3)	30(97)
	Broadside	3(25)	9(75)	8(57)	6(43)

- 전체 데이터 중에서 반응변수인 '차량손상정도'의 범주별 구성은 '경미함 (minor)'이 520건, '중대함 (major)'이 109건이다.
- 주행모드 (자율/수동)에 따른 중대사고 사고 발생비율은 큰 차이가 없다. - 수동모드일 때 18%, 자율주행모드일 때 17%.
- 밝을 때에 비해 어두울 때 중대한 사고 비율이 높게 나타나며, 어두운 경우의 중대사고 비율은 자율주행모드 보다 수동모드일 때가 더 높다.
- '차대 차' 사고인 경우가 '차대 기타' 사고인 경우보다 중대사고 비율이 높으며, 이는 주행모드와는 큰 상관성이 없다.
- 도로표면이 미끄러울 때는 '수동모드'일 때보다 '자율주행모드'일 때 중대한 사고 비율이 더 높다. 다만, 자율주행모드인 경우 도로표면의 건조 여부는 중대한 사고 비율에 차이를 보이지 않는다.

- ‘수동모드’인 경우 부상자가 보고된 사고일 때 중대한 사고 비율이 매우 높다. 반면, ‘자율주행모드’인 경우는 ‘부상자 여부’에 따른 중대한 사고의 비율 차이는 거의 없다.
- 가해차량이 상대방인 경우에 중대사고 비율이 근소하게 높지만, 자율주행 차량의 주행모드에 따른 차이는 거의 없다.
- ‘자율주행 차량 충돌 직전 움직임’과 관련해서는, ‘수동모드’에서는 ‘자율주행 차량 충돌 직전 움직임’이 ‘주행’인 경우 중대한 사고 비율이 가장 높으며, ‘자율주행모드’에서는 ‘차선 및 방향 전환’인 경우가 가장 높게 나타났다. 반면, ‘자율주행 차량 충돌 직전 움직임’이 ‘주차 및 후진’인 경우는 ‘주행모드’에 관계없이 중대한 사고의 비율이 가장 낮게 나타났다.
- ‘상대 차량의 충돌 유형’ 관련해서는, ‘주행모드’에 관계없이 ‘측면 충돌’일 때 중대한 사고의 비율이 가장 높게 나타났다. ‘수동모드’에서는 ‘상대 차량의 충돌 유형’이 ‘후방 충돌’일 때 중대한 사고 비율이 가장 낮게 나타났으며, ‘자율주행모드’에서는 ‘차량 간 충돌이 아닌 경우’일 때 중대한 사고의 비율이 가장 낮게 나타났다.

4. 모델링

로지스틱회귀모형, 혼합효과 로지스틱회귀모형, 배깅, 랜덤포레스트 및 부스팅 방법을 통해 자율주행 차량 사고 시 차량 손상 정도 (경미/중대)를 예측하는 이진 분류모형을 구축하여 비교하였다. 앙상블 모형에서는 모두 의사결정나무를 100개 사용하였다. 총 629건의 데이터를 반응변수 기준의 층화추출로 6:4 비율로 랜덤하게 나누어 378건의 훈련용 데이터와 251건의 테스트 데이터로 구성하였다. 이러한 작업을 100번 반복하여 모형별 테스트 데이터에 대한 예측 성능 지표들의 평균과 표준편차를 구하여 Table 4.1에 제시하였다.

정확도와 특이도는 랜덤포레스트가 각각 0.701, 0.766로 가장 높고, 민감도는 배깅이 0.599로 가장 높으며, 정밀도는 혼합효과 로지스틱모형이 0.261로서 가장 높게 나왔고, 민감도와 정밀도의 조화평균인 F1-스코어는 배깅이 0.358로 가장 높게 나타났다. ROC 곡선의 아래 면적을 나타내는 AUC 값은 배깅이 0.622로 가장 높고 혼합효과 로지스틱회귀모형이 0.610으로 그 다음으로 높게 나타났다. 본 연구에서는 예측모형의 해석가능성 및 차량 손상이 중대한 경우를 잘 예측하는 것을 더 중요하게 생각하여 민감도와 정밀도가 상대적으로 높게 나온 혼합효과 로지스틱회귀모형을 대상으로 구축된 모형의 결과와 해석을 진행하였다.

