

자율주행 안전성 검증 시나리오 개발 활용을 위한 교통사고보고서 개선방향에 관한 연구

Study on the Improvement of Traffic Accident Report for Automated Vehicle Test Scenarios

오 경 택* · 고 우 리** · 박 지 혁*** · 윤 일 수**** · 소 재 현*****

* 주저자 : 아주대학교 교통공학과 석사과정

** 공저자 : 아주대학교 교통공학과 석·박사통합과정

*** 공저자 : 아주대학교 교통공학과 석사과정

**** 공저자 : 아주대학교 교통시스템공학과 교수

***** 교신저자 : 아주대학교 교통시스템공학과 조교수

OH, Gyungtaek* · KO, Woori* · PARK, Jihyeok* · YUN, Ilsoo** ·
SO, Jaehyun (Jason)**

* Dept. of Transportation Eng., Univ. of Ajou

** Dept. of Transportation System Eng., Univ. of Ajou

† Corresponding author : SO Jaehyun(Jason), jso@ajou.ac.kr

Vol. 21 No.2(2022)
April, 2022
pp.167~182

pISSN 1738-0774
eISSN 2384-1729
<https://doi.org/10.12815/kits.2022.21.2.167>

Received 8 February 2022
Revised 24 February 2022
Accepted 12 April 2022

© 2022. The Korea Institute of
Intelligent Transport Systems. All
rights reserved.

요 약

교통사고보고서 상의 교통사고 관련 자료 속성들은 그 원인을 파악하고자 하는 전통적인 교통안전 관련 연구에서 뿐만 아니라 최근 자율주행자동차의 안전성 검증 시나리오 개발을 위한 연구에서도 활용될 수 있다. 다만, 교통사고보고서의 자료 속성들은 교통사고 상황 재현 및 시나리오 개발만을 위해 정의된 항목들이 아니므로, 본 연구에서는 확대된 활용성 측면의 교통사고보고서의 개선방향을 제시하고자 한다. 교통사고보고서의 개선방향은 각각 교통사고 발생 이전 상황(pre-crash), 교통사고 중 상황(on-crash), 교통사고 발생 이후 상황(post-crash)로 구분하여 제시하였으며, 각 구분에 따른 추가 자료 항목 또는 자료 처리 방법에 대하여 제시하였다. 또한, 정형화된 형태의 교통사고자료 외에 비정형화된 서술 형태의 교통사고 경위자료로부터 추출 가능한 정보항목들을 도출하여 제시하였다.

핵심어 : 교통사고보고서, 교통사고 경위자료, 교통사고심각도, 교통사고 원인, 비정형자료

ABSTRACT

The accident data attributes of the traffic accident report are used not only in traditional traffic safety-related research to identify the cause of traffic accidents, but also as basis data for the development of the automated vehicle driving performance verification scenarios. However, since the data attributes of the traffic accident report are limited for the purpose of reconstructing the traffic situation and developing scenarios, this study aims to provide the directions for improvement of traffic accident report, ultimately for its expanded usability for the automated vehicle test scenarios. The directions for improvement of the traffic accident report are provided by categorizing the traffic situation before the accident (pre-crash), the situation immediately before or during the accident (on-crash), and the situation after the accident (post-crash), respectively. Additional data items or data processing methods are presented. Furthermore, data elements that can be extracted from the traffic accident process data in the unstructured narrative form are explored and provided.

Key words : Traffic accident report, Traffic accident attributes, Accident severity, Structured accident data, Unstructured accident data

I. 서 론

교통안전성 향상을 위해 널리 활용되어 온 교통사고보고서 상의 교통사고 관련 정보는 최근 위험한 교통 상황에서의 자율주행자동차의 안전성 검증을 위한 시나리오 개발에 활용되고 있으며, 자율주행자동차의 상용화가 촉진됨에 따라서 교통사고자료의 활용성이 확대되고 있다(Park et al., 2019). 「도로교통법」 제54조에 따르면 차 또는 노면전차의 운전 등으로 사람이 사상하거나 물건이 손괴된 경우 교통사고로 정의한다. 이때 「도로교통법 시행령」 제32조에 따라 경찰공무원은 대통령령으로 정하는 교통사고 발생일시 및 장소, 교통사고 피해 상황, 교통사고 현장 상황, 교통사고 관련자, 차량등록 및 보험가입 여부, 운전면허의 유효 여부, 술에 취하거나 약물을 투여한 상태에서의 운전 여부 및 부상자에 대한 구호 조치 등 필요한 조치의 이행 여부, 운전자의 과실 유무, 그 밖에 차, 노면전차 또는 교통안전시설의 결함 등 교통사고 유발 요인 및 「교통안전법」 제55조에 따라 설치된 운행기록장치 등 증거의 수집 등과 관련하여 필요한 사항들을 조사해야 한다. 이에 따라 교통사고가 발생하는 경우 경찰공무원은 현장조사와 교통사고 처리가 모두 끝난 후 앞에 언급된 각종 교통사고와 관련된 현황자료를 기입한 교통사고보고서를 작성한다. 기존의 연구에서는 교통사고자료를 활용하여 교통사고 원인 분석 및 대응방안에 관한 연구(Jung, 2019), 충돌시험을 통한 교통사고 재현(Kim et al., 2013), 교통사고 심각도 분류분석(Sohn and Shin, 2001) 등 과거에서부터 현재에 걸쳐 교통의 전반적인 분야를 분석하는 목적으로 활용되어왔다. 최근 연구에서는 자율주행자동차의 안전성 검증을 위한 위험 교통상황 시나리오 개발 연구에서도 교통사고자료가 활용되었고, 자율주행자동차 시대에 교통사고자료의 활용성은 시나리오 개발 또는 자율주행차 교통사고의 원인 분석 등에 다양하게 활용될 것으로 예상된다(Park et al., 2021).

다만, 현재의 교통사고보고서의 조사항목은 교통사고의 원인 분석에 초점을 맞추어 정의되어 있어, 교통사고 상황에 대한 재현 및 교통사고 상황에 기반한 자율주행자동차 시나리오 개발 등에 활용하는데엔 한계점이 존재한다. 가령, 교통사고 직전의 사고차량들의 속도나 진행방향, 사고차량 주변의 교통상황 등은 교통사고보고서에 기록되지 않으며, 교통사고의 원인과 결과간 인과관계를 명확히 파악하는데 도움이 되는 교통사고 후 피해정도에 대한 내용이 포함되어 있지 않다. 또한, 교통사고보고서에 이러한 사고 전후 상황에 대해 교통사고 조사 담당 경찰공무원이 교통사고 상황에 대해 조사자의 시각에서 글의 형식으로 설명한 교통사고 경위자료가 있지만, 자연어 형식의 자료에 대한 분석상 어려움으로 그동안 교통사고 원인, 교통상황 재현 등 많은 교통안전 관련 연구에 자주 활용되진 못하고 있는 실정이다.

