

교통안전 빅데이터와 Transformer-LSTM을 활용한 시계열 교통류 예측 분석 및 Xgboost 기반 실시간 고속도로 화물운송경로 위험도 평가 방법론 개발

박동혁¹, 박준영^{2*}, 김덕녕³

¹한양대학교 ERICA 교통물류공학과,

²한양대학교 ERICA 교통물류공학과 및 스마트시티공학과, ³한국도로공사 도로교통연구원

A Methodology for Real-time Risk Assessment of Freeway Freight Routes using Transformer-LSTM Time-series Prediction with Traffic Safety Big Data

Donghyeok Park¹, Juneyoung Park^{2*}, Ducknyung Kim³

¹Department of Smart City Engineering, Hanyang University ERICA

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

³Korea Expressway Corporation Research Institute

Freight vehicle crashes with physical damage and human casualties can significantly affect lead times at the supply chain level, leading to decreased reliability and lower reorder rates, making it necessary to consider freight vehicle traffic accidents as a potential risk factor. This study aims to develop a real-time freight transportation route risk assessment methodology that can predict short-term traffic flows and detect crashes in advance based on factors that are complexly related to traffic safety through traffic, weather, and mobile data. First, based on real-time traffic data, Transformer-LSTM was used to predict time series traffic flow. Next, Xgboost based real-time crash prediction model for freight vehicles was developed. As a result of the analysis, a false positive rate of 5.23% was obtained, and it is judged to be effective for real-time risk assessment of cargo transportation routes. The results of the study can be used to provide a safe route guidance service for freight vehicle drivers in the future. Moreover, real-time crash risk warning services can be applicable to prevent freight vehicle crashes through cooperation with private navigation companies.

Keywords: Freight vehicle, Freight routes, Real-time risk assessment, Deep learning, Traffic data

논문접수일 : 2023.10.21. 논문수정일 : 2024.02.28. 게재확정일 : 2024.03.15.

이 논문은 한국도로공사 도로교통연구원의 “교통안전 빅데이터 실시간 분석 및 모니터링 기술 개발” 연구 영역의 지원과 과학기술정보통신부의 재원으로 한국연구재단의 지원을 받아 수행된 연구임(No. 2022R1A2C1093424).

1. 한양대학교 ERICA 교통물류공학과 박사과정

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

3. 한국도로공사 도로교통연구원 책임연구원

1. 서론

1.1 연구배경 및 목적

화물차는 승용차 대비 무거운 차체와 고중량의 화물이 있어 사고가 발생할 때 물리적 충격량이 크며 고속도로는 지방도나 국도와 같은 다른 위계의 도로보다 평균 통행 속도가 높기 때문에 고속도로에서 발생하는 화물차의 사고는 물리적 피해와 인명 피해가 크다(Kang et al., 2019). 화물차는 승용차와 주행행태가 서로 달라 도로 위에 공존할 경우 사고 개연성이 높아지기 때문에 화물차 지정차로제를 도입하여 교통류를 분리하고 있다. 화물차 통행량과 물동량은 시간이 지남에 따라 꾸준히 증가하고 있으며 특히, COVID-19 팬데믹으로 인한 국내 교통량이 감소하는 상황에서도 물류 산업이 활성화되어 이전보다 20% 증가한 것으로 나타났다(Park and Park, 2022).

경제적 효율성, 서비스의 품질 및 신뢰성은 화물 및 물류 시스템의 목표이지만 다른 차량과 도로 네트워크를 공유하는 화물차의 경우 안전은 중요한 사항이다(Hall and Mukherjee, 2008). 화물차 사고는 제품의 파손 뿐만 아니라 공급망 차원에서 리드 타임에 크게 영향을 미치기 때문에 잠재적인 피해 또한 고려할 필요가 있다. 공급망에서 고객이 제품을 수령하는데 소요되는 시간이 증가할 경우 고객과의 신뢰 저하와 재주문을 하락으로 이어져 회복하기 어렵기 때문에 화물차 교통사고를 잠재적 위험 요소로 판단해야 한다. 교통사고는 임의로 드물게 발생하지만 도로교통의 구성요소인 운전자, 차량, 도로, 환경 등이 교통안전과 복합적으로 연관되어 있어 서로 상충하는 상황이 발생할 때 사고로 이어질 가능성이 높아진다. 본 연구는 교통·기상·모바일 데이터를 통해 화물차 교통안전과 복합적으로 연관된 요소를 기반으로 단기 교통류를 예측하고 사고 징후 인자와 머신러닝을 활용하여 실시간 화물 운송 경로 위험도 평가 방법론을 개발하는 것을 목적으로 한다. 우선, 실시간 교통 데이터를 기반으로 시계열 예측에 우수한 성능을 보이는 Transformer-LSTM (Transformer-Long Short-Term Memory)을 활용하여 교통류의 변화를 예측하였다. 그 다음으로 Xgboost(eXtreme Gradient Boosting) 기반의 사고 예측모형을 개발하여 실시간 화물운송경로 위험도를 도출하였다. 연구의 결과물은 향후 화물차 운전자를 위한 안전경로안내 서비스를 제공하는 데 참고할 수 있다. 또한, 민간 내비게이션 서비스 업체와 협력을 통해 화물차 사고 예방을 위한 실시간 사고 위험 경고 서비스 제공에 활용할 수 있다.

본 연구는 다음과 같이 구성되어 있다. 서론의 다음 장에서는 시계열 분석 기반의 교통류 예측 연구와 실시간 사고 예측모형 개발 연구를 중심으로 기존 문헌을 고찰하고 기존 연구와의 차별성과 본 연구의 의의에 대해 다루었다. 본론에서는 Transformer-LSTM 기반의 시계열 단기 교통류 예측 방법론과 Xgboost 기반의 실시간 사고 예측 방법론에 대해 설명하고 수집된 데이터에 대해 서술하였다. 또한, 수집된 데이터에 대해 설명하고 연구 범위에 대해 다루었다. 마지막으로 분석된 결과를 제시하고 해석하였다. 결론에서는 본 연구의 결과를 요약하였으며 향후 연구계획에 대해 서술하였다.

1.2 기존 문헌 고찰

화물차 사고는 다른 사고 대비 인명 및 금전 피해가 높고 장기적으로 공급망 리스트 측면에서 신뢰성과 화물에 포함된 상품 가치의 손실로 이어지기 때문에 화물차를 중심으로 운송 경로에 대한 잠재적인 위험성을 평가하는 것은 중요하다. 기존 문헌에 따르면 실시간 사고 예측모형을 개발하여 교통안전 모니터링 연구가 진행된 사례는 있었으나 화물차를 대상으로 특화하여 실시간 사고 예측 분석을 수행한 연구는 부족하였다. 본 연구는 시계열 예측 분석 기법을 활용하여 단기 교통류를 예측하고 화물차를 대상으로 실시간 화물차 운송 경로 위험도 평가를 위한 사고 예측모형 개발을 목적으로 한다. 본 장에서는 시계열 분석 기반의 교통류 예측 연구와 실시간 사고예측모형 개발 관련 연구에 대해 문헌 고찰을 수행하였다.