Table 4.1 The mean and standard deviation of performance metrics over 100 iterations

Prediction models	Accuracy	Sensitivity	Specificity	Precision	F1-score	AUC
Logistic regression	0.652 (0.034)	0.539 (0.093)	0.675 (0.042)	0.254 (0.035)	0.344 (0.048)	0.607 (0.043)
Mixed effect logistic regression	0.663 (0.034)	0.531 (0.084)	0.690 (0.042)	0.261 (0.035)	0.349 (0.046)	0.610 (0.042)
Bagging	0.637 (0.045)	0.599 (0.122)	0.644 (0.060)	0.257 (0.040)	0.358 (0.057)	0.622 (0.055)
Random forest	0.701 (0.032)	0.388 (0.097)	0.766 (0.042)	0.253 (0.048)	0.304 (0.062)	0.577 (0.044)
AdaBoost	0.693 (0.029)	0.364 (0.095)	0.761 (0.038)	0.236 (0.048)	0.285 (0.061)	0.562 (0.044)

4.1. 혼합효과 로지스틱회귀모형

혼합효과 로지스틱회귀모형은 군집화된 자료나 반복 측정된 자료에서 군집간 이질성이나 군집내 자료들의 상관성을 고려할 수 있는 로지스틱회귀모형이다. 기존의 로지스틱회귀모형의 입력변수는 고정효과 변수들이고 군집의 랜덤효과를 추가적으로 모형에 산입한다. 흔히 군집 간 차이를 랜덤 절편모형 (random intercept model)으로 설명하거나 고정효과로 들어간 특정 설명변수의 기울기가 군집 간에 차이를 나타내는 랜덤효과로 하여 랜덤 기울기모형 (random slope model)을 구축한다.

이진 범주 (즉, 0과 1)를 갖는 반응변수를 $(Y_i)_{i=1, \dots, N}$ 라고 하자. 고정효과와 랜덤효과를 포함한 선형 예측자(linear predictor) $\eta_i = \mathbf{x}_i^T \boldsymbol{\beta} + \mathbf{z}_i^T \mathbf{u}$ 가 주어졌을 때 $Y_i = 1$ 일 확률이 다음과 같은 관계가 있다고 가정하자.

$$P(Y_i = 1 | \eta_i) = \frac{\exp(\eta_i)}{1 + \exp(\eta_i)}, \quad \text{단 } \eta_i = \mathbf{x}_i^T \boldsymbol{\beta} + \mathbf{z}_i^T \mathbf{u}, \quad (4.1)$$

여기서, \mathbf{x}_i 는 i 번째 관측치의 관찰값인 p 차원 벡터이고, $\boldsymbol{\beta}$ 는 고정효과를 나타내는 p 차원 모수로서 추정 대상이다. \mathbf{u} 는 랜덤효과를 나타내는 r 차원 확률벡터로서 평균이 $\mathbf{0}$ 이고 공분산행렬 Σ 를 갖는 다변량 정규분포를 따른다. \mathbf{z}_i 는 랜덤효과에 대응되는 설명변수 값을 나타내는 r 차원 벡터이다. 흔히 절편만 있는 경우에는 군집간 절편이 랜덤인 랜덤절편 모형이 되며, 특정 설명변수를 포함하는 경우 랜덤 기울기 모형으로서 해당 설명변수의 기울기도 군집에 따라 변동이 있음을 설명하는 모형이 된다. 한편, 각 관측치의 확률 예측값은 랜덤효과와 조건부 기댓값을 반영하여 추정된 회귀계수를 사용한다.

본 연구에서는 ‘자율주행 차량 충돌 직전 움직임’과 ‘상대 차량의 충돌 유형’을 군집으로 간주하여, 랜덤효과로서 군집간 랜덤절편 및 이에 따른 ‘도로 표면’과 ‘사고 원인 차량’의 랜덤기울기를 포함한 랜덤절편-랜덤기울기 모형으로 분석을 진행하였다. ‘도로 표면’과 ‘사고 원인 차량’은 범주형 변수로서 0 또는 1의 값을 갖는 이진 변수들이지만 연속형으로 취급하여 랜덤효과를 구성하였다. 흔히 혼합효과모형은 랜덤효과와 유의성을 먼저 살펴보고, 유의할 경우 각 고정효과 변수들의 유의성을 탐색하면서 AIC 또는 BIC 기준의 최종모형을 선택한다. 여기서는 AIC 기준으로 선택된 최종 모형 결과만을 Table 4.2에 제시하며 고정효과 변수 중에서 유의하지 않은 변수도 해석 및 비교분석의 편리를 위해 포함시켰다. 모든 모형구축 과정은 R 패키지 lme4에 있는 glmer() 함수를 이용하여 진행하였다.