따라서, 본 연구의 주요 목적은 교통사고의 원인 분석 또는 교통사고 시나리오 개발 등 다양한 안전연구를 위해서는 교통사고보고서에 더욱 상세한 내용이 포함되어야 한다는 전제 하에 교통사고 속성자료의 개선방향 및 글 형식의 교통사고 경위자료의 활용 등 교통사고보고서의 개선방향을 도출하고자 한다. 이를 위해 먼저 기존 교통사고보고서에 포함된 교통사고 관련 자료들의 활용성 극대화를 위해 글 형식의 교통사고 경위자료의 수기 분석을 통해 경위자료로부터 추출 가능한 정보항목들을 제시한다. 두 번째로, 교통상황 재현 및 교통사고 시나리오 개발을 위해 교통사고 전후 상황을 더욱 정확히 묘사하기 위해 필요한 교통사고보고서의 속성자료의 추가 및 수정과 같은 개선방향을 도출한다. 본 연구를 통해 교통사고보고서의 조사항목이 개선된다면 기존의 교통사고 원인 분석 연구 및 자율주행자동차의 교통사고 시나리오 개발 등 다양한 교통안전 관련 연구에 활용가능한 최상의 교통사고자료 자료 셋을 보유할 수 있을 것으로 기대된다.

II. 선행 연구 고찰

1. 교통사고자료 기반 상황별(pre-/on-/post-crash) 교통안전 연구

교통사고자료 중 교통사고 전 상황(pre-crash)에 관한 연구는 주로 교통사고 예측 모형 개발과 교통사고 원인 분석을 위해 활용되어 왔다. Lee et al.(2011), Kang and Lee(2002), Hwang et al.(2010), Char and Serre(2020)의 연구는 교통사고 예측 모형 개발을 위해 pre-crash에 해당하는 교통사고 자료를 활용하였다. 그 결과, 교통사고에 영향을 미치는 요소는 환경요소, 인적요소, 차량요소 등 다양한 요인이 존재하고 정확하고 세밀한 분석을 통해 신뢰성 높은 모형을 구축하기 위해서는 모든 요소를 고려해야 한다 결론지었다. Kim et al.(2005), Yoon and Lee(2018), Kim et al.(2006), Song(2021), Jung(2019), Tian et al.(2010)은 교통사고 원인 분석을 진행하였다. 교통사고 원인 분석을 통해 사고 지점에 대한 문제점을 확인하여 예방 및 안전한 도로 환경 조성에 활용하였고, 인적요인과 차량요인을 추가적으로 고려해야 하며, 도로의 환경요소를 더 세분화하여야 한다고 결론지었다. 이외에도 Shi et al.(2018)은 교통사고 전 위험 노출을 평가하기 위해 하이브리드 평가 지표를 개발하였고 이를 기반으로 시나리오를 재구성하여 교통사고 원인 및 영향 요인을 조사하였다.

교통사고 중 상황(on-crash)에 관한 연구는 주로 교통사고 자료를 통한 재현 또는 PC-Crash의 시뮬레이션을 활용해 분석을 진행하였다. Ha and Han(2003), Han(2021), Kim et al.(2013), Chen et al.(2019)의 연구는 교통사고 상황을 재현하여 교통사고가 사고차량의 최종 자세에 어떤 영향을 미치는지 확인하였고, 차량 속도별로 파손상태와 부상위험을 분석하였다. 그 결과, 차량 특성 외에도 인간 행동 조사가 필요하며 충돌 전 차량 자세를 검증하여 교통사고 원인을 규명할 필요가 있다고 확인하였다. 이 외에도 Shanthi and Ramani(2011)의 연구에서는 차량 충돌 패턴을 분류하는 규칙을 도출하여 충돌 방법을 예측하는데 교통사고자료를 활용하였다.

교통사고 후 상황(post-crash)에 관한 연구는 교통사고 심각도 및 교통사고 비용에 관한 연구들이 존재했다. Post-crash 자료를 활용한 교통사고 심각도에 관한 연구는 Lee et al.(2008), Jeong et al.(2017), Sohn and Shin(2001), Ha et al.(2005), Jung and Bae(2017)의 연구가 해당했다. 교통사고 심각도 분석 결과, 운전자 정보의 추가와 기하구조의 세분화 및 체계적인 데이터베이스화가 필요함을 확인하였다. Shim et al.(2018)은 교통수단별 교통사고 비용 추정을 통해 교통안전 정책 결정 지표로 활용하였고, 물적피해만 존재한 경우에도 데이터베이스화가 필요하다고 결론지었다. 이 외에도 Lee et al.(2015)의 연구에서는 교통사고 처리시간 예측모형을 개발하였고, Cryer et al.(2001)의 연구에서는 경찰의 교통사고보고서와 병원 데이터베이스가 연계되어야 한다고 주장하였다.

2. 교통사고자료 기반 자율주행자동차 시나리오 개발 연구

교통사고자료는 일부 연구에서 교통사고자료의 교통상황 설명력에 기반하여 자율주행자동차 시나리오 개발을 시도된 바 있다. Park et al.(2017)은 자율주행 실험도시인 K-City 커뮤니티부에서 실제 안전상황에 근거하여 자율주행자동차 안전성 평가를 위한 시나리오를 개발하기 위해 경찰청 교통사고 자료를 활용하였다. 그 결과, TF-IDF를 이용하여 보행자, 오토바이, 자전거 등 키워드와 교차로, 차로의 안전상황을 도출하였으며 이를 기반으로 대표 시나리오들을 구축하였다. Park et al.(2019)의 연구는 K-City 커뮤니티부도로를 중심으로 자율주행자동차의 안전성을 평가하기 위한 적절한 평가시나리오를 개발하기 위해 경찰청 교통사고 자료를 활용하였다. 또한, 실차 실험에 적용시키기 위해 5-레이어 포맷을 적용하였다. 그 결과, TF-IDF를 이용하여 주요 키워드를 도출하고 교통안전상황을 도출하여 24개의 functional 시나리오를 개발하였으며 이를 구성하

기 위해 5-레이어 모델기반으로 레이어별 변수 및 범위를 도출하였다. Park et al.(2018)의 연구는 자율주행자동차의 제어권 전환 안전성 평가를 위해 고속도로 구간과 주행 중 발생가능한 상황을 조합해 주행 시나리오를 구분하기 위해 2014년 발생한 고속도로 교통사고 경위자료를 활용하였다. 그 결과, 6개의 제어권 전환 고속도로 시나리오를 도출하였고, 실험 시 고려해야할 사항인 변수 36개를 정의하고 그 범위를 제시하였다. Eo et al.(2019)의 연구는 사고회피 시나리오를 작성하기 위해 교통사고분석시스템(TAAS)에서 제공하는 국내사고 자료를 활용하였으며 PreScan 시뮬레이터를 이용하여 정보조합에 따른 효과성 실험을 진행했다. 그 결과, 유니모달보다 멀티모달이 사고회피에 효과적이고 청각모달의 영향이 필수적임을 확인하였으며 다양한 조합의 정보제공 방법 중 1순위는 시각+청각+촉각임을 도출하였다. Nitsche et al.(2017)의 연구는 T자형 및 4지 교차로의 pre-crash 시나리오를 추출하기 위해 영국 교통부와 고속도로청이 의뢰한 On The-Spot(OTS) 프로젝트에서 가져온 1056건의 교차로 사고 자료를 활용하였으며 K-medoid 기법을 통해 분석 및 시각화를 포함하여 새로운 데이터 분석 방법을 제시하였다. 그 결과, T교차로에서 13개, 4지 교차로에서 6개의 충돌 클러스터를 생성했으며 부상 결과가 높은 클러스터를 고려하여 시뮬레이션에서 평가할 운전 상황의 핵심 모집단을 구성하는 12개의 시나리오를 도출하였다.

