

## Driver Group Clustering Technique and Risk Estimation Method for Traffic Accident Prevention

Tae-Wook Kim\*, Ji-Woong Yang\*\*, Hyeon-Jin Jung\*\*\*, Han-Jin Lee\*\*\*, Ellen J. Hong\*\*\*\*

\*Student, Division of Software, Yonsei University, Wonju, Korea

\*\*Student, Dept. of Artificial Intelligence Semiconductor Engineering, Hanyang University, Seoul, Korea

\*\*\*Student, Dept. of Computer Science, Yonsei University, Wonju, Korea

\*\*\*\*Professor, Division of Software, Yonsei University, Wonju, Korea

### [Abstract]

Traffic accidents are not only a threat to human lives but also pose significant societal costs. Recently, research has been conducted to address the issue of traffic accidents by predicting the risk using deep learning technology and spatiotemporal information of roads. However, while traffic accidents are influenced not only by the spatiotemporal information of roads but also by human factors, research on the latter has been relatively less active. This paper analyzes driver groups and characteristics by applying clustering techniques to a traffic accident dataset and proposes and applies a method to calculate the Risk Level for each driver group and characteristic. In this process, the preprocessing technique suggested in this paper demonstrates a higher Silhouette Score of 0.255 compared to the commonly used One-Hot Embedding & Min-Max Scaling techniques, indicating its suitability as a preprocessing method.

▶ **Key words:** Traffic accident, Clustering, Risk level, Embedding, Preprocessing

### [요 약]

교통사고는 인간의 생명뿐만 아니라 사회적으로 큰 비용을 발생시키는 문제이다. 최근에는 교통사고 문제를 해결하기 위하여, 딥러닝 기술과 도로의 시공간적 정보를 통해 교통사고 위험도를 예측하는 연구가 진행되었다. 그러나 교통사고는 도로의 시공간적 정보뿐만 아니라 인적요소 또한 교통사고에 매우 큰 영향을 미치지만 이에 대한 연구는 상대적으로 활성화되지 않았다. 본 논문은 교통사고 데이터셋을 바탕으로 클러스터링 기법을 적용하여 운전자 그룹 및 특성을 분석하였으며, 각 운전자 그룹 및 특성에 대한 위험도를 산출하는 방법을 제시 및 적용하였다. 이 과정에서 본 논문에서 제시한 전처리 기법이 기존에 일반적으로 사용되었던 원-핫 임베딩, Min-Max Scaling 기법보다 더 높은 성능을 보임으로써 더 적합한 전처리 기법임을 보였다.

▶ **주제어:** 교통사고, 클러스터링, 위험도, 임베딩, 전처리

- 
- First Author: Tae-Wook Kim, Corresponding Author: Ellen J. Hong
  - \*Tae-Wook Kim (xodnr8436@yonsei.ac.kr), Division of Software, Yonsei University
  - \*\*Ji-Woong Yang (jiwoong0412@hanyang.ac.kr), Dept. of Artificial Intelligence Semiconductor Engineering, Hanyang University
  - \*\*\*Hyeon-Jin Jung (wjd5985@yonsei.ac.kr), Dept. of Computer Science, Yonsei University
  - \*\*\*Han-Jin Lee (han-.jin@yonsei.ac.kr), Dept. of Computer Science, Yonsei University
  - \*\*\*\*Ellen J. Hong (ellenhong@yonsei.ac.kr), Division of Software, Yonsei University
  - Received: 2024. 06. 11, Revised: 2024. 07. 19, Accepted: 2024. 07. 26.

## I. Introduction

교통사고는 인간의 안전과 생명에 직결되었을 뿐만 아니라, 도로교통공단 보도자료[1]에 따르면 2021년 도로교통사고로 인한 사회적 비용은 약 26조 9,987억 원이 발생할 정도로 국가에 사회적, 경제적으로 큰 손실을 초래하는 문제이다.

최근에는 이러한 교통사고 문제를 해결하기 위하여 딥러닝 기술을 기반으로 사고위험도를 예측하는 연구가 주로 수행되었다. 이 과정에서 교통 속도나 교통량 등의 시공간적 요인과 날씨 등의 환경적 요인을 통해 교통사고 위험도를 예측하여 교통사고를 예방하고자 하였다.