4.2. 혼합효과 로지스틱회귀모형 결과 해석

혼합효과 로지스틱회귀모형을 통해 ‘자율주행 차량 손상 정도’에 미치는 영향을 분석한 결과를 Table 4.2에 제시하였다. 먼저 유의한 영향력을 가지는 고정효과를 살펴보면, ‘주행모드’는 ‘자율주행모드’일 때 ‘자율주행 차량 손상 정도’가 증대할 확률이 낮아지는 것으로 나타났으며 ($\beta = -0.588, p = 0.030$), ‘조명’은 어두울수록 ‘자율주행 차량 손상 정도’가 증대할 확률이 높아지는 것으로 나타났다 ($\beta = 0.785, p = 0.001$).

‘부상자 여부’는 부상자가 보고되지 않는 경우보다 부상자가 보고된 경우가 ‘자율주행 차량 손상 정도’가 증대할 확률이 높아지는 것으로 나타났다 ($\beta = 1.217, p < .001$). 또한 ‘부상자 여부’를 모르는 경우도 ‘자율주행 차량 손상 정도’가 증대할 확률이 높아지는 것으로 나타났다 ($\beta = 0.807, p = 0.029$). ‘자율주행 차량 제조사’는 ‘Chevrolet’일 때에 비해 ‘Chrysler’인 경우 ‘자율주행 차량 손상 정도’가 증대할 확률이 낮아지는 것으로 나타났다 ($\beta = -1.917, p = 0.001$). 반면, ‘Cruise’인 경우 ‘자율주행 차량 손상 정도’가 증대할 확률이 높아지는 것으로 나타났다 ($\beta = 0.833, p = 0.028$). 다음으로 랜덤효과를 살펴보면 ‘자율주행차량 충돌 직전 움직임’과 ‘상대 차량의 충돌 유형’의 교호작용은 랜덤절편이고, 분산이 20.472 표준편차가 4.525로 고정효과와 절편과 비교했을 때 각 군집 별로 절편의 변동성이 큰 것을 볼 수 있다. ‘도로 표면’과 ‘사고 원인’은 랜덤기울기이고, 각각 분산이 8.322, 8.386 표준편차가 2.885,

Table 4.2 Mixed effect logistic regression

Category	Variables	β	SE	z statistics	p-values
Fixed effect					
	(Intercept)	-1.4751	2.4908	-0.592	0.554
Driving mode	Conventional	-	-	-	-
	Autonomous	-0.588	0.271	-2.168	0.030*
Lighting	Daylight	-	-	-	-
	Darkness	0.785	0.241	3.254	0.001**
Involved in the accident	Vehicle	-	-	-	-
	Pedestrian, Object	-1.253	1.129	-1.110	0.267
Roadway surface	Dry	-	-	-	-
	Slippery	-1.186	1.416	-0.838	0.402
Injury	Not reported	-	-	-	-
	Reported	1.217	0.343	3.551	<.001***
	Unknown	0.807	0.368	2.189	0.029*
Accident cause	Others	-	-	-	-
	Autonomous vehicle	0.994	1.029	0.966	0.334
AV_Maker	Chevrolet	-	-	-	-
	Chrysler	-1.917	0.574	-3.338	0.001**
	Cruise	0.833	0.380	2.192	0.028*
	Jaguar	-0.484	0.430	-1.125	0.261
	Toyota	0.106	0.408	0.261	0.794
	Others	0.828	0.432	1.916	0.055
Random effect					
	Variables	Variance		SD	
AV_MPC*OV_Type of collision	(Intercept)	20.472		4.525	
	Roadway surface	8.322		2.885	
	Accident cause	8.386		2.896	

p<0.05 : * , p<0.01 : ** , p<0.001 : ***

2.896으로 고정효과의 기울기와 비교했을 때 각 군집 별로 기울기의 변동성이 큰 것을 볼 수 있다. 이는 랜덤효과를 도입한 것에 대한 타당성을 간접적으로 말해주는 것으로 간주할 수 있다.