3. 비정형자료 분석 기법

교통사고보고서는 사전에 정해진 코드 형식의 정형자료와 자유로운 글(text) 형식의 비정형자료가 포함된다. 교통사고 경위에 대해 설명하는 비정형자료는 교통상황에 대한 조사자의 시각을 통해 교통사고 상황을 이해할 수 있다는 점에서 도움이 되는 반면, 자연어의 분석 측면의 어려움으로 그간 교통안전 연구에 자주 활용되지는 못했다. 다만, 최근 데이터마이닝(data mining) 기법의 발전으로 타 분야에서는 자연어 처리 기법을 활용한 연구가 다수 수행되어 왔다. Scheurwegs et al.(2016)은 진료기록의 정형화된 자료와 글 형태의 비정형자료를 결합한 임상진단 방법론을 개발하였다. 이 연구는 2년간의 병원진료기록 자료를 사용하였으며, Bayesian network 기반 hill climbing 기법을 통한 각종 환자 관련 검사수치 등의 정형화된 자료와 의사의 소견이 포함된 비정형자료 분석방법을 개발하였다. 그 결과, hill climbing 기법 기반의 진료기록 자료의 구조화된 학습에 의한 임상진단의 정확도가 7.7% 향상된 것으로 나타났다. Hur et al.(2019)은 보안 기업 회사에서 사이버 공격을 예측하고 대응하기 위해 대량의 비정형 Security Intelligence Report(SIR)을 기반으로 키워드 추출, 토픽 모델링, 문서 요약, 유사도 문서 검색 5가지 분석기술이 적용된 프레임워크를 제안하였다. 이후 이를 정형화하여 효율적으로 주요 정보를 추출하고자 하였다. 그 결과 SIR에서 자동으로 위협정보를 인지하고 추출하는 정확도는 73.3%로 높은 정확도를 가진다는 결과를 도출하였다. Ha and Ahn(2019)은 직업훈련 중 중도탈락을 예측하기 위해 정형자료와 비정형자료를 동시에 고려하는 모형을 구축하였다. 정형자료에서 유의한 변수를 선택하고, 비정형자료에서 텍스트마이닝(text mining)과 합성곱 신경망(Convolutional neural network, CNN) 분석을 통해 얻은 벡터값을 이중결합하여 정형자료와 비정형자료를 퓨전(fusion)하였다. 그 결과, 정형자료만 활용하여 예측했을 때에 비해 최대 약 20%까지 예측 정확도가 향상되었고, 검증 자료를 이용해 검증한 결과 예측 정확도가 최대 약 96%로 우수한 성능을 확인하였다. Zhang et al.(2020)은 의사의 clinic notes와 같은 비정형자료를 환자의 성별, 나이와 같은 정형자료와 결합하기 위해 Fusion-CNN과 Fusion-LSTM(long short term memory networks)을 이용하였다. 정형자료와 비정형자료의 퓨전을 통해 환자의 사망률, 재입원 확률, 장기입원 확률 등을 도출하였고, 퓨전된 새로운 자료를 이용하는 방법이 더 나은 결과를 도출함을 확인하였다.

4. 시사점

관련 연구 고찰 결과, 교통사고자료를 활용한 교통사고 원인 파악 연구가 다수 수행되었으며, 자율주행자

<Table 1> Comparison of Structured Elements Traffic Accident Reports in Korea and Abroad

Traffic Accident Report	Korea Police Agency	California	Texas	Europe
Structured Elements	Accident type Violation of the law Call the police Weather Road type Road shape Accident level Including the elderly and children Classification Perpetrators and victims Vehicle type Vehicle usage Drink Run away Have a license Damage classification Accident factor Injured area protective equipment Level of damage to the car	Vehicle movement Involved in the accident Describe vehicle damage Weather Lighting Roadway surface Roadway conditions Movement preceding collision Type of collision Other associated factor	Roadway system Roadway part street prefix, direction from int. or ref. marker Street suffix Unit description Vehicle color Body style Driver license/ID type Driver license class Commercial driver License endorsements Driver license restrictions Person type Seat position Injury severity Ethnicity Sex Ejected Restraint used Airbag Helmet use Solicitation Alcohol specimen type Drug specimen type Drug test result Drug category Financial responsibility type Vehicle damage rating Vehicle operation Carrier ID type Roadway access Vehicle type Hazardous material class number Cargo body style Trailer type Sequence of events Factors and conditions Vehicle defects Weather condition Light condition Entering roads Roadway type Roadway alignment Surface condition Traffic control	Injury Material damage Vehicle movement Does the policy cover material damage to the vehicle Drink Drug Has an official report been drawn up Send insurer Immobilized vehicle Reugular driver Deduct the V.A.T. regarding the damaged good

동차 운행성능 검증을 위한 시나리오 개발을 포함한 교통사고상황 재현 연구에 대한 활용도는 낮은 것으로 분석되었다. 교통사고자료의 항목별 활용성 측면에서는 정형화된 데이터의 활용성이 높았는데, 이는 정형화된 자료가 활용성 측면에서 용이하기 때문인 것으로 파악된다. 반면, <Table 1>과 같이 정형화된 교통사고자료 항목을 국외 교통사고보고서와 비교한 결과 국내 항목은 교통사고 발생시 인적요인 및 차량요인에 관한 내용이 부족하고 교통환경 요인과 기하구조 측면에서 역시 관련 속성데이터에 대한 세부적 분류가 부족했다. 이는 교통사고 원인 분석에 있어 단편적인 해석만 가능함을 의미하며, 교통사고상황 재현 및 자율주행자 동차 시나리오 개발에 필요한 자료가 기록되지 않아 많은 제한점들이 존재했을 것으로 예상된다. 따라서, 본 연구에서는 교통사고보고서의 기존 항목을 세분화하거나 추가적으로 필요한 항목을 도출하여 교통사고보고서의 개선점을 제시하고자 한다. 그 외에도, 기존의 비정형자료 분석기법을 활용한 타 분야 연구들에 따르면, 정형자료와 비정형자료의 퓨전 또는 비정형자료의 자연언어처리 기법을 적용한 정형화 연구가 다수 수행되어왔다. 본 연구에서는 비정형자료로부터 추출가능한 정보항목을 확인하고, 이 자료들을 퓨전하여 활용하는 방안에 대해 제시하고자 한다. 비정형자료를 정형화해 정형자료와 퓨전할 경우 교통사고보고서 내용을 더욱 구체화할 수 있게된다. 이러한 측면에서 교통사고보고서 내의 정형자료와 새롭게 정형화된 자료의 퓨전은 기존 정형자료에서 알 수 없었던 교통사고 상황의 세분화가 가능한 동시에, 비정형자료를 추가적으로 분석하지 않아도 된다는 장점을 가질 것으로 기대한다.

III. 연구 방법론

1. 교통사고자료 수집

본 연구에서는 2014년 대구지역에서 발생한 경찰청 교통사고자료 총 20,776건 중 약 10%에 해당하는 2,000건의 교통사고자료를 활용하였다. 교통사고자료는 경찰공무원이 교통사고현장 및 상황에 대해서 수기로 작성한 비정형자료와 교통사고유형, 법규위반, GPS 좌표, 도로종류 및 형태, 기상상태 등 정형자료로 구성되어 있다. 이 중 3중 이상의 다중충돌 교통사고, 비정형자료 내용과 정형자료의 입력값이 다른 교통사고, 자료가 불충분한 교통사고는 분석에서 제외하였다. 그 이유는 입력값이 다른 데이터는 신뢰성이 떨어지며, 자료가 불충분한 데이터는 차량의 사고 직전 움직임·충격위치·사고유형 등에 누락이 존재해 분석에 어려움이 있었기 때문이다. 그 결과 2,000건의 데이터 중 약 12%에 해당하는 243건을 제외하였다.