그러나 교통사고분석시스템(TAAS) 2022년 통계 보고서[2]에 따르면 2021년 1년 동안 발생한 교통사고에 대하여 가해 운전자의 법규위반 중 55%가 안전 운전 불이행에 해당할 정도로 인적요소가 교통사고에 매우 큰 영향을 미침에도 불구하고 인적요소를 활용한 사고위험도 관련 연구가 상대적으로 활발하게 수행되지 않았다.

본 논문에서는 운전자 특성에 기반한 교통사고 위험도를 산출하기 위해 클러스터링을 통해 유사한 특성을 가진 여러 운전자 그룹을 형성하고, 각 운전자 그룹이 지닌 특징을 바탕으로 그룹별 교통사고 위험도 산출 방법을 제시하고자 한다.

## II. Related Works

운전습관을 차량 운행 데이터를 바탕으로 클러스터링을 적용한 연구[3]는 차량의 가속계 및 위치 정보로 구성된 NGSIM 데이터셋을 바탕으로 횡방향 안정성 지표, 종방향 안정성 지표 등 4개의 지표를 추출하여 K-Means 클러스터링을 통해 4개의 위험적, 공격적, 안정적, 보수적 운전자 그룹으로 나누어 운전자 특성을 분석하였다. 트럭 4,357대의 GPS 기록, 수하물 정보, 차량 내부 모니터링 데이터를 기반으로 위험 운전 성향을 분석한 연구[4]는 K-Means 클러스터링과 주성분 분석을 바탕으로 2가지 수하물 조건에 따른 6개의 운전자 클러스터를 형성하였다. 이후, 다변량 분산 분석과 분산 분석을 통해 운전자 성향과 위험도 간의 관계를 파악하였다. 차량 내 가속계 센서의 시계열 정보를 클러스터링하여 운전자 특성을 파악한[5]는 Dynamic Time Warping과 Hidden Markov Model을 사용하여 시계열 데이터를 클러스터링하여 운전자 습관을 분석하였다. 스마트폰 센서 정보를 바탕으로 운전자 습관

을 분석한 [6]은 K-Means 클러스터링 알고리즘을 바탕으로 3개의 운전자 습관을 형성하였고, 각 운전자 습관의 가속 범위에 따라서 각 운전자 습관에 대한 위험 정도를 분석하였다. 여러 특성 선택 방법과 운전 습관 분석을 위해 클러스터링 방법을 적용하여 운전자 특성을 분석한 연구 [7]는 18가지의 특성 선택 방법을 도입하고 속도의 표준편차, 종방향 가속도 등으로 구성된 72개의 특성 후보를 생성하였다. 이러한 주요 특징을 바탕으로 클러스터링을 통해 3개의 초보 운전자, 경험 있는 신중한 운전자와 경험 있는 무모한 운전자 그룹으로 나누어 운전자 특성을 분석하였다. 더불어 운전 행동 설문지(DBQ)를 바탕으로 운전자 특성 분석 결과를 평가하여 제시한 방법이 효과적임을 보였다. 차량 내 센서를 바탕으로 공격적인 운전 행동을 식별하기 위한 프레임워크를 제시한 연구[8]는 국내 대도시에서 운행된 43대의 차량 운전 기록 데이터에서 의미있는 이벤트 정보를 추출하여 급격한 변화를 감지하고 오토 인코더를 통해 특징을 추출한 뒤, 두 개의 클러스터링 알고리즘을 통해 공격적인 운전의 높은 잠재성을 가진 클러스터를 식별할 수 있었다. 이로써 발견된 클러스터는 운전 행동의 표준을 나타내며 다른 운전자의 행동을 진단하는데 참조자료로 활용될 수 있음을 시사하였다. 이 밖에도, 버스 운전 GPS 기록을 사용하여 Fuzzy C-Means 클러스터링 알고리즘을 통해 버스 운전자의 위험 정도에 대한 분석[9], 전기 버스의 속도, 엔진 온도, 브레이크 및 엑셀 페달 작동 등의 데이터를 바탕으로 K-Means 클러스터링과 DBSCAN 기반의 버스 운전자 습관 분석[10], 나이와 차량 가속 관련 정보들을 바탕으로 K-Means 클러스터링과 계층적 클러스터링을 도입하여 운전자의 특징을 분석하고 실시간 위험도를 산출하기 위한 프레임워크[11], 차량 가속 정보 기반 운전자의 차선 변경 특징[12] 등의 연구가 수행되었다.