1) 시계열 분석 기반의 교통류 예측 관련 문헌 고찰

Ma et al.(2015)은 MVDS에서 수집한 2분 단위 교통류 속도 데이터를 활용하여 LSTM-NN(LSTM-Neural Network) 기반의 단기 속도 예측 분석을 수행하였으며 비교 모델로 Elman-NN 등의 신경망 모델, ARMA(Autoregressive Moving Average)와 같은 시계열 모형, 비 모수적(Non-parametric) 방법론인 SVM(Support Vector Machine), 모수적(Parametric) 방법론인 Kalman-filter를 활용하였다. Wang et al.(2016)은 교통류 속도 예측을 위해 eRCNN(Recurrent Convolutional Neural Network) 구조를 이용한 딥러닝 방법을 제안하였다. 인접한 도로 구간의 시공간 교통 속도를 통합하고 첨두시간에 발생하는 반복적인 교통 지·정체로 인해 급작스럽게 발생하는 비반복적인 정체 상황을 예측 오류로 설정하여 모형의 성능을 개선하였다. Jia et al.(2016)은 교통류의 확률적 특성을 반영하기 위한 단기 속도 예측 방법론으로 딥러닝 기법인 DBN(Deep Belief Network) 모델을 제안하였으며 BPNN(Back-Propagation Neural Network)와 ARIMA(Autoregressive Integrated Moving Average) 모델을 비교군으로 활용하였다. Zhao et al.(2019)은 교통류의 평균 속도와 통행 시간 지연 변수를 기반으로 군집 분석을 통해 교통류 혼잡 상태를 구분하고 LSTM과 SVR(SVM regression)을 활용하여 혼잡 정도에 따른 교통류의 속도를 예측하였다.

2) 실시간 사고예측모형 관련 문헌 고찰

기존 연구에 따르면 고속도로 구간 내 교통류 상태 및 특성과 실시간 사고 위험 간 관계를 규명하기 위해 사고를 사전에 감지하는 지표를 개발하고 계량 경제학, 머신러닝, 딥러닝 등 다양한 분류(Classification) 및 예측 방법론을 제안하였다. Lee et al.(2002)은 고속도로에서 사고를 유발하는 교통류 특성에 대해 조사하여 사고가 발생하기 전에 관찰되는 지표인 사고 전구체(Crash precursors) 개념을 제안하였다. Abdel-Aty and Pemmanaboina(2006)는 실시간 교통류를 수집할 수 있는 ATMIS (Advanced Traffic Management and Information Systems)와 강수 정보를 활용하여 PCA(Principal Component Analysis)와 로지스틱 회귀모형을 통해 실시간 사고 가능성을 식별할 수 있는 예측모형을 개발하였다. 사고 발생 전 5-10분의 강수 상태, 하류부의 5분간 평균 점유율 및 교통량의 표준편차, 상류부의 5분간 속도 변동계수가 사고 발생에 가장 큰 영향을 끼치는 것으로 나타났다. Kwak and Kho(2016)는 고속도로의 다양한 구간 유형과 교통류 상태에 따른 실시간 사고 위험 예측 모델을 개발하기 위해 루프 검지기에서 수집한 교통류 데이터를 기반으로 시간 또는 공간 간격에 대한 매개변수인 교통량, 속도, 점유율을 평균, 표준편차, 변동계수, 변동률 등을 산출하여 교통류 특성과 사고 위험도 간의 관계를 추정하였다. 또한, 조건부 로지스틱 회귀모형과 유전자 프로그래밍을 활용하여 실시간 사고 위험을 예측하였으며 제안된 모형의 예측 성능을 평가하기 위해 ROC Curve(Receiver Operating Characteristics Curve)와 AUC(Area Under the ROC Curve)를 활용하였다. Wang et al.(2015)은 서로 다른 교통류 간의 상충으로 인한 병목현상이 자주 발생하는 고속도로 엇갈림구간을 대상으로 MVDS(Microwave Vehicle Detection System)를 통해 수집된 실시간 교통류 데이터, 기상 데이터 그리고 도로 기하 구조 정보를 기반으로 다중수준 베이지안 로지스틱 회귀모형을 제안하였다. 분석 결과 진입부와 진출부 교통류 사이에 높은 속도 차이는 후미 추돌 위험을 증가시키는 것을 통해 속도 차가 사고 위험을 추정하는데 중요한 변수임을 확인하였다. 최근에는 로지스틱 회귀모형 외 머신러닝 및 딥러닝을 활용한 실시간 사고 예측 연구가 다수 수행되었다. Yu and Abdel-Aty(2013)는 과적합을 방지하기 위해 CART(Classification and Regression Tree)를 활용하여 변수 중요도를 기반으로 변수를 선정하였으며 머신러닝 기법 중 하나인 SVM을 활용한 실시간 사고 위험 예측모형을 개발하였다. Parsa et al.(2020)은 교통, 인구통계학, 토지 이용 및 날씨 특성으로 구성된 실시간 데이터를 기반으로 예측 성능이 우수한 것으로 알려진 머신 러닝 방법론인 Xgboost를 활용하여 실시간 사고 예측 연구를 수행하였다. Li et al.(2020)은 교통류 특성, 신호 체계 및 날씨 조건과 같은 다양한 변수를 기반으로 2가지 딥러닝 기법을 결합한 LSTM-CNN(Long Short-Term Memory Convolutional Neural Network)을 활용하여 도시 간선 도

로를 대상으로 실시간 사고 위험 예측 모델을 제안하였다. 사고는 무작위적이고 드물게 발생하는 위험 상황에 대한 결과로 사고 상황과 일반 상황 간에 데이터 불균형이 발생하기 때문에 SMOTE(Synthetic Minority Over-Sampling Technique)를 활용하여 데이터 증강을 수행하였으며 예측 모델의 성능을 비교하기 위해 베이지안 로지스틱 회귀모형, Xgboost, LSTM, CNN, Sequential LSTM-CNN을 기준 모델로 제시하였다. Li et al.(2022)은 실시간 사고 위험 예측 분석을 수행하기 위해 교통류 특성과 차량 궤적 데이터를 융합하였다. 최근 ChatGPT(Chat Generative Pre-trained Transformer)를 통해 알려진 어텐션 메커니즘(Attention mechanism)을 적용한 TA-LSTM(Temporal Attention Based LSTM)으로 사고 위험 예측 모델을 개발하였다. 기존 문헌 고찰에 대한 내용은 Table 1에 제시하였다.