2. 교통사고자료 항목별 분류

교통사고보고서는 사고 원인에 해당하는 요인만을 작성자가 주관적으로 기록한다. 이는 사고 상황 및 사고 후 상황에 대한 자료의 부재로 교통사고시나리오 및 재현 목적으로 활용되기 어렵다. 따라서, 교통사고자료의 기존 항목을 분석하기 위해 교통사고보고서에서 조사하는 항목 중 활용목적 측면의 가능성을 검토하였다. 각 항목의 특성에 따라 교통사고 전 상황에 대한 분석(pre-crash), 교통사고 상황에 대한 분석(on-crash), 교통사고 후 상황에 대한 분석(post-crash)으로 그 활용목적을 분류하였다. 교통사고보고서는 교통사고(접수)번호, 교통사고 일자, 시간, 장소, 좌표, 교통사고(상황)개요, 교통사고유형, 법규위반 등의 교통사고 일반정보와 성명, 주민등록번호, 차량번호 등의 운전자 정보, 회사명 등의 사업용자동차 정보, 피해구분, 차량손상정도, 차량수리비의 피해물 정보로 구분되어 총 35개 항목이 포함되어 있다. 총 35개의 교통사고자료 항목 중 정의

가 모호하거나 세분화가 필요한 항목을 도출하여 명확하게 정의하고, 하위 구분을 도출하였다.

3. 교통사고 경위자료 분석 방안

교통사고자료는 교통사고에 관한 정보가 정형자료와 비정형자료로 기술되어 있다. 정형자료의 항목에는 교통사고 일시, 장소, 차종, 유형 등 교통사고에 관한 개략적 정보가 코드로 구분되어 기술되며, 비정형자료 항목은 교통사고가 발생한 과정과 교통사고 관련 객체의 움직임, 비정상적인 노면 상태 등의 정보가 기술되어 있다. 정형자료의 경우 코드로 구분되는 등 일정한 형식 및 규칙을 가지기 때문에 해석이 용이하고, 상황이 객관화되어 분석이 쉬우므로 빈번하게 활용된 것으로 판단된다. 반면, 비정형자료의 경우 일정한 규칙 없이 저장된 자료로 정형자료에 반영하기 어려운 세부 정보의 저장이 가능하지만 분석이 어렵다. 교통연구에서는 연구목적에 따라 정형자료와 비정형자료 중 더 적합한 자료를 각각 활용하였지만 주로 정형자료의 활용이 더 빈번하였다. 이에 세부적 교통사고 상황 정보가 포함된 비정형자료의 연구 활용성을 높이기 위해 본 연구에서는 정형자료를 비정형자료와 수기로 비교 및 분석하여 비정형자료에서 추가로 추출 가능한 항목을 정형화 후 두 자료를 퓨전하여 기존 교통사고보고서의 개선 항목을 도출하고자 한다.

4. 교통사고보고서 항목 검토 방안

본 연구에서는 기존 교통사고 경위자료와 새로운 퓨전 자료를 기반으로 목적(pre-/on-/post-crash)별 연구를 진행할 경우 추가적으로 필요한 항목 및 요인을 연구하였다. 교통사고 발생원인 및 대응방안에 초점을 둔 연구 분야는 ‘pre-crash’, 교통사고 재현이나 충돌의 진행체적을 분석하는 분야는 ‘on-crash’, 교통사고비용 추정이나 심각도 분류 등 사후처리는 ‘post-crash’로 구분한다. 이를 위해 목적별로 기존 연구들을 참고하여 각각의 연구들에서 향후 연구가 필요한 부분들과 한계점에 대해서 분석하였다. 목적별 연구 분석 후 연구들을 통해 기존 연구들을 보완할 수 있는 정보항목을 정의하고 이를 기반으로 교통사고의 발생 원인을 파악하는데 활용 가능한 항목, 교통사고 상황 설명이 가능한 항목, 교통사고 발생형태에 따른 심각도 또는 교통사고 후 발생한 교통사고 비용 항목으로 구분하였다. 통합된 교통사고 경위자료를 기반으로 각 연구 목적별 부족한 점을 분석하고, 설명력을 높이기 위해 필요한 추가적인 정보 항목을 제안하고자 한다.

IV. 교통사고자료 개선 항목 도출 결과

1. 비정형자료 분석 기반 신규 정형화 자료 항목 도출

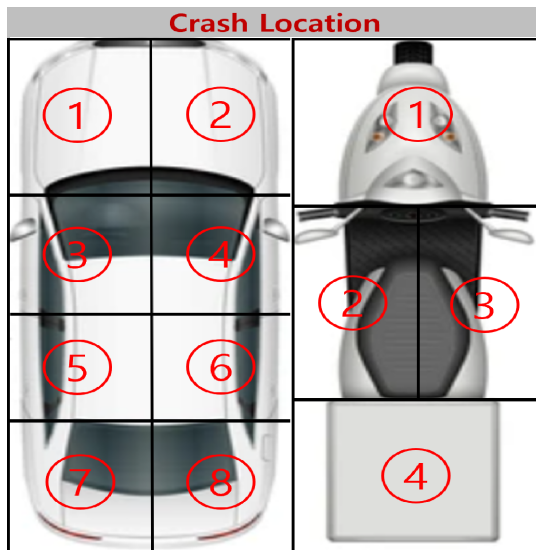
본 연구에서는 비정형자료 분석을 위해 총 20,776건의 교통사고자료 중 다중사고 자료와 누락 자료를 제외한 후 전체 자료의 약 10%에 해당하는 2,000여 건의 교통사고 개요를 랜덤 샘플링하였다. 이후 각각 사고에 대해 <Table 2>와 같이 비정형자료와 정형자료 항목을 비교하였다. 비교 결과, 교통사고보고서에서 정의한 코드로 구성된 정형자료에 비해 비정형자료는 교통사고 관련 객체 움직임, 교통사고 원인 등의 교통사고 관련 세부 정보를 포함하고 있는 것으로 나타났다. 즉, 정형자료에서는 교통사고 상황에 대한 교통사고 종류, 원인, 날씨조건, 도로 종류 등의 객관적인 정보를 포함하고 있는데 반해, 비정형자료를 통해서는 교통사고 시 차량의 충돌위치, 교통사고 직전 차량의 움직임 등 교통사고 피주체의 교통사고 직전 움직임과 결과적

인 충돌위치와 같은 미시적인 정보들을 포함하고 있는 것으로 나타났다.