위 관련 연구들은 차량의 가속계 센서로부터 추출된 정보를 바탕으로 운전자의 공격적인 특성을 감지하거나 클러스터링을 통해 운전자 특성을 분석하였다. 그러나 사용된 차량 가속계 데이터셋들은 사고와 직접적인 관계가 없으므로 교통사고 위험도를 산출하는 것은 매우 어려운 문제이다. 본 논문은 교통사고 당시 가해자와 피해자의 인적 사항 및 주변 요소들이 포함된 사고 데이터셋과 클러스터링 기법을 활용하여 운전자 그룹 및 특성에 따른 교통사고 위험도를 산출하고자 한다.

### III. Dataset Construction & Preprocessing

본 논문에서는 운전자 특성을 분석하기 위하여 TAAS에서 제공하는 서울특별시 2021년 교통사고 데이터셋[13]을 사용한다. 해당 데이터셋은 사고일시, 사고유형, 법규위반, 도로 형태, 가해자 및 피해자 차종, 성별, 상해 정도 등 교통사고 정보에 대한 다양한 특성으로 구성된다.

데이터셋의 모든 특성을 클러스터링에 사용할 경우, 성능이 저하될 수 있으므로 다음과 같이 운전자 특성 분석에 유용한 특성을 추출하였다. 사고번호, 시군구, 요일, 부상 신고자 수, 노면 상태, 기상 상태와 피해 운전자 관련 특성은 교통사고를 유발하는 가해 운전자의 특성을 분석하는 과정에서 불필요한 요소라고 판단되어 제거하였다.

사고유형, 사고내용, 운전자 상해 정도 법규위반의 경우, 본 논문이 제시하는 클러스터링과 위험도 산출 방법을 실제에 적용할 때 사전에 알 수 없는 특성이므로 클러스터링의 입력으로 사용될 수 없어 제거하였다. 사망자 수, 중상자 수, 경상자 수도 마찬가지로 사전에 알 수 없는 정보이므로 제거하였으나, 클러스터링 이후 위험도 산출 과정에서는 클러스터링 기법을 통해 형성된 각 운전자 그룹별 사망, 중상, 경상에 대한 규모를 측정 및 활용하기 위해 사용되었다.

사고일시의 경우, TAAS 교통사고 통계 보고서[2]에 따르면 시간이라는 요소 또한 교통사고에 영향을 미치는 중요한 요소이나 운전자의 특성이 시, 분 단위로 변화하지 않는다고 판단하여 월(Month)만 특성으로 선택하였다.

최종적으로 클러스터링에 사용되는 특성은 Month, 도로 형태, 가해 운전자 차종, 가해 운전자 성별, 가해 운전자 연령이며 위험도 수준 산출 과정에서 사망자 수, 중상자 수, 경상자 수를 추가적으로 사용하였다.

## IV. Experiments

### 1. Preprocessing based on Statistical Data

클러스터링 알고리즘은 데이터셋의 전처리 과정에 따라 성능에 큰 영향을 받을 수 있으므로 적합한 전처리 방법을 선택하는 것은 매우 중요한 문제이다. 3장에서 구축된 데이터셋의 5가지 특성 중 Month, 도로 형태, 가해 운전자 차종, 가해 운전자 성별은 범주형 특성에 해당된다. 클러스터링 기법에 범주형 특성을 사용하기 위해선 원-핫 임베딩을 적용하여 처리하는 것이 일반적이다. 그러나 원-핫

임베딩은 데이터의 차원 수를 증가시켜 클러스터링 과정에서 좋지 않은 성능을 유발할 수 있다. 또한, 원-핫 임베딩 처리된 특성들은 0 또는 1의 값을 가지지만 연령과 같은 특성은 상대적으로 매우 큰 값을 가지기 때문에 특성 간 거리를 기반으로 작동되는 클러스터링 알고리즘은 연령을 기준으로 그룹을 나눠 형성될 가능성이 크며 이를 해결하기 위해 Min-Max Scaling을 적용할 경우 상대적으로 연령의 중요도가 무시되어 그룹들이 형성될 가능성이 존재한다. 따라서 각 특성들이 먼저 동일한 단위 혹은 관점으로 변환되어야 할 필요가 있다.