Table 1. Literature Review

| Subject | Author | Detailed |
|----------------------------|-----------------------------------|--|
| Traffic flow prediction | Ma et al.(2015) | <ul style="list-style-type: none"> Short-term traffic flow speed prediction study using MVDS data [Method] LSTM-NN, Elman-NN, ARMA, SVM, Kalman-filter |
| | Wang et al.(2016) | <ul style="list-style-type: none"> Research on improving traffic flow speed prediction performance by setting non-repetitive congestion situations as errors [Method] eRCNN |
| | Jia et al.(2016) | <ul style="list-style-type: none"> Short-term speed prediction study applying stochastic characteristics of traffic flow [Method] DBN, BPNN, ARIMA |
| | Zhao et al.(2019) | <ul style="list-style-type: none"> Research on predicting traffic flow speed according to the congestion status [Method] LSTM, SVR |
| Real-time crash prediction | Lee et al.(2002) | <ul style="list-style-type: none"> Introduction to the crash precursor concept |
| | Abdel-Aty and Pemmanaboina (2006) | <ul style="list-style-type: none"> Real-time traffic crash prediction based on traffic flow changes between upstream and downstream segments [Method] PCA-Logistic regression model |
| | Kwak and Kho(2016) | <ul style="list-style-type: none"> Development of loop detector-based traffic flow parameters and real-time crash risk prediction research [Method] genetic programming, Logistic regression model |
| | Wang et al.(2015) | <ul style="list-style-type: none"> Real-time crash prediction for freeway weaving segments using traffic flow, weather, and road geometry data [Method] Bayesian logistic regression model |
| | Yu and Abdel-Aty (2013) | <ul style="list-style-type: none"> Variable selection based on CART and real-time crash prediction using SVM [Method] CART, SVM |
| | Parsa et al.(2020) | <ul style="list-style-type: none"> Real-time crash prediction using traffic flow, demographic, land use, and weather condition data [Method] Xgboost |
| | Li et al.(2020) | <ul style="list-style-type: none"> Real-time crash risk prediction for arterial roads using data augmentation [Method] LSTM-CNN, Xgboost, Bayesian logistic regression model, LSTM, CNN, Sequential LSTM-CNN |
| | Li et al.(2022) | <ul style="list-style-type: none"> Real-time crash prediction combining vehicle trajectory data and traffic flow data [Method] Attention mechanism, TA-LSTM |

1.3 기존 연구와의 차별성 및 의의

화물차 사고는 직접적으로 물리적 피해가 발생하고 간접적으로 물류 시스템 내 신뢰성에도 부정적인 영향을 미친다. 화물차 사고는 물류 공급망 내 관리해야 할 리스크로써 교통사고를 예방하는 것이 중요하다. 통계적 이론 기반에서 딥러닝까지 다양한 방법론을 활용하여 단기 교통류를 예측하였으며 다양한 기법들을 통해 실시간 사고 예측 연구가 많이 수행되었으나 화물차 중심의 사고 예측모형의 개발 연구가 미비하다. 또한, 선행 연구에서는 현재 교통 상황을 기반으로 현재의 사고 위험을 판단하는 연구를 수행하였으나 미래 교통 상황을 기반으로 미래의 사고 위험을 예측하는 연구는 미흡하다. 본 연구는 최근 기법 중 하나인 Transformer-LSTM을 기반으로 실시간 단기 교통류 예측을 수행하였다. 또한 교통, 기상, 모바일 데이터를 활용하여 화물차 중심의 사고 예측모형을 개발하고 실시간 화물운송경로의 위험도 평가 방법론을 제시하였다.

2. 본론

2.1 연구 방법론

1) Framework

본 연구는 단기 교통류를 시계열 예측하는 모듈과 화물차 중심의 실시간 사고 위험을 예측하는 모듈로 구성되어 있다. 단기 교통류를 예측하기 위해 도로 위에 설치된 차량 검지기(VDS; Vehicle Detection System)를 기반으로 수집한 실시간 교통류 데이터를 활용하여 시계열 예측 기법인 LSTM을 통해 단기 교통류 속도, 교통량, 점유율 값을 얻는다. 실시간 사고 예측을 수행하기 위해 사고에 영향을 줄 수 있는 교통류 특성 데이터, 기상 데이터, 모바일 기반의 급감속 이벤트 데이터, 사고 이력 자료를 기반으로 데이터 셋을 구축하고 사고를 사전에 감지할 수 있는 교통류 파라미터를 산출하였다. 다음으로 머신러닝 기법 중 예측 성능이 뛰어난 것으로 알려진 Xgboost를 기반으로 화물차 중심 사고 예측 모델을 구축하였다. 마지막으로 단기 미래 교통류 데이터를 입력 값으로 활용하여 화물운송경로 위험도 평가를 위한 실시간 사고 위험 예측모형을 개발하였다. 연구의 수행과정은 Figure 1에 제시하였다.