랜덤효과에 대한 조건적인 해석은 그룹별 랜덤절편 및 랜덤기울기의 조건부 기대값을 통해 해석할 수 있다. 첫째, ‘자율주행 차량의 충돌 직전 움직임’이 ‘정지’인 경우를 나타내는 Table 4.3을 살펴보면 다음과 같은 사실을 알 수 있다. 즉, ‘상대 차량의 충돌 유형’이 ‘차량 간 사고가 아닌 경우’ 자율주행 차량의 손상이 증대할 확률이 낮아진다. 이때, ‘도로 표면’이 미끄러운 경우 또는 자율주행 차량이 사고의 원인이 되는 경우 차량 손상이 증대할 확률이 높아진다. ‘상대 차량의 충돌 유형’이 ‘전방 충돌’인 경우 자율주행 차량의 손상이 증대할 확률이 높아진다. 이때, ‘도로 표면’이 미끄러운 경우 차량 손상이 증대할 확률이 낮아진다. ‘상대 차량의 충돌 유형’이 ‘측면 스칩’인 경우 자율주행 차량의 손상이 증대할 확률이 높아진다. 이때, ‘도로 표면’이 미끄러운 경우 차량 손상이 증대할 확률이 낮아진다. ‘상대 차량의 충돌 유형’이 ‘후방 충돌’인 경우 자율주행 차량의 손상이 증대할 확률이 낮아진다. 이때, ‘도로 표면’이 미끄러운 경우 또는 자율주행 차량이 사고의 원인이 되는 경우 차량 손상이 증대할 확률이 높아진다. ‘상대 차량의 충돌 유형’이 ‘측면 충돌’인 경우 자율주행 차량의 손상이 증대할 확률이 높아진다. 이때, ‘도로 표면’이 미끄러운 경우 또는 자율주행 차량이 사고의 원인이 되는 경우 차량 손상이 증대할 확률이 낮아진다.

둘째, ‘자율주행 차량의 충돌 직전 움직임’이 ‘주행’인 경우 Table 4.4로부터 다음과 같은 해석을 도출할 수 있다. ‘상대 차량의 충돌 유형’이 ‘차량 간 사고가 아닌 경우’ 자율주행 차량의 손상이 증대할 확률이 낮아진다. 이때, ‘도로 표면’이 미끄러운 경우 또는 자율주행 차량이 사고의 원인이 되는 경우 차량 손상이 증대할 확률이 높아진다. ‘상대 차량의 충돌 유형’이 ‘전방 충돌’인 경우 자율주행 차량의 손상이

증대할 확률이 높아진다. 이때, ‘도로 표면’이 미끄러운 경우 차량 손상이 증대할 확률이 낮아진다. ‘상대 차량의 충돌 유형’이 ‘측면 스침’인 경우 자율주행 차량의 손상이 증대할 확률이 낮아진다. 이때, ‘도로 표면’이 미끄러운 경우 차량 손상이 증대할 확률이 높아진다. ‘상대 차량의 충돌 유형’이 ‘후방 충돌’인 경우 자율주행 차량의 손상이 증대할 확률이 낮아진다. 이때, 자율주행 차량이 사고의 원인이 되는 경우 차량 손상이 증대할 확률이 높아진다. ‘상대 차량의 충돌 유형’이 ‘측면 충돌’인 경우 자율주행 차량의 손상이 증대할 확률이 높아진다. 이때, 자율주행 차량이 사고의 원인이 되는 경우 차량 손상이 증대할 확률이 낮아진다.