본 논문에서는 TAAS 2022년 교통사고 통계 보고서[2]를 기반으로 전처리 과정을 제안하고자 한다. 이는 데이터셋 내의 모든 특성에 대하여 각 특성의 값을 통계자료의 사고 건수 기준 구성비로 임베딩 하는 것이다. 가령, 데이터셋 내 특정 샘플의 연령이 21세라면 통계자료에 따라 21-30세의 사고 건수 구성비인 3.4로 임베딩된다. 이를 통해 각 특성들의 값이 동일한 단위와 범위를 가지게 되며, 클러스터링 알고리즘은 모든 특성을 구성비라는 하나로 공통된 위험도 관점에서 그룹을 형성할 수 있다. 이후 임베딩된 데이터셋에 대해서 구성비의 최대값인 100으로 나누어 Scaling을 진행한다.

본 논문에서 제시한 통계자료 기반 전처리 기법과 일반적인 원-핫 임베딩 및 Min-Max Scaling 기법을 비교하기 위해, 각 전처리 기법이 적용된 2개의 데이터셋에 대해서 클러스터링 알고리즘인 K-Means++를 바탕으로 클러스터링을 진행하여 그 성능을 비교한다.

### 2. Clustering Algorithm Performance Comparison

클러스터링 결과는 사용되는 알고리즘에 따라 성능과 결과가 상이하므로, 더 좋은 성능을 위해 다양한 알고리즘 중에서 가장 성능이 뛰어난 알고리즘을 선택할 필요가 있다. 본 논문에서는 기계 학습 Python Library인 Scikit-Learn[14]에서 제공되는 클러스터링 알고리즘 7개를 사용하여 성능 비교를 수행한다. 실험은 Ubuntu 20.04.6 LTS에서 진행하였다. 성능 비교를 위한 평가지표로 실루엣 스코어를 사용한다. 이때 실루엣 스코어는 각 데이터와 주변 데이터 간의 거리 계산을 바탕으로 군집 내 데이터들 간 비유사성이 작고 군집 간 비유사성이 클수록 높은 값을 가지는 클러스터링 알고리즘 평가지표이다. 또한, 각 클러스터링 알고리즘의 성능은 하이퍼파라미터에 따라 달라지므로, Table 1과 같은 Search Space를 바탕으로 가장 성능이 뛰어난 하이퍼파라미터를 선택하여 성능 비교를 수행한다. 이후, 클러스터링 결과를 T-SNE[15]를 통해 시각화하여 분석한다.

Table 1. Hyper-parameter Search Space

Algorithm	Search Space
K-Means++	n_cluster={2, 3, ..., 20}, init='k-means++', __n_init=50, max_iter=600
DBSCAN	min_samples={1000, 2000, ..., 10000}, epsilon={0.1, 0.2, ..., 0.5}
GaussMixture	n_compnents={2, 3, ..., 20}, __n_init=50, max_iter=600
BIRCH	n_cluster={2, 3, ..., 20}, hreshold={0.1, 0.2, ..., 0.9}
MeanShift	max_iter=600
OPTICS	min_samples={1000, 2000, ..., 10000}

### 3. Risk Estimation Method

클러스터링 기법을 통해 형성된 여러 개의 운전자 그룹별 위험도를 산출하기 위해, 먼저 각 그룹의 사고 빈도수(Frequency)와 사고 심각도(Severity)를 구한다.

사고 빈도수는 그룹에 포함된 샘플 수 및 사고 건수를 나타내며, 특정 그룹의 사고 빈도수가 높을수록 사고가 자주 발생하는 특성을 가지는 그룹으로 해석할 수 있다. 사고 심각도는 식 (1)과 같이 그룹 내의 총 사망자(Death), 중상자(Serious), 경상자(Minor) 수에 대한 가중치 평균을 통해 산출된다. 사고 심각도는 그룹 내 사망자 수가 많을수록 큰 값을 가진다. 사고 심각도가 클수록 사고가 발생하였을 때 큰 규모의 피해가 발생하는 특징을 가진 운전자 그룹으로 해석할 수 있다. 이때 가중치  $W_1$ ,  $W_2$ ,  $W_3$ 의 값은 사망, 중상, 경상에 대한 가중치 및 비용으로서 도로교통공단 보도자료[1]의 사망, 중상, 경상에 따른 1인당 평균 사고비용을 바탕으로  $W_1=563604$ ,  $W_2=74045$ ,  $W_3=5291$ 를 각 가중치 값으로 가진다.

$$\begin{aligned}
 Severity &= W_1 \times Death \\
 &+ W_2 \times Serious \\
 &+ W_3 \times Minor
 \end{aligned} \quad (1)$$

사고 빈도수를 바탕으로 여러 그룹 중 교통사고가 빈번히 발생하는 그룹을 나타낼 수 있으며, 사고 심각도를 바탕으로 사고 빈도수는 낮으나 사고의 규모 및 피해가 높은 그룹을 효과적으로 나타낼 수 있다. 이후, 산출된 사고 빈도수와 심각도를 바탕으로 그룹별 순위를 부여한 뒤 두 순위의 평균을 기준으로 종합적인 순위를 부여하여 클러스터별 위험도를 산출한다. 이때, 위험도의 값이 높을수록 그룹의 교통사고 위험도가 높음을 의미한다.