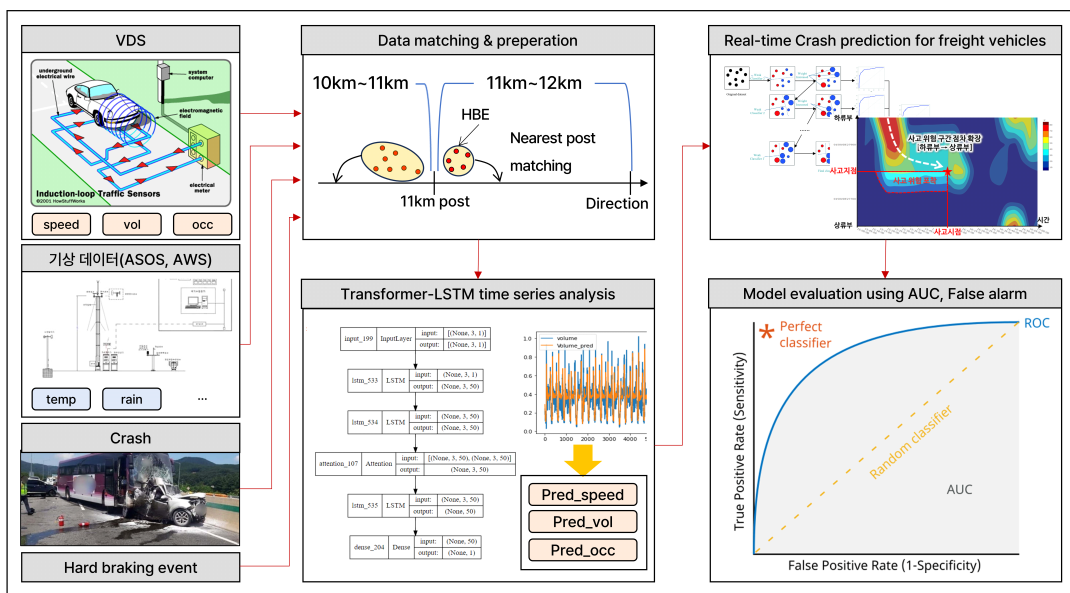


Figure 1. Research framework

2) Transformer-LSTM 시계열 예측모형

LSTM은 순환신경망(RNN; Recurrent Neural Network)에서 가중치를 업데이트하는 역전파 과정(Back-propagation)에서 손실(Loss)이 비정상적으로 감소하는 그래디언트 소실 문제(Gradient vanishing problem)를 해결하기 위해 1997년 Hochreiter와 Schmidhuber가 제안한 방법론이다(Hochreiter and Schmidhuber, 1997). 순환신경망 계열의 방법론에서 그래디언트 소실 문제는 시계열의 시퀀스가 길어질수록 과거의 정보를 제대로 전달하지 못하는 장기 의존성 문제(Long term dependency problem)를 유발한다. LSTM은 은닉층(Hidden layer)의 메모리 셀에 입력 게이트(Input gate), 삭제 게이트(Forget gate), 출력 게이트(Forget gate)를 추가하여 불필요한 정보 지우고, 전달할 정보를 결정한다. 이를 통해 장기기억과 단기 기억을 결정하여 셀 상태(Cell state)의 값으로 출력한다.

Transformer는 자연어 처리(NLP; Natural Language Processing) 분야에서 기존 인코더(Encoder)-디코더(Decoder) 기반 신경망 기계 번역(Translation) 시스템을 개선하기 위해 제안되었으며 번역 품질의 향상에 기여한 방법론이다. 기존 시계열 모델은 입력 시퀀스(Sequence)의 모든 정보를 처리하기 위해 고정된 크기의 벡터에 압축하기 때문에 정보의 손실이 발생하는 문제가 있다. 그러나, Transformer는 출력값을 예측하는 타임 스텝마다 전체 입력 값 대신 해당 타임 스텝에서 연관성이 높고 중요한 부분에 집중(Attention)하는 어텐션 매커니즘을 활용한다(Vaswani et al., 2017).

본 연구에서는 연관성이 높고 중요한 부분에 집중할 수 있는 Transformer와 그래디언트 소실 문제를 해결한 LSTM을 결합하여 단기 교통류 시계열 예측 분석을 수행하였다. 예측된 단기 교통류 값의 결과를 기반으로 정량적인 예측성능을 도출하여 비교하기 위해 MAPE(Mean Absolute Percentage Error)와 RMSE(Root Mean Squared Error)를 산출하였다. 관측값 y_i 와 예측값 \hat{y}_i 간의 오차를 기반으로 측정하는 지표로 수식 (1)과 (2)에 제시하였다.

$$MAPE = \frac{1}{N} \times \sum_{i=1}^n \left| \frac{y_i - \hat{y}_i}{y_i} \right| \quad (1)$$

$$RMSE = \sqrt{\sum_{i=1}^n \frac{(y_i - \hat{y}_i)^2}{N}} \quad (2)$$

3) Xgboost 기반 실시간 사고 예측모형

Xgboost는 하나로 약한 분류기 또는 회귀 모델의 예측 결과를 결합함으로써 단일 모델 대비 신뢰성 높은 예측값을 얻을 수 있는 앙상블 학습 모델 중 하나인 부스팅(Boosting) 기반의 머신러닝 모델이다(Chen and Guestrin, 2016). 또한, 데이터의 병렬처리를 통해 예측성능과 자원의 효율성이 우수한 것으로 알려져 있다(Parsa et al., 2020). 기존에 수행된 다수의 실시간 사고 예측 연구에서 Xgboost는 다른 방법론 대비 예측 성능이 뛰어나 비교 모델로 많이 활용되고 있다(Li et al., 2020; Li et al., 2023).

본 연구에서는 예측 성능이 뛰어난 Xgboost 방법론을 활용하여 실시간 화물차 사고 예측모형을 개발하였다. 예측된 모형의 성능을 비교하기 위해서 AUC curve(Area under the ROC curve)와 오경보율(False-alarm ratio)을 활용하였다. ROC curve는 모형의 분류 기준으로 활용되는 여러 임계 값 설정에 따른 위양성률에 대한 양성률을 의미한다. 양성률(TPR; True Positive Rate)은 재현율(Recall) 또는 민감도(Sensitivity)로도 불리며 실제 발생한 사고 중 모형이 사고로 정확히 예측한 비율을 의미하며 수식 (3)에 제시하였다. 오경보율이라고도 불리는 위양성률(FPR; False Positive Rate)은 실제로 발생하지 않은 사고 중 모형이 잘못 분류하여 사고로 예측한 비율을 의미하며 수식 (4)와 같이 전체 중 특이도(Specificity)간의 차로 산출한다.

$$TPR = sensitivity = \frac{TP}{TP + FN} \quad (3)$$

$$FPR = 1 - specificity = \frac{FP}{FP + TN} = False\ alarm\ rate \quad (4)$$

머신러닝을 활용한 실시간 사고 예측 모형은 입력 데이터를 기반으로 0~1 사이의 값을 예측 값(predict value)으로 도출하며 임계 값(threshold)을 기준으로 예측 값이 크거나 같을 경우 사고 발생(1), 작을 경우 사고 미발생(0)으로 분류한다. 즉, 임계 값을 낮게 조정할 경우 사고 발생으로 판단하는 경우가 많아지고 반대로, 높게 조정할 경우 사고 미발생으로 판단하는 경우가 많아진다. 임계 값이 지나치게 낮아질 경우 사고로 예측하는 빈도가 많아지지만 도로 이용자나 관리자 측면에서 잘못된 정보로 판단하여 신뢰성과 위험 정보에 대한 순응률이 감소한다. 반면, 임계 값이 지나치게 높아질 경우 실제로 사고 발생 확률이 높은 상황임에도 불구하고 사고 미발생으로 판단하여 위험 정보를 제공하지 못해 사전에 미리 대처하지 못할 수 있다. 따라서, 사고를 판단하는 기준인 임계 값을 적절하게 결정하는 것은 사고 예측에서 중요한 문제로 AUC와 오경보율을 예측 성능의 기준으로 설정하였다.