셋째, ‘자율주행 차량의 충돌 직전 움직임’이 ‘주차 및 후진’인 경우를 나타내는 Table 4.5를 통해 다음과 같은 시사점을 도출할 수 있다. ‘상대 차량의 충돌 유형’이 ‘차량 간 사고가 아닌 경우’ 자율주행 차량의 손상이 증대할 확률이 높아진다. 이때, ‘도로 표면’이 미끄러운 경우 또는 자율주행 차량이 사고의 원인이 되는 경우 차량 손상이 증대할 확률이 낮아진다. ‘상대 차량의 충돌 유형’이 ‘전방 충돌’인 경우 자율주행 차량의 손상이 증대할 확률이 높아진다. 이때, 자율주행 차량이 사고의 원인이 되는 경우 차량 손상이 증대할 확률이 낮아진다. ‘상대 차량의 충돌 유형’이 ‘측면 스침’인 경우 자율주행 차량의 손상이 증대할 확률이 낮아진다. 이때, ‘도로 표면’이 미끄러운 경우 또는 자율주행 차량이 사고의 원인이 되는 경우 차량 손상이 증대할 확률이 높아진다. ‘상대 차량의 충돌 유형’이 ‘후방 충돌’인 경우 자율주행 차량의 손상이 증대할 확률이 낮아진다. 이때, ‘도로 표면’이 미끄러운 경우 차량 손상이 증대할 확률이 높아진다. ‘상대 차량의 충돌 유형’이 ‘측면 충돌’인 경우 자율주행 차량의 손상이 증대할 확률이 낮아진다. 이때, ‘도로 표면’이 미끄러운 경우 차량 손상이 증대할 확률이 높아진다.

끝으로, ‘자율주행 차량의 충돌 직전 움직임’이 ‘차선 및 방향 전환’인 경우인 Table 4.6을 살펴보면, ‘상대 차량의 충돌 유형’이 ‘차량 간 사고가 아닌 경우’ 자율주행 차량의 손상이 증대할 확률이 낮아진다. 이때, 자율주행 차량이 사고의 원인이 되는 경우 차량 손상이 증대할 확률이 높아진다. ‘상대 차량의 충돌 유형’이 ‘전방 충돌’인 경우 자율주행 차량의 손상이 증대할 확률이 높아진다. 이때, 자율주행 차량이 사고의 원인이 되는 경우 차량 손상이 증대할 확률이 낮아진다. ‘상대 차량의 충돌 유형’이 ‘측면 스침’인 경우 자율주행 차량의 손상이 증대할 확률이 높아진다. 이때, 자율주행 차량이 사고의 원인이 되는 경우 차량 손상이 증대할 확률이 낮아진다. ‘상대 차량의 충돌 유형’이 ‘후방 충돌’인 경우 자율주행 차량의 손상이 증대할 확률이 낮아진다. 이때, ‘도로 표면’이 미끄러운 경우 차량 손상이 증대할 확률이 높아진다. ‘상대 차량의 충돌 유형’이 ‘측면 충돌’인 경우 자율주행 차량의 손상이 증대할 확률이 낮아진다. 이때, 자율주행 차량이 사고의 원인이 되는 경우 차량 손상이 증대할 확률이 높아진다.

Table 4.3 Conditional random effects for intercept and slopes (AV_MPC = 'Stopped')

AV_MPC*OV_Type of collision	(Intercept)	Roadway surface	Accident cause
Stopped*Not between vehicles	-2.099	0.154	0.841
Stopped*Head-on	2.348	-2.391	0.919
Stopped*Side swipe	1.254	-0.864	0.111
Stopped*Rear end	-4.312	0.226	1.799
Stopped*Broadside	2.739	-0.201	-1.097

Table 4.4 Conditional random effects for intercept and slopes (AV_MPC = 'Driving')

AV_MPC*OV_Type of collision	(Intercept)	Roadway surface	Accident cause
Driving*Not between vehicles	-3.822	1.201	1.500
Driving*Head-on	4.716	-3.234	0.896
Driving*Side swipe	-1.654	4.796	-2.373
Driving*Rear end	-2.475	-2.318	4.005
Driving*Broadside	5.789	0.648	-3.496

Table 4.5 Conditional random effects for intercept and slopes (AV_MPC = 'Parked and backing')

AV_MPC*OV_Type of collision	(Intercept)	Roadway surface	Accident cause
Parked and backing*Not between vehicles	5.218	-1.567	-0.996
Parked and backing*Head-on	2.951	1.269	-3.235
Parked and backing*Side swipe	-3.411	1.111	0.248
Parked and backing*Rear end	-3.088	1.927	-0.707
Parked and backing*Broadside	-0.380	1.301	-1.195

Table 4.6 Conditional random effects for intercept and slopes (AV_MPC = 'Changing lanes and direction')