<Table 2> Sample Comparison of Structured and Unstructured Accident Data

Accident Description (Unstructured)										
#1 The accident occurred when the left front of the motorbike collided with the right front bumper of the #2 vehicle, which was moving in the same direction after changing its lane from Bokhyeon Ogeori to the left side of the motorbike.										
vs.										
Accident Record (Structured)										
Item name	Accident type	Violation of the law	Vehicle type	Call the police	Weather Condition	Road type	Road shape	Alcohol	Run away	
Meaning	Side right angle collision.	Not getting a safe distance.	Motor	Yes	Sunny	Metropolitan road	Etc	No	No	etc
Code	220	03	06	01	01	03	08	02	02	

교통사고 상황을 재현하기 위해서는 비정형자료와 같은 미시적인 정보들이 필요하다. 따라서 비정형자료를 통해 얻을 수 있는 교통사고 세부 정보를 정형화하여 활용하는 것이 필요하다 판단했고, 교통사고 경위자료를 통해 추가로 추출가능한 정보항목들에 대해서는 <Table 3>을 통해 제시하였다. 그 대표적인 사례로, 비정형화된 교통사고 경위자료의 해석을 통해 <Fig. 1>과 같이 교통사고 상황에서 교통사고 객체간 충돌위치에 대한 정보를 확인할 수 있었다. 또한, 비정형자료 분석 결과 한 건의 사고에서도 차량에서 다방면 충돌위치가 확인되었기 때문에 충돌위치에 대한 복수기입이 가능하게 정형화가 필요하다. 교통사고 객체간 충돌위치는 교통상황 재현 및 실험시나리오 개발에 있어 교통사고 직전의 객체간 움직임을 묘사할 수 있어 활용성이 높다. 특히 충돌위치에 따라 교통사고상해정도가 달라질 수 있으므로 교통사고 직전의 객체들의 움직임과 교통사고 발생 후 상해도와의 인과관계를 파악하는데 역시 도움이 될 것으로 보인다.



<Fig. 1> Proposal of Collision Location Schematic

추가적으로 진로변경 중, 2차로에서 3차로로 진로변경, 우회전 중, 좌회전 중 등 차로변경과 회전 중 교통사고가 빈번하게 발생하는 것을 확인할 수 있었다. 하지만 기존 교통사고유형은 해당 교통사고와 관련된 항목이 존재하지 않는다. 이로 인해 관련 교통사고의 교통사고유형 기록은 기타, 측면 직각충돌 등 비관련자가 정형자료를 사용할 경우 오해할 수 있는 상황이 발생한다. 이를 방지하기 위해 차선변경과 회전 중을 신규 항목으로 추가하여 명확한 교통사고유형 규정을 통해 교통사고 상황의 신속한 판단을 가능하게 할 것으로 보인다. 또한, 직진하던 중, 후진하던 중, 정차 중 등 교통사고와 연관된 객체들의 개별적인 교통사고 직전 움직임들을 확인할 수 있었다. 교통사고가 발생하기 전 개별적인 객체들의 교통사고 직전 움직임은 앞선 층

<Table 3> New elements based on unstructured accident descriptions

Classification	Item name	Detailed elements	The Code defined in this study
Traffic Accident Type	Car-to-Car	Lane change	250
		Roatating	260
Movement just before an individual accident	Car, Motorcycle, etc	Go straight	11
		Go staright(Speeding down)	12
		Back	13
		Stop	14
		Left turn	15
		Right turn	16
		Lane change	17
		U-turn	18
		etc	19
	Pedestrian, Bicycle, etc	Walking(sidewalk)	21
		Standing(sidewalk)	22
		Crossing(sidewalk)	23
		etc(unknown)	29
Collision location	Car	Left-1	1
		Left-2	3
		Left-3	5
		Left-4	7
		Right-1	2
		Right-2	4
		Right-3	6
		Right-4	8
		Non-contact	9
	etc(unknown)	99	
	Motorcycle/Bicycle	Forward	1
		Mid-Left	2
		Mid-Right	3
		Backwards	4
etc(unknown)		99	

돌위치와 연관된다. 이는 노면 흔적물이 없는 경우 사용하는 PC-Crash의 충돌 전 진행 자세를 명시해주어 현장조사 과정을 생략해주기 때문에 교통사고 재현 연구에 도움이 될 수 있을 것으로 보인다.

비정형자료 분석을 통해 새롭게 추가 가능한 항목을 객관화된 범례에 의해 구분하면 <Fig. 1>, <Table 3>와 같다. <Table 3>과 같이 비정형자료를 정형화해 자료 항목을 새롭게 정의한 후 <Table 4>와 같이 기존 정형자료에 푸전한다면 비정형자료를 분석하기 위한 텍스트 마이닝 등의 과정을 생략 가능하며 동시에 비정형자료 항목에서만 얻을 수 있던 자료도 모두 쉽게 분석 가능하게 된다. 즉, 푸전된 교통사고자료는 비정형자료와 같이 구체적으로 묘사된 교통사고 상황을 정형화된 자료로 보유해 현실적인 교통사고 재현이 가능한 동시에 분석이 용이하다는 장점을 지닌다. 따라서 본 연구에서는 비정형자료의 자연어 분석을 통해 새로운 정형화된 형태의 자료 항목을 추출하였다.

<Table 4> Proposed set of traffic accident data elements

Classification	Item name	Detailed elements	Previous + New Code
Traffic accident type	Car-to-Pedestrian	Crossing	110
		Passing the road	120
		Passing through the edge of the road	130
		Walking on the sidewalk	140
		etc(unknown)	199
	Car-to-Car	Head-on collision	210
		Crash in another direction	220
		Collision(in progress)	230
		Collision(stop/parking)	240
		Lane change	250
		Roatating	260
	Only Car	etc(unknown)	299
		Collision of a structure	310
		Off the road(fall)	320
		Off the road(others)	330
		Parking car crash	340
		Overtum	350
	Crossing the Railroad	etc(unknown)	399
		Breaking through the breaker	410
		Ignore the alarm	420
Go straight		430	
Movement just before an individual car accident	Car, Motorcycle, etc	etc(unknown)	499
		Go straight	11
		Go staright(Speeding down)	12
		Back	13
		Stop	14
		Left turn	15
		Right turn	16
		Lane change	17
		U-turn	18
	etc(unknown)	19	
	Pedestrian, Bicycle, etc	Walking(sidewalk)	21
		Standing(sidewalk)	22
		Crossing(sidewalk)	23
		etc(unknown)	29

<Table 4> Proposed set of traffic accident data elements (continuation)

Classification	Item name	Detailed elements	Previous + New Code
Collision location	Car	Left-1	1
		Left-2	3
		Left-3	5
		Left-4	7
		Right-1	2
		Right-2	4
		Right-3	6
		Right-4	8
		Non-contact	9
		etc(unknown)	99
	Motorcycle/Bicycle	Forward	1
		Mid-Left	2
		Mid-Right	3
		Backwards	4
etc(unknown)		99	

2. 분석목적별 교통사고자료 개선방향 제시

교통사고 전 상황(pre-crash), 교통사고 중 상황(on-crash), 교통사고 후 상황(post-crash)의 세 가지 분석목적별 교통사고자료의 활용 측면의 개선방향을 제시한다. Pre-crash의 경우 기존 연구들은 도로의 기하구조, 인적, 차량, 환경요인 등을 분석항목으로 연구를 진행하였다. 하지만 교통사고 다발지점이나 특정 지역으로 국한되는 점과 교통사고 당시 주변 상황을 고려할 수 있는 요소가 부족한 한계가 연구고찰을 통해 발견되었다. 이에 한계점을 보완할 수 있는 항목들로 교통량, 시설물 설치 현황, 도로 선형을 제시한다. 교통량, 시설물 설치 현황, 도로 선형에 대한 정보가 제공되면 교통사고 당시 주변 상황에 대한 파악이 더욱 세부적으로 가능해지며, 이러한 세부 정보를 통해 교통사고가 도출된 복합적인 원인에 대해 심층분석이 가능해진다. 본 연구의 분석 범위인 대구시의 경우 대구종합교통정보 홈페이지를 통해 교통량 정보 취득이 가능하다. 이 외에도 서울특별시의 경우 서울시 교통정보 시스템(TOPIS), 수원시의 경우 수원교통정보와 같은 지자체의 교통정보 포털을 통해 도로 종류 및 구간에 해당하는 차종 구성비, 첨두시 교통량과 같은 정보를 취득할 수 있다. 도로 시설물의 경우 교량, 터널, 지하차도의 정보를 건설 Continuous Acquisition & Life-cycle Support(CALS) 포털, 도로 선형의 경우 지자체 도로 관리 시스템 또는 한국건설기술연구원의 국토 관리 시스템을 통해 정보의 취득이 가능하다. 따라서 시간, 위치, 지점 명, 유입 방향 등 key 값을 입력하면 교통량, 도로 시설물, 도로 선형 자료가 포털 시스템과 연계되어 각각의 정형자료 항목에 입력되는 방안을 제시한다.