### 4. Experiment Results

본 논문에서 제시한 전처리 기법과 원-핫 임베딩 및 Min-Max Scaling을 적용하는 전처리 기법 간의 K-Means++ 클러스터링 알고리즘 실루엣 스코어 비교 결과는 Table 2와 같다. 원-핫 임베딩 전처리 기법의 경우 실루엣 스코어의 최대값은 0.566이었으나, 본 논문에서 제시한 전처리 기법은 0.821이었으며 전반적인 실루엣 스코어가 본 논문에서 제시한 기법이 더 우수함을 확인할 수 있다.

이러한 전처리 기법을 바탕으로 효과적인 클러스터링 알고리즘을 선택하기 위해 수행한 성능 비교 결과는 Table 3과 같다. OPTICS를 제외한 나머지 알고리즘의 경우 6~8개의 클러스터가 형성되었으며, 이 중 BIRCH와 MeanShift가 가장 높은 성능을 보였다. 또한, 7개의 알고리즘에 대한 T-SNE 시각화 결과는 fig. 1과 같이 비슷한 양상을 보이는 것을 확인하였다.

가장 높은 성능을 보인 MeanShift와 BIRCH 중 클러스터 수와 같은 별도의 하이퍼파라미터 없이 높은 성능을 달성할 수 있는 MeanShift를 기준으로 위험도를 산출한 결과는 Table 4와 같다. Cluster 0의 경우 가장 높은 사고 빈도수를 보였으며, Cluster 1의 경우 가장 높은 사고 심각도를 보였다. 특히, Cluster 1의 경우 사고 빈도수 또한 2번째로 높았기에 최종적으로 위험도 8에 해당하여 가장 위험한 운전자 그룹 및 특성으로 나타났다.

Table 2. K-Means++ clustering algorithm performance comparison

# of clusters	Silhouette Score	
	One-hot	Proposed
2	<b>0.566</b>	0.643
3	0.465	0.771
4	0.411	0.796
5	0.377	0.809
6	0.342	0.813
7	0.311	<b>0.821</b>
8	0.317	0.686
9	0.313	0.546
10	0.313	0.560

Table 3. Performance comparison of clustering algorithms

Algorithm	# of clusters	Silhouette Score
K-Means++	7	0.820
DBSCAN	7	0.820
GaussMixture	6	0.813
BIRCH	8	0.822
MeanShift	8	0.822
OPTICS	3	0.766
Agglomerative	7	0.820

Table 4. Risk level calculation results by cluster (F: Frequency, S: Severity, F\_Rank: Frequency Rank, S\_Rank: Severity Rank, RL: Risk Level)

Cluster	F	S	F_Rank	S_Rank	RL
0	40,224	19659.72	8	4	7
1	24,023	27785.21	7	8	8
2	13,411	18396.10	6	3	4
3	2,829	17317.31	5	2	3
4	1,777	25161.40	4	6	5
5	1,480	26472.97	3	7	5
6	971	17257.45	2	1	1
7	198	22694.94	1	5	2