AV_MPC*OV_Type of collision	(Intercept)	Roadway surface	Accident cause
Changing lanes and direction* Not between vehicles	-1.194	-0.188	0.609
Changing lanes and direction*Head-on	3.778	1.703	-3.089
Changing lanes and direction*Side swipe	2.867	0.069	-1.444
Changing lanes and direction*Rear end	-1.522	1.003	-0.485
Changing lanes and direction*Broadside	-1.772	-2.256	3.234

5. 결론 및 논의

본 연구에서는 자율주행 차량의 사고 발생 시 사고 유형과 환경 등이 자율주행 차량 손상에 미치는 영향을 알아보기 위해 다양한 앙상블 기법들과 혼합효과 로지스틱회귀모형을 비교하여 최적의 모델을 선정하였으며, 선정된 최적의 모델인 혼합효과 로지스틱회귀모형을 통해 구축 모형을 해석하였다. 분석결과 '주행모드'가 '자율주행모드'일 때, 차량의 손상 정도가 낮아지는 경향을 확인하였다. 따라서, 전반적으로 사고 발생 시 자율주행 차량은 사람이 운전하는 것에 비해 더욱 안전한 것으로 생각된다. '조명'은 어두울 때, 차량의 손상 정도가 높아지는 경향을 확인하였다. 하지만, 어두울 때의 중대한 사고 비율을 확인한 결과 '자율주행모드'가 '수동모드'에 비해서 안전한 것으로 생각된다. '부상자 여부'는 부상자가 보고 되었을 때, 차량 손상 정도가 높아지는 경향을 확인하였다. '자율주행 차량 제조사'는 'Chrysler'가 'Chevrolet'에 비해 차량의 손상 정도가 낮아지는 경향을 보였으며, 'Cruise'는 'Chevrolet'에 비해 차량의 손상 정도가 높아지는 경향을 확인하였다.

본 연구의 분석결과는 국내 자료가 아닌 해외 자료만을 이용한 결과로서 분석결과를 그대로 국내 여건에 적용하기에는 분명한 한계가 있다. 또한 가용 데이터의 한계로 인해 Kwon 등 (2021)과 Kwon 등 (2023)이 고려한 겨울철 도로살얼음이나 집중 호우 등에 대한 상황을 고려하지 못한 점도 아쉬운 부분이다. 한편, 국내 자율주행 사고 데이터는 아직 체계적으로 수집되어 활용되지는 못하고 있는 상황으로 보인다. 향후 국내 자료 활용이 가능해지면 우리 교통환경과 실정에 맞는 유의미한 분석도 가능할 것이다. 특히, 국내 자율주행 사고 자료 구축시 사고주체를 명확히 하고 책임비율 등을 함께 수집한다면 추후 연구에서 활용도가 매우 높을 것으로 판단된다.

References

- Boggs, A. M., Arvin, R. and Khattak, A. J. (2020). Exploring the who, what, when, where, and why of automated vehicle disengagements. *Accident Analysis and Prevention*, **136**.
- Chen, H., Chen, H., Liu, Z. and Zhou, R. (2020). Analysis of factors affecting the severity of automated vehicle crashes using XGBoost model combining POI data. *Journal of Advanced Transportation*.
- Chung, M. S. and Choi, B. Y. (2024). A study on the trend of traffic psychology using text mining: Applying topic modeling. *Journal of the Korean Data Analysis Society*, **26**, 725-741.
- Das, S., Dutta, A. and Tsapakis, I. (2020). Automated vehicle collisions in California: Applying Bayesian latent class model. *IATSS Research*, **44**, 300-308.