On-crash의 기존 연구들은 실제 발생한 교통사고의 현장조사자료를 이용한 충돌시험이나 PC-Crash 프로그램을 사용하여 교통사고 재현을 진행하였다. 충돌시험이나 프로그램에 사용되는 요소들은 차량 최종 정지 자세, 충돌 직전 속도, 충돌 위치, 노면 흔적 등이다. 충돌 직전 속도는 PC-Crash 프로그램의 경우 차량의 무게중심 속도와 각속도를 이용하여 프로그램 내부의 알고리즘을 통해 산출하고(Ha and Han, 2003), 충돌시험의 경우 실차 실험을 통해 속도별 실험 차량과 교통사고 차량의 손상상태 비교를 통해 속도 범위를 추정한다(Kim et al., 2013). 기존 연구들은 충돌 직전 속도 추정에 있어 다양한 변수를 이용한 수식을 사용하거나 특정 장비를 통해 추정하므로 전문성이 필요하여 많은 시간과 노력이 소요된다. On-crash에서 다양하게 활용되는 충돌 직전 속도를 보다 효율적으로 분석할 수 있다면 연구 목표를 이루는 데 집중할 수 있게 된다. 따

라서, 충돌 직전 속도의 분석을 위하여 기존 속도 추정보다 상대적으로 직관적인 블랙박스 자료 사용을 제시한다. 블랙박스 자료는 영상과 함께 비영상 자료인 녹화 시간, 충격량, GPS 정보 등이 함께 저장되어(An and Lee, 2015) 충돌 직전 속도 추정 정확도가 제고된다고 판단하였다. 블랙박스 자료를 사용하면 이동 거리와 이동 시간을 통해 누구든 속도를 추정할 수 있으며, GPS 정보에 포함된 위치와 시간 정보를 통해 대상 차량의 위치나 동선을 확인 가능해 이동 속도 파악에 유용하다. 실제로 이를 기반으로 차량의 이동을 시뮬레이션으로 재현할 수 있는 연구가 진행되었다(Rajendran et al., 2011). 따라서 본 연구에서는 정확한 교통사고의 재현을 위해 블랙박스 자료를 연계한 교통사고 직전 개별차량 속도를 추가 항목으로 제시한다.

Post-crash의 경우 기존 연구들은 교통사고 심각도 분석과 교통사고 비용 분석의 두 가지 경우로 나뉜다. 교통사고 심각도 분석의 경우 사상자 수를, 교통사고 비용 분석의 경우 인적, 물적, 행정적, 심리적 피해비용을 분석한다. 현재 인적 피해 정도는 의사의 진단서를 경찰서에 제출하여 입력된다. 교통사고가 경미하여도 의사가 3주 이상의 진단을 내리면 중상이 되는 것이고 교통사고가 심각하여도 3주 미만의 진단이면 경상으로 입력되며 중상의 경우 성별, 나이 등으로 인해 교통사고 후 인적 피해가 사망으로 변경되기도 하여 정보가 불확실하다. 또한 경미한 교통사고는 신체를 다쳐도 상대가 보험처리를 해주어 경찰조사로 이어지지 않으면 사건이 종결되어 진단서를 제출하지 않아 인적 피해 교통사고가 아닌 물피 교통사고로 종결되는 경우가 발생하여 교통사고의 정보 또한 불명확하다(Song, 2021). 따라서 기존 항목인 사상자 수 항목을 그대로 유지하되 경찰, 보험사, 병원을 연계하여 실제 사상자 수를 조사하기 위한 자료 연계가 필요하다고 판단한다. 또한, 교통사고 처리시간을 추가 항목으로 제시한다. 교통사고 처리시간은 교통사고 처리뿐만 아니라 교통사고로 인해 발생한 지체가 해소되어 본래 흐름으로 회복하기까지 걸린 시간을 의미한다. 현재는 각 돌발상황이 교통사고에 어떤 영향을 얼마나 주는지 추정하는데 검증이 불가하지만, 교통사고 처리시간을 추가 항목으로 가

<Table 5> Summary of Traffic Accident Data Reformation

	Proposed new/ revised elements for		
	Pre-crash analysis	On-crash analysis	Post-crash analysis
Proposal of new elements	<ul style="list-style-type: none"> • Road-traffic characteristics and situations - (Elements) traffic volume, road geometry, and traffic facilities near a crash - (Data collection) by external linkage to road inventory or traffic information system of road authorities • Non-compliance with safe driving - (Elements) retaliatory driving, reckless driving, and falling loads - (Data collection) by manual investigation on site • Non-compliance with safety behavior - (Elements) using a nomadic devices while driving, by manual investigation on site - (Data collection) by manual investigation on site • Driving course violation - (Elements) violation of overtaking prohibition and cautions - (Data collection) by manual investigation on site 	<ul style="list-style-type: none"> • Vehicle speed just before the accident - (Elements) vehicle speed in collision - (Data collection) post-investigation of in-car blackbox storage 	<ul style="list-style-type: none"> • Fatalities and property damages - (Elements) Fatalities, injuries, and property damages - (Data collection) post-investigation of police agency, insurance companies, and hospitals • Traffic accident recovery duration on roadways - (Elements) traffic recovery time - (Data collection) post-investigation of local traffic management system (CCTVs)
Extraction from unstructured accident description	<ul style="list-style-type: none"> • Detailed driving maneuver and movement just before the accident - Stop, deceleration, lane change, turnings • Detailed traffic accident type - In lane-change, overturn, road departure 	<ul style="list-style-type: none"> • Collision location on object - Exact collision location on the object surface 	

지면 교통사고 발생 유형에 따른 영향도 척도로 활용할 수 있다. 교통사고 처리시간의 경우 도로 관리 시스템의 도로 정보 수집 장치로 수집한 영상 자료를 분석하여 도출이 가능하며, 교통사고 처리시간을 분석하면 교통사고 심각도에 따라 다른 값을 가지는 관계가 있을 것으로 판단되어 하나의 신규 항목으로 제시한다.