4) 시·공간 교통류 지표 산출

기존 연구에서는 사고 발생 이전에 포착할 수 있는 사고 전구체(Crash precursors) 개념을 활용하여 교통류 특성의 변화를 지표로 산출하여 사고 예측 연구에 활용하였다(Lee et al., 2002; Abdel-Aty and Pemmanaboina, 2006; Kwak and Kho, 2016; Wang et al., 2015; Yu and Abdel-Aty, 2013; Li et al., 2020; Li et al., 2022). 본 연구에서는 교통류 특성을 나타내는 VDS 데이터를 기반으로 시·공간적 교통류 특성 차이를 지표로 산출하여 사고 전조 현상을 포착하는 입력변수로 적용하였다. 시간적 교통류 차이 지표(TXD; Temporal traffic flow difference)는 동일한 구간 내 사고 발생 시간과 사고 5분 전 교통량, 점유율, 평균 속도 등 교통류 지표의 차이로 정의하였으며 수식 (5)에 산출식을 제시하였다.

$$TXD_i = Traffic_{ref} - Traffic_{before(i)} \quad (X: speed, volume, occupancy) \quad (5)$$

2.2 연구 범위와 자료

1) 연구 범위

경부고속도로는 대한민국을 대표하고 교통량과 사고 건수가 가장 많은 고속도로 중 하나로 구서 나들목부터 양재 나들목까지 416.1km 전 구간을 공간적 범위로 설정하였다. 또한, 중부내륙고속도로는 화물차 비율이 다른 고속도로 대비 높은 노선으로 내서 분기점에서 양평 나들목까지 301.7km 구간을 공간적 범위에 추가하였다. 본 연구의 공간적 범위는 Figure 2에 제시하였다. 본 연구는 2021년~2022년 동안 경부고속도로와 중부내륙고속도로에서 발생한 화물차 교통사고 1,196건, 급감속 위험 운전 이벤트 932,408건을 분석에 활용하였다. 연구에서 활용한 데이터에 대한 기술통계는 Table 2에 제시하였으며, 히스토그램은 Figure 3에 제시하였다. 데이터는 1km 단위로 집계하였으며 통합 데이터 셋을 구축하였다.

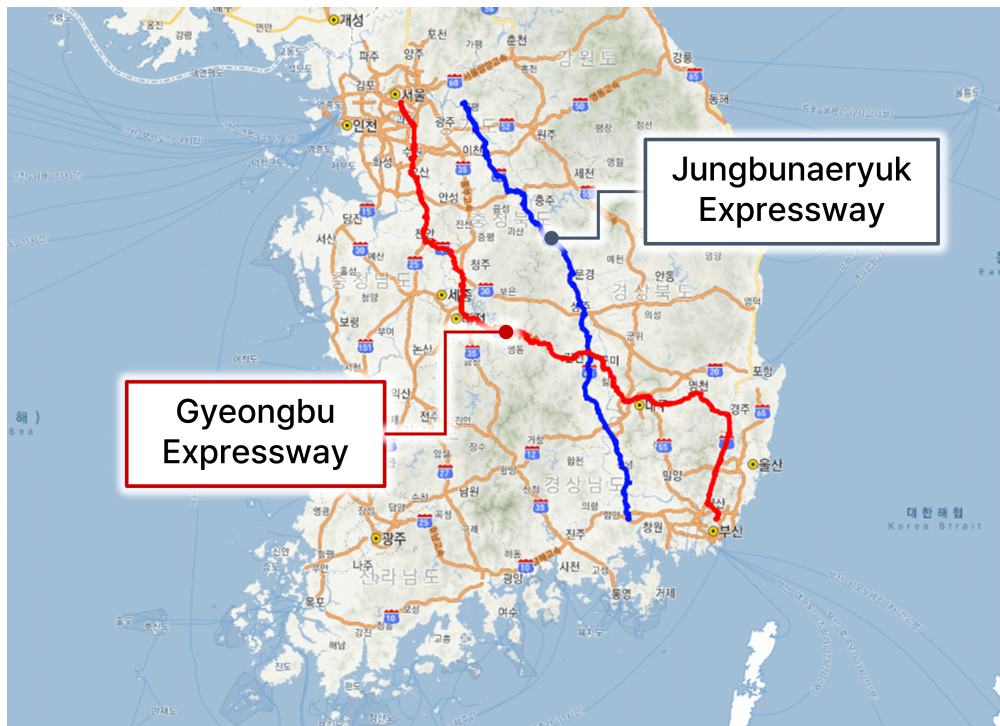


Figure 2. Research scope

Table 2. Data description

| Variable | Mean | 85 PCTL | S.D. | Min | Max |
|-------------------|--------|---------|--------|-------|--------|
| Volume(veh/5min) | 141.33 | 290 | 117.77 | 0 | 487 |
| Occupancy(%) | 4.01 | 7 | 3.15 | 0 | 49 |
| Speed(km/h) | 95.59 | 103.8 | 8.91 | 22.24 | 121.76 |
| Temperature(℃) | 14.10 | 24.6 | 10.38 | -14 | 32.9 |
| Precipitation(mm) | 0.29 | 0 | 1.17 | 0 | 14.5 |