- Jin, H. J., Joo, A. R., Lee, D. K. and Cheon, S. Y. (2023). Autonomous driving research trend analysis using topic modeling. *Journal of the Korean Data Analysis Society*, **25**, 1671-1682.
- Khattak, Z. H., Fontaine, M. D. and Smith, B. L. (2019). An exploratory investigation of disengagements and crashes in autonomous vehicles. *Proceedings of the Annual Meeting of the Transportation Research Board*.
- Kim, H. S., Han, H. R., You, Y. S., Cho, M. J., Hong, J. H. and Song, T. J. (2024). A comprehensive traffic accident investigation system for identifying causes of the accident involving events with autonomous vehicle. *Journal of Advanced Transportation*.
- Ko, Y. S. and Lee, J. S. (2023). Take-over time and driving performances by driver's age group, types of non-driving related tasks, and road complexity in automated driving. *Journal of the Korean Data Analysis Society*, **25**, 1177-1192.
- Kuo, P. F., Hsu, W. T., Lord, D. and Putra, I. G. B. (2024). Classification of autonomous vehicle crash severity: Solving the problems of imbalanced datasets and small sample size. *Accident Analysis and Prevention*, **205**.
- Kutela, B., Das, S. and Dadashova, B. (2021). Mining patterns of autonomous vehicle crashes involving vulnerable road users to understand the associated factors. *Accident Analysis and Prevention*, **165**.
- Kwon, O. K., Kim, B. S. and Jung, S. K. (2021). A study on traffic impact by heavy rain using betweenness centrality analysis. *Journal of the Korean Data And Information Science Society*, **32**, 49-61.
- Kwon, T. Y., Yoon, H. C., Choi, Y. H. and Yoon, S. H. (2023). A comparative study on the prediction of road surface temperature. *Journal of the Korean Data & Information Science Society*, **34**, 751-761.
- Lee, S., Arvin, R. and Khattak, A. (2023). Advancing investigation of automated vehicle crashes using text analytics of crash narratives and Bayesian analysis. *Accident Analysis and Prevention*, **181**.
- Leilabadi, S. H. and Schmidt, S. (2019). In-depth analysis of autonomous vehicle collisions in California. *IEEE Intelligent Transportation Systems Conference*, 889-893.
- Wang, T., Chen, J., Li, W. and Ye, X. (2024). A precrash scenario analysis comparing safety performance across autonomous vehicle driving modes. *Journal of Advanced Transportation*.
- Zhou, R., Zhang, G., Huang, H., Wei, Z., Zhou, H., Jin, J., Chang, F. and Chen, J. (2024). How would autonomous vehicles behave in real-world crash scenarios? *Accident Analysis and Prevention*, **202**.

A statistical analysis of autonomous vehicle accident patterns and vehicle damage[†]

Ki-Yeong Sim¹ · In-Gyu Lee² · Jung-A Yang³ · Kyupil Yeon⁴

^{1,2,3}Department of Data Science, Hoseo University

⁴Division of Big Data and AI, Hoseo University

Received 8 October 2024, revised 12 November 2024, accepted 20 November 2024

Abstract

As autonomous vehicles (AV) are commercialized, they are expected to grow as a core technology field of the 4th industrial revolution, leading to major changes that bring new experiences to users. However, as the technology advances, the risk of autonomous vehicle accidents is inevitable. Accidents of autonomous vehicles cannot be judged as the fault of the autonomous vehicle, and traffic accidents or fatalities may occur due to unexpected circumstances or inadequate response of the technology. Therefore, this study aims to investigate the effects of accident types and environment on the degree of vehicle damage in autonomous vehicle accidents through mixed effect logistic regression. The data was collected using the autonomous vehicles collision reports provided by the California DMV. We found that autonomous driving mode tends to mitigate the damage of AV in accidents while some factors related on the accident environments such as the darker lighting condition, vehicle-to-vehicle accident, whether there were any injuries, and whether the AV has caused the accident, tend to increase the damage of a AV. We expect this study can help to better understand the causes of AV accidents and provide a guide to the development of autonomous vehicle technology.

Keywords: Autonomous vehicle, autonomous vehicle accident, California DMV, mixed effects logistic regression model.

[†] This work was supported by the National Research Foundation of Korea (NRF) grant funded by the Korea government (MSIT) (No. 2020R1F1A1A01073456). This research was supported by "Regional Innovation Strategy (RIS)" through the National Research Foundation of Korea (NRF) funded by the Ministry of Education (MOE) (2021RIS-004).

¹ Graduate student, Department of Data Science, Hoseo University.

² Graduate student, Department of Data Science, Hoseo University.

³ Graduate student, Department of Data Science, Hoseo University.

⁴ (Corresponding Author) Professor, Division of Big Data and AI, Hoseo University. 20, Hoseo-ro 79beon-gil, Baebang-eup, Asan-si, Chungnam, 31499, Korea. E-mail: kpyeon1@hoseo.edu