V. 결론

도로에서 발생하는 교통사고는 자료화되어 교통사고보고서에 저장된다. 교통사고 연구는 교통사고 원인, 교통사고 상황, 교통사고 후와 같이 연구 목적별로 교통사고 자료에 기록된 교통사고 경위자료를 기초 자료로 사용한다. 하지만 현재 교통사고자료는 실제 교통사고 상황을 재현하기에 부족하다. 이에 교통사고 연구의 전 분야에서 사용되는 교통사고자료 개선 연구가 필요하다고 판단하였다. 본 연구에서는 2014년 대구지역에서 수집된 교통사고자료를 수기로 분석하여 연구를 진행하였다. 우선 수집된 자료 중 비정형자료의 텍스트를 분석해 정형자료와 비교했다. 그 결과 교통사고보고서의 항목 중 교통사고유형 항목에서 차선변경, 회전 중을 추가하고, 신규 항목으로 충돌 위치와 충돌 직전 개별차량 움직임 추가를 제안하였다. 또한, 외부 시스템의 연계를 통해 pre-/on-/post-crash의 목적별로 진행되었던 기존 연구들을 고찰하여 추가적으로 필요한 정형자료 항목들을 제시하였다. Pre-crash는 교통사고 원인을 파악하는 연구로 교통사고 당시 주변 상황을 더 폭넓게 반영하기 위해 포탈 시스템과 연계되어 교통량 정보, 시설물 설치 현황, 도로 선형을 정형자료의 추가 항목으로 제시하였다. On-crash는 교통사고 중 상황을 재현하기 위해 블랙박스 자료를 연계한 교통사고 직전 개별차량 속도를 추가 항목으로 제시하였으며, post-crash는 경찰, 보험사, 병원 자료가 연계된 사상자 수와 도로 정보 수집 장치로 분석 가능한 교통사고처리 시간을 신규 항목으로 제시하였다.

본 연구는 향후 교통사고보고서의 교통상황 재현 및 자율주행자동차 시나리오 개발 등에서의 폭넓은 활용을 위한 기초연구라는데 의의가 있다. 현재의 교통사고보고서는 교통사고 상황에 대한 기록 또는 교통사고의 발생원인 분석을 위해 개발된 기록문서로서, 교통사고에 대한 원인 분석 외의 목적으로 활용하기엔 한계가 있다. 특히, 교통사고가 자주 발생하는 상황에 대한 재현 및 이에 기반한 자율주행자동차 교통사고 시나리오 개발을 위해서는 교통사고보고서에 교통사고 발생 당시의 상황을 묘사할 수 있는 더욱 자세한 정보가 필요하다. 기존 자율주행 안전성 검증 시나리오 개발 연구에서는 교통사고보고서의 다양한 자료 중 비정형으로 작성된 사고경위를 TF-IDF, K-medoid와 같은 별도의 모델을 이용하여 분석하였다. 이러한 분석방법은 연구를 진행함에 있어 자료의 전처리과정이 요구되어 많은 시간 소요를 유발한다. 반면, 본 연구는 비정형자료 기반 정형자료를 추가로 제시함으로써 시나리오 개발 연구에 필요한 데이터 탐색에 있어 시간적, 비용적 효율 확보가 기대된다. 본 연구는 이러한 목적 하에 추가적으로 필요한 교통사고조사항목 및 교통사고 경위자료로부터 추출가능한 정보항목 등을 규정함으로써 교통사고보고서의 활용성 증대를 위해 필요한 개선방향을 제시하였다. 또한, 향후 국내 교통사고자료 자체에 대한 가치를 높이고, 교통사고에 대한 정밀한 분석 및 교통사고자료 기반 실험시나리오 개발 등에 교통사고자료가 널리 활용될 수 있는 기반이 될 것으로 기대된다.

다만, 본 연구는 교통사고보고서에 대한 저자들의 분석의견을 바탕으로 개선방향이 제시되어, 다소 저자들의 주관이 반영되었다는 한계가 있다. 따라서, 향후 연구를 통해 교통사고보고서의 활용목적 등을 교통사고 원인 분석, 자율주행자동차 실험 시나리오 개발 등으로 구분하여, 각 분야별 전문가들을 대상으로 설문조사를 진행함으로써, 교통사고보고서의 개선방향에 대한 객관적 근거를 마련할 필요가 있다. 또한, 자율주행 자동차 실험 시나리오에 필요한 자료의 경우 향후 관련 연구들을 진행하며 추가하는 것이 필요하다. 마지막

으로, 개선된 교통사고보고서 자료 기반 자율주행자동차 실험 시나리오와 기존 교통사고보고서 자료 기반의 시나리오와의 비교검증을 통해 교통사고보고서의 개선방향에 대한 검증이 가능할 것이다.

ACKNOWLEDGEMENTS

본 과제(결과물)은 교육부와 한국연구재단의 재원으로 지원을 받아 수행된 디지털 신기술 인재양성 혁신 공유대학사업의 연구 결과입니다.

REFERENCES

- An, H. and Lee, S.(2015), “The analysis of data structure to digital forensic of dashboard camera”, *Journal of the Korea Institute of Information Security & Cryptology*, vol. 25, no. 6, pp.1495-1502.
- Char, F. and Serre, T.(2020), “Analysis of pre-crash characteristics of passenger car to cyclist accidents for the development of advanced drivers assistance systems”, *Accident Analysis & Prevention*, vol. 136, 105408.
- Chen, T., Shi, X. and Wong, Y. D.(2019), “Key feature selection and risk prediction for lane-changing behaviors based on vehicles’ trajectory data”, *Accident Analysis & Prevention*, vol. 129, pp.156-169.
- Cryer, P. C., Westrup, S., Cook, A. C., Ashwell, V., Bridger, P. and Clarke, C.(2001), “Investigation of bias after data linkage of hospital admissions data to police road traffic crash reports”, *Injury Prevention*, vol. 7, no. 3, pp.234-241.
- Econovill, <http://www.econovill.com/news/articleView.html?idxno=352294>, 2021.11.18.
- Eo, H., Xu, G. and Lee, S.(2019), “Effective Multimodal Design for Safe Hand Over in Autonomous Vehicles: With a Focus on Avoiding Accident Scenarios”, *Journal of the Human Computer Interaction Society of Korea*, vol. 13, no. 3, pp.21-28.
- Ha, M. and Ahn, H.(2019), “A Machine Learning-Based Vocational Training Dropout prediction Model Considering Structured and Unstructured Data”, *The Journal of the Korea Contents Association*, vol. 19, no. 1, pp.1-15.
- Ha, O. K., Oh, J. T., Won, J. M. and Sung, N. M.(2005), “The study on the accident injury severity using ordered probit model”, *Journal of Korean Society of Transportation*, vol. 23, no. 4, pp.47-55.
- Ha, W. S. and Han, S. Y.(2003), “Establishment of Important Impact Parameters of Traffic Accident Reconstruction Program‘PC-CRASH’”, *Journal of Korean Society of Transportation*, vol. 21, no. 2, pp.155-164.
- Han, C. P.(2021), “Analysis of vehicle progress before and after a collision using simulation”, *Journal of the Korea Academia-Industrial Cooperation Society*, vol. 22, no. 1, pp.402-408.
- Hur, Y., Lee, C., Kim, G. and Lim, H.(2019), “Topic Automatic Extraction Model based on Unstructured Security Intelligence Report”, *Journal of the Korea Convergence Society*, vol. 10, no. 6, pp.33-39.