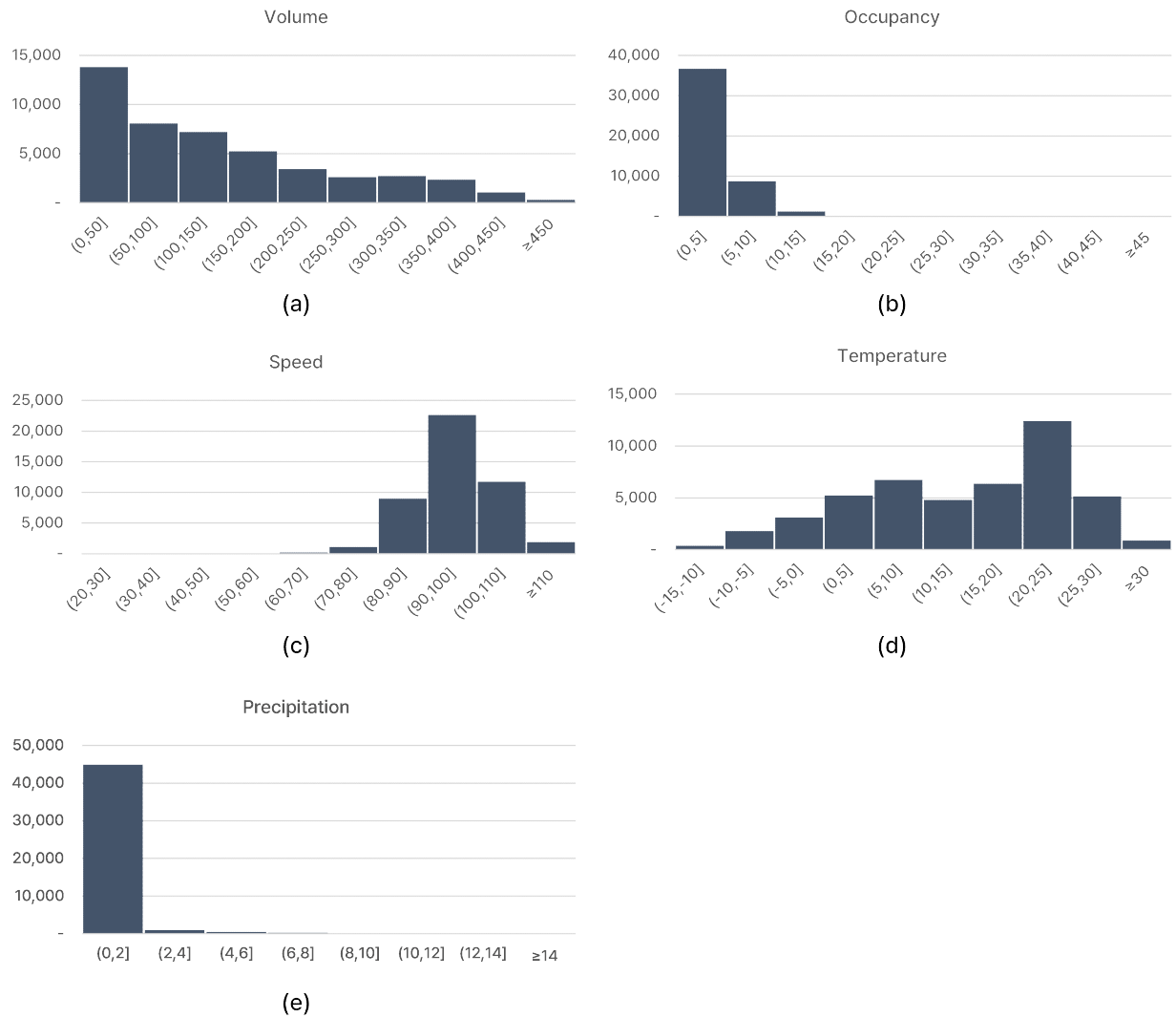


Figure 3. Histograms of traffic and weather data

2) 자료 수집 및 처리

본 연구는 단기 교통류 시계열 예측을 활용한 실시간 화물차 사고 예측 연구를 수행하기 위해 고속도로 사고 이력자료, 교통류 데이터, 기상 데이터 그리고 모바일 데이터를 수집하였다. 한국도로공사에서는 관할 고속도로를 대상으로 VDS를 설치 및 관리하고 있으며 30초, 5분 등 단위로 고속도로 공공데이터 포털을 통해 데이터를 제공하고 있다. 교통류 데이터는 지점 검지기에서 수집되는 VDS 원시자료로 본 연구에서는 5분 단위 평균속도, 교통량, 점유율을 활용하였다. 기상청에서 관리하고 있는 기상 관측 장비로는 자동기상 관측장비(Automatic Weather Station, AWS)와 종관기상관측시스템(Automated Synoptic observing System, ASOS)이 있다. AWS는 국지적인 기상현상 관측 데이터를 수집하는 장비로 전국 510개 지점에 설치되어 있으며 바람, 습도, 기온, 기압 등의 기상 요소를 1시간 단위로 수집할 수 있다. ASOS는 종관 규모의 날씨를 파악하기 위해 정해진 시각에 모든 관측소에서 동시에 실시하는 지상관측을 의미하는 종관기상 관측을 수행하는 장비로 전국 103개 지점에 설치되어 있으며 1시간 단위로 데이터 수집이 가능하다. 모바일 데이터는 SKT T-map에서 제공하고 있는 급감속 위험 운전 이벤트 데이터를 활용하였다. 급감속 위험 운전 이벤트는 차량이 3초간 40km/h 이상 감속하는 이벤트가 발생할 때 시간과 위치를 기록한다. 즉, 개별 차량 내 스마트폰 GPS 센서는 차량의 가속도가 약 -3.7m/s^2 이하일 때 위험 상황을 감지하고 SKT T-map 급정거 알림 서비스가 제공된다. 본 연구에서는 위경도를 기준으로 고속도로 구간에 가장 근접한 거리의 기상 관측 장비를 매칭시켜 사고 이력자료, 실시간 교통류 데이터, 날씨 데이터를 하나의 데이터 셋으로 구축하였다.

2.3 분석 결과

1) Transformer-LSTM 기반 교통류 시계열 예측 결과

본 연구는 VDS 데이터를 활용하여 단기 교통류 시계열 예측 분석을 수행하였으며 Transformer-LSTM과 LSTM 방법론의 시계열 예측 성능을 비교하였다. 예측된 속도, 교통량, 점유율 값은 실시간 화물차 사고 예측모형 개발에 활용되었다. 두 가지 시계열 예측모형에 15분 전 데이터를 기반으로 5분 이후 데이터를 예측하였다. 모형 학습을 위한 파라미터로 에포크는 20, 배치 크기는 32로 하였으며 학습 데이터, 검증 데이터, 테스트 데이터의 비율은 6:2:2로 설정하였다. Transformer-LSTM의 구조는 3개의 LSTM 층과 1개의 Attention 층을 반영하였으며 LSTM의 구조는 3개의 LSTM 층으로 구성하였다. 분석 결과, Transformer-LSTM 교통류 시계열 예측 모형이 LSTM 모형보다 오차 비율을 대표하는 MAPE 값이 낮아 예측성능이 더 우수한 것으로 나타났다. 오차의 크기를 의미하는 RMSE 기준에서도 Transformer-LSTM 모형이 비교 모형보다 교통량을 제외한 나머지 변수에 대해 예측성능이 뛰어나다. 단기 교통류 시계열 예측 결과는 Table 3에 제시하였다. 본 연구에서는 Transformer-LSTM을 기반으로 예측된 교통류 시계열 데이터를 반영하여 화물차 실시간 사고 예측 모형 개발에 활용하였다.

Table 3. Analysis results of short-term traffic flow time series forecast

| Performance measure | Transformer-LSTM | | | LSTM | | |
|------------------------|------------------|--------|-----------|-------|--------|-----------|
| | Speed | Volume | Occupancy | Speed | Volume | Occupancy |
| MAPE(%) | 2.731 | 14.951 | 17.878 | 2.946 | 15.953 | 20.131 |
| RMSE | 3.134 | 14.346 | 0.825 | 3.318 | 14.001 | 0.924 |

2) Xgboost 기반 실시간 사고 예측 결과

본 연구에서는 단기 교통류 시계열 예측 값을 활용한 실시간 사고위험 예측모형을 개발하였다. 학습 데이터, 검증 데이터, 테스트 데이터의 비율은 6:2:2로 설정하였으며 파라미터 최적화를 위해 가우시안 프로세스를 활용한 베이지안 최적화를 수행하였다. Xgboost 파라미터 중 learning rate, max depth, subsample 등을 최적화 대상으로 선정하였으며 과적합(Over-fitting)을 방지하기 위해 정칙화 기법을 활용하였다. 모형의 임계값(Threshold)은 안정적인 예측 결과를 도출하기 위해 TPR과 FPR이 교차하는 지점으로 선정하였다. 최적 임계값을 적용한 현재 상태의 교통류를 기준으로 한 모형의 위양성률은 11.05%가 도출되었으며 단기 미래 예측 상태의 교통류를 기준으로 한 모형의 위양성률이 5.23%가 도출되었다. 모형의 성능을 비교하기 위해 도출한 AUC의 경우 단기 미래 예측 상태의 교통류를 기준으로 한 모형은 0.9444, 비교 모형은 0.9154가 도출되었다. 각 모형의 ROC curve는 Figure 4(a)와 Figure 4(b)에 제시하였다.

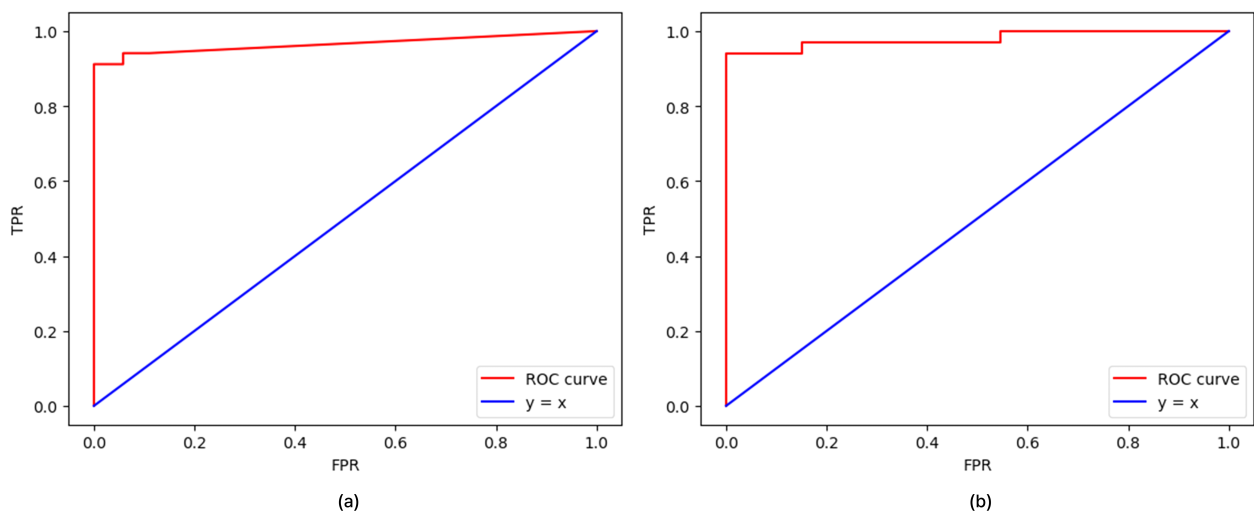


Figure 4. Analysis result of real time crash prediction model for freight vehicles

3. 결론

본 연구는 교통·기상·모바일 데이터를 통해 화물차 교통안전과 복합적으로 연관된 요소를 기반으로 단기 교통류를 예측하고 사전에 사고를 포착할 수 있는 실시간 화물 운송 경로 위험도 평가 방법론을 개발하였다. 먼저, 실시간 교통 데이터를 기반으로 시계열 예측에 우수한 성능을 보이는 Transformer-LSTM을 활용하여 시계열 교통류를 예측하였다. 분석 결과, Transformer-LSTM 기반 시계열 예측 모형이 비교 모형 대비 성능이 뛰어난 것으로 도출되었다. 예측된 교통류 데이터를 기반으로 Xgboost(eXtreme Gradient Boosting) 기반의 화물차 사고 예측모형을 개발하였다. 분석 결과 단기 미래 예측 상태의 교통류를 기반 예측 모형의 오경보율은 5.23%, AUC는 0.9444가 도출되었으며 비교 모형의 오경보율은 11.05%, AUC는 0.9154가 산출되었다. AUC가 높을수록 사고 발생/미발생을 구분하는 성능이 우수한 것을 의미하며, 이는 단기 교통류 예측 값을 활용한 실시간 사고 예측 모형이 비교 모형 대비 분류 성능이 우수한 것을 의미한다. 본 연구를 통해 제안한 모형이 비교 모형보다 오경보율이 낮게 도출된 것은 잘못된 정보를 제공하여 도로 이용자나 관리자 측면에서 신뢰성과 위험 정보에 대한 순응률이 감소할 우려가 적다는 것을 의미한다. 본 연구의 결과로 미래의 교통 상황을 기반으로 미래의 사고 위험 상태를 예측하는 모형을 개발한 것으로 기존 모형 대비 우수한 성능을 도출하여 실시간 화물차 교통사고 모니터링 기술에 대한 응용 가능성

을 확인하였다. 향후 화물차 운전자를 위한 안전경로안내 서비스를 제공하는 데 활용이 가능하며, 민간 내비게이션 서비스 업체와 협력을 통해 화물차 사고 예방을 위한 실시간 사고위험 경고 서비스로 발전이 가능하다.