- Hwang, G. S., Choe, J. S., Kim, S. Y., Heo, T. Y., Jo, W. B. and Kim, Y. S.(2010), “Freeway Crash Frequency Model Development Based on the Road Section Segmentation by Using Vehicle Speeds”, *Journal of Korean Society of Transportation*, vol. 28, no. 2, pp.151-159.
- Jeong, H., Kim, H., Park, S., Han, E., Kim, K. H. and Yun, I.(2017), “Prediction of Severities of Rental Car Traffic Accidents using Naive Bayes Big Data Classifier”, *The Journal of The Korea Institute of Intelligent Transport Systems*, vol. 16, no. 4, pp.1-12.
- Jung, K. Y. and Bae, S. H.(2017), “Forecasting of Probability of Accident by Analyzing the Traffic Accident Data: Main Intersections on Arterial Roads in Busan”, *Korean Society of Civil Engineers Journal of Civil and Environmental Engineering Research*, vol. 37, no. 1, pp.111-117.
- Jung, W. J.(2019), “A Study on the Cause Analysis and Countermeasures of Traffic Accidents-Focused on Chungnam and Sejong Regions”, *Korea Community Welfare Policy Association*, vol. 30, no. 1, pp.143-164.
- Kang, J. G. and Lee, S.(2002), “Traffic accident prediction model by freeway geometric types”, *Journal of Korean Society of Transportation*, vol. 20, no. 4, pp.163-175.
- Kim, G., Lim, J., Park, I., Chun, Y. and Cho, C.(2013), “A Study on Traffic Accident Reconstruction through Vehicle Crash Test”, *Transactions of the Korean Society of Automotive Engineers*, vol. 21, no. 6, pp.58-63.
- Kim, J. H. and Kim, S. S.(2021), “A Study on the Analysis of Agricultural R&D Keywords Using Textmining Method”, *Journal of the Korea Academia-Industrial Cooperation Society*, vol. 22, no. 2, pp.721-732.
- Kim, J. U., Nam, G. M., Kim, J. H. and Lee, S. B.(2006), “Development of traffic accidents prediction model with fuzzy and neural network theory”, *Journal of Korean Society of Transportation*, vol. 24, no. 7, pp.81-90.
- Kim, S. H., Jang, J. A. and Choe, G. J.(2005), “A Hierarchical Approach for Diagnose of Safety Performance and Factor Identification for Black Spots (Black on Suwon-city)”, *Journal of Korean Society of Transportation*, vol. 23, no. 1, pp.9-20.
- Kim, Y. Y., Cho, K. H. and Kim, Y.(2020), “Analysis of risk factors for traffic accidents in Daegu area”, *The Korean Data & Information Science Society*, vol. 31, no. 3, pp.503-510.
- Korea Transport Institute(2017), *Estimation of Transport Accident Costs in 2015*, pp.62-67.
- Lee, H. R., Kum, K. J. and Son, S. N.(2011), “A study on the factor analysis by grade for highway traffic accident”, *International Journal of Highway Engineering*, vol. 13, no. 3, pp.157-165.
- Lee, J. Y., Chung, J. H. and Son, B. S.(2008), “Analysis of traffic accident severity for Korean highway using structural equations model”, *Journal of Korean Society of Transportation*, vol. 26, no. 2, pp.17-24.
- Lee, S. B., Han, D. H. and Lee, Y. I.(2015), “Development of freeway traffic incident clearance time prediction model by accident level”, *Journal of Korean Society of Transportation*, vol. 33, no. 5, pp.497-507.
- Lee, S. C. and Oh, J. S.(2006), “The understanding of traffic accidents’ patterns: The analysis of driver’s behavior and location at the moment of traffic accident”, *The Journal of the Korea Data Analysis Society*, vol. 8, no. 6, pp.2457-2471.
- Lee, S. H., Jeung, W. D. and Woo, Y. H.(2012), “Comparative Analysis of Elderly’s and Non-Elderly’s Human Traffic Accident Severity”, *The Journal of the Korea Institute of*

- Intelligent Transport Systems*, vol. 11, no. 6, pp.133-144.
- Nitsche, P., Thomas, P., Stuetz, R. and Welsh, R.(2017), “Pre-crash scenarios at road junctions: A clustering method for car crash data”, *Accident Analysis & Prevention*, vol. 107, pp.137-151.
- Park, S. M., Han, E., Hong, Y. S., So, J. H. and Yun, I. S.(2017), “Development of Safety Evaluation Scenarios for Autonomous Vehicles in Community Roads Using a Text Mining Technique”, *Proceedings of Korea Institute of Intelligent Transport Systems Fall Conference 2017*, pp.189-192.
- Park, S., Jeong, H., Kim, K. H. and Yun, I.(2018), “Development of safety evaluation scenario for autonomous vehicle take-over at expressways”, *The Journal of The Korea Institute of Intelligent Transport Systems*, vol. 17, no. 2, pp.142-151.
- Park, S., Park, S., Jeong, H., Yun, I. and So, J. J.(2021), “Scenario-mining for level 4 automated vehicle safety assessment from real accident situations in urban areas using a natural language process”, *Sensors*, vol. 21, no. 20, p.6929.
- Park, S., So, J. J., Ko, H., Jeong, H. and Yun, I.(2019), “Development of Safety Evaluation Scenarios for Autonomous Vehicle Tests Using 5-Layer Format (Case of the Community Road)”, *The Journal of the Korea Institute of Intelligent Transport Systems*, vol. 18, no. 2, pp.114-128.
- Rajendran, G., Arthanari, M. and Sivakumar, M.(2011), “GPS Tracking Simulation by Path Replaying”, *International Journal of Innovative Technology & Creative Engineering*, vol. 1, no. 1, pp.20-26.
- Scheurwegs, E., Luyckx, K., Luyten, L., Daelemans, W. and Van den Bulcke, T.(2016), “Data integration of structured and unstructured sources for assigning clinical codes to patient stays”, *Journal of the American Medical Informatics Association*, vol. 23, no. e1, pp.e11-e19.
- Shanthi, S. and Ramani, R. G.(2011), “Classification of vehicle collision patterns in road accidents using data mining algorithms”, *International Journal of Computer Applications*, vol. 35, no. 12, pp.30-37.
- Shi, X., Wong, Y. D., Li, M. Z. F. and Chai, C.(2018), “Key risk indicators for accident assessment conditioned on pre-crash vehicle trajectory”, *Accident Analysis & Prevention*, vol. 117, pp.346-356.
- Shim, J., Yu, J., Park, J. and Park, B.(2018), *Estimation of transport accident costs in 2016*, Sejong: The Korea Transport Institute, pp.29-54.
- Sohn, S. Y. and Shin, H. W.(2001), “Data mining for road traffic accident type classification”, *Ergonomics*, vol. 44, pp.107-117.
- Song, J. M.(2021), *A Study on Efficient Use of Traffic Accident Statistics Tables and Improvement Measures*, Master’s Thesis, Hanbat National University, pp.56-57.
- Tian, R., Yang, Z. and Zhang, M.(2010), “Method of road traffic accidents causes analysis based on data mining”, *In 2010 International Conference on Computational Intelligence and Software Engineering*, pp.1-4.
- Yoon, J. and Lee, S.(2018), “Analysis of Pedestrian Traffic Accident Factors around the Exclusive Median Bus Lane Station Area: Focused on TAAS (2014 - 2016) Data in Seoul”, *Journal of Korea Planning Association*, vol. 53, pp.123-142.
- Zhang, D., Yin, C., Zeng, J., Yuan, X. and Zhang, P.(2020), “Combining structured and unstructured data for predictive models: A deep learning approach”, *BMC Medical Informatics and Decision Making*, vol. 20, no. 1, pp.1-11.