향후 연구에서는 단기 교통류를 예측하기 위해 활용한 시계열 데이터가 적절한 방법론인지 검증하는 절차를 추가하여 모델 선택의 당위성을 확보할 필요가 있다. 또한, 시계열 교통류 예측 모델인 Transformer-LSTM의 구조를 결정하는 파라미터를 최적화하여 모형을 고도화하고 예측 성능을 개선할 필요가 있다. 실시간 사고 위험을 예측하기 위해 활용하는 모형에 대해 적절성을 판단하기 위해 Xgboost 외 머신러닝 기법에 해당하는 Random Forest, Support Vector Machine과 딥러닝 기법인 DNN(Deep Neural Network) 등 다양한 방법론을 적용하여 비교할 필요가 있다. 본 연구에서 활용한 입력 변수에 대해 향후 민감도 분석을 수행하여 변수의 적정성을 고려해야 한다. 또한, 머신러닝 및 딥러닝은 입력 변수가 모형에 미치는 영향에 대해 설명이 어려운 것으로 알려져 있으나 SHAP(Shapley Additive exPlanations), LIME(Local Interpretable Model-agnostic Explanation) 등 설명가능한 AI (XAI; eXplainable AI) 방법론을 통해 보완이 가능하다. 따라서, 교통류, 기상, 모바일 데이터를 기반으로 수집된 다양한 입력 변수에 대해 XAI 기법을 활용하여 입력 데이터 별 모형에 미치는 영향력의 크기, 방향성, 변수 간 의존성에 대해 확인할 필요가 있다. 터널, 교량, 진·출입부, 엇갈림구간 등은 도로 시설물 및 기하구조 요인으로서 교통사고 발생에 영향을 미칠 수 있다. 터널의 경우 음영 지역에 노면 결빙이 생성되어 화물차량의 제동력에 영향을 줄 수 있으며, 교량은 강풍으로 인해 화물차의 주행 안전성 저하를 유발할 수 있다. 또한, 진·출입부 및 엇갈림구간은 차량 별 주행 방향이 서로 달라 상충(conflict)이 다수 발생하는 지역이다. 따라서, 향후 연구에서는 도로 기하구조 요인을 입력 변수로 활용할 필요가 있다. 본 연구를 통해 제시된 사고위험 예측모형을 통해 1km/5분 단위 실시간 사고위험도를 산출하고 평균, 백분위수, 임계값을 활용한 면적 산출 등의 집계 방법과 계절/월/요일/시간대 등 시간적 특성을 고려하여 시·공간적 특성을 고려한 사고위험구간 식별 연구로 확장할 필요가 있다.

참고문헌

- Abdel-Aty, M., Pemmanaboina, R.(2006), Calibrating a real-time traffic crash-prediction model using archived weather and ITS traffic data, *IEEE Transactions on Intelligent Transportation Systems*, 7(2), 167-174
- Chen, T., Guestrin, C.(2016), XGBoost: A scalable tree boosting system, *arXiv 2016*, arXiv preprint arXiv:1603.02754, 11.
- Hall, R. W., Mukherjee, A.(2008). Bounds on effectiveness of driver hours-of-service regulations for freight motor carriers, *Transportation research part E: logistics and transportation review*, 44(2), 298-312
- Hochreiter S., Schmidhuber J.(1997), Long short-term memory, *Neural computation*, 9(8), Institute of Electrical and Electronics Engineers Inc., 1735-1780
- Jia, Y., Wu, J., Du, Y.(2016, November), Traffic speed prediction using deep learning method, In 2016 IEEE 19th international conference on intelligent transportation systems (ITSC), (pp. 1217-1222). IEEE
- Kang, C., Chung, Y., Chang, Y. J.(2019). Injury severity analysis of truck-involved crashes on Korean freeway systems using an ordered probit model, *KSCE Journal of Civil and Environmental Engineering Research*, 39(3), 391-398
- Kwak, H., Kho, S.(2016), Predicting crash risk and identifying crash precursors on Korean expressways using loop detector data, *Accident Analysis & Prevention*, 88, 9-19
- Lee, C., Saccomanno, F., Hellinga, B.(2002), Analysis of crash precursors on instrumented freeways, *Transportation Research Record*, 1784(1), 1-8
- Li, P., Abdel-Aty, M., Yuan, J.(2020), Real-time crash risk prediction on arterials based on LSTM-CNN, *Accident Analysis & Prevention*, 135, 105371
- Li, P., Abdel-Aty, M.(2022), Real-time crash likelihood prediction using temporal attention-based deep learning and trajectory fusion, *Journal of transportation engineering, Part A: Systems*, 148(7), 04022043
- Ma, X., Tao, Z., Wang, Y., Yu, H., Wang, Y.(2015), Long short-term memory neural network for traffic speed prediction using remote microwave sensor data, *Transportation Research Part C: Emerging Technologies*, 54, 187-197
- Park, J., Kim, T., Yu, D.(2007), Correlation analysis and estimation modeling between road environmental factors and traffic accidents (The Case of a 4-legged Signalized Intersections in Cheongju), *Journal of Korean Society of Transportation*, 25(2), 63-72
- Park, S., Park, J.(2022), Multilevel Mixed-Effects Models to Identify Contributing Factors on Freight Vehicle Crash Severity, *Sustainability*, 14(19), 11804
- Parsa, A. B., Movahedi, A., Taghipour, H., Derrible, S., Mohammadian, A. K.(2020), Toward safer highways, application of XGBoost and SHAP for real-time accident detection and feature analysis, *Accident Analysis & Prevention*, 136, 105405
- Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., Kaiser, Ł., Polosukhin, I.(2017), Attention is all you need, *Advances in neural information processing systems*, 30

- Wang, J., Gu, Q., Wu, J., Liu, G., Xiong, Z.(2016, December), Traffic speed prediction and congestion source exploration: A deep learning method, In 2016 IEEE 16th international conference on data mining (ICDM), (pp. 499-508). IEEE
- 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, 1-10
- Yu, R., Abdel-Aty, M.(2013), Utilizing support vector machine in real-time crash risk evaluation, Accident Analysis & Prevention, 51, 252-259
- Zhao, J., Gao, Y., Bai, Z., Wang, H., Lu, S.(2019), Traffic speed prediction under non-recurrent congestion: Based on LSTM method and BeiDou navigation satellite system data, IEEE Intelligent Transportation Systems Magazine, 11(2), 70-81

요약문

물리적 피해와 인명 피해가 큰 화물차 사고는 공급망 차원에서 리드 타임에 크게 영향을 미쳐 신뢰도 저하와 재주문을 하락으로 이어져 화물차 교통사고를 잠재적 위험 요소로 판단해야 한다. 본 연구는 교통·기상·모바일 데이터를 통해 교통안전과 복합적으로 연관된 요소를 기반으로 단기 교통류를 예측하고 사전에 사고를 포착할 수 있는 실시간 화물 운송 경로 위험도 평가 방법론을 개발하는 것을 목적으로 한다. 우선, 실시간 교통 데이터를 기반으로 Transformer-LSTM을 활용하여 시계열 교통류를 예측하였다. 다음으로 Xgboost 기반의 화물차 사고 예측모형을 개발하였다. 분석 결과 위양성률이 5.23%로 도출되어 실시간 화물운송경로 위험도 평가에 효과적일 것으로 판단된다. 연구의 결과물은 향후 화물차 운전자들을 위한 안전경로안내 서비스를 제공할 수 있으며 민간 내비게이션 업체와 협력을 통해 화물차 사고 예방을 위한 실시간 사고위험 경고 서비스 제공에 활용할 수 있다.

주제어: 화물차, 화물운송경로, 실시간사고위험평가, 딥러닝, 교통데이터