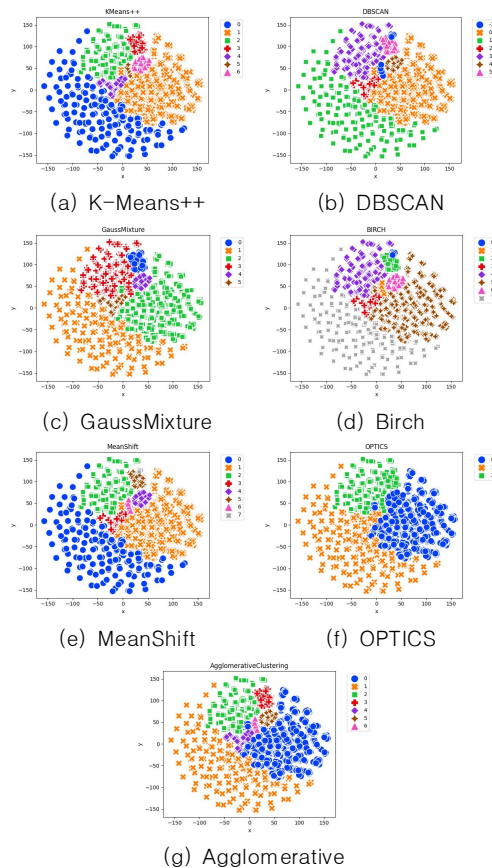


Fig. 1. T-NSE of clustering algorithm

## V. Conclusions

본 논문은 교통사고 데이터셋을 바탕으로 클러스터링 기법을 적용하여 각 운전자 그룹 및 특성에 따른 위험도를 산출하였다. 이 과정에서 통계자료에 기반한 임베딩 및 Scaling 기법이 기존에 일반적으로 사용되는 전처리 기법보다 클러스터링 성능 측면에서 더 뛰어난 결과를 보였다. 또한, 클러스터링 결과를 바탕으로 각 운전자 그룹 및 특성별 위험도를 산출하는 방법을 제시하였으며, 그 결과 여러 클러스터링 알고리즘 중 실루엣 스코어를 기준으로 성능 비교를 통해, 가장 높은 성능을 보인 MeanShift 알고리즘을 사용하여 각 운전자 그룹 및 특성별 위험도를 산출 및 분석하였다. 위와 같은 결과를 바탕으로 기존에 연구되었던 교통사고 위험도 예측 연구들과 결합하여, 도로의 시공간적 요인과 운전자의 인적 요인을 동시에 고려한 교통사고 위험도 예측 모델 구축을 향후 연구로 진행할 예정이다.

## ACKNOWLEDGEMENT

This work was supported by a National Research Foundation of Korea (NRF) grant funded by the Korea government (2022R1F1A1074273).

## REFERENCES

- [1] "In 2021, about KRW 26,998.7 trillion in social costs due to road traffic accidents was incurred" Accessed: Feb. 13, 2024. [Online]. Available: [https://www.koroad.or.kr/main/board/6/87791/board\\_view.do?&cp=1&listType=list&bdOpenYn=Y&bdNoticeYn=N](https://www.koroad.or.kr/main/board/6/87791/board_view.do?&cp=1&listType=list&bdOpenYn=Y&bdNoticeYn=N)
- [2] "Traffic Accident Analysis System(TAAS) Accessed: Feb. 13, 2024. [Online]. Available: [https://taas.koroad.or.kr/web/bdm/srs/select/StaticReportsList.do?menuId=WEB\\_KMP\\_IDA\\_SRS\\_TAA](https://taas.koroad.or.kr/web/bdm/srs/select/StaticReportsList.do?menuId=WEB_KMP_IDA_SRS_TAA)
- [3] S. Chen, K. Cheng, J. Yang, X. Zang, Q. Luo, and J. Li, "Driving Behavior Risk Measurement and Cluster Analysis Driven by Vehicle Trajectory Data," *Applied Sciences*, Vol. 13, No. 9, pp. 5675-7495, 2023, doi: 10.3390/AP13095675.
- [4] C. Zhang, Y. Ma, A. J. Khattak, S. Chen, G. Xing, and J. Zhang, "Driving style identification and its association with risky driving behaviors among truck drivers based on GPS, load condition, and in-vehicle monitoring data," *Journal of Transportation Safety & Security*, Vol. 16, No. 5, pp. 507-541, 2024, doi: 10.1080/19439962.2023.2233089
- [5] Y. Yao, X. Zhao, Y. Wu, Y. Zhang, and J. Rong, "Clustering

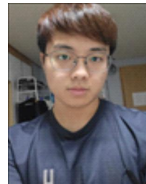
driver behavior using dynamic time warping and hidden Markov model,” *Journal of Intelligent Transportation Systems*, Vol. 25, No. 3, pp. 249–262, 2021.

- [6] A. R. Anil and J. Anudev, "Driver behavior analysis using K-means algorithm," 3rd International Conference on Intelligent Computing Instrumentation and Control Technologies (ICICT), pp. 1555-1559, 2022, doi: 10.1109/ICICT54557.2022.9917899.
- [7] Y. Chen, K. Wang, and J. J. Lu, "Feature selection for driving style and skill clustering using naturalistic driving data and driving behavior questionnaire," *Accident Analysis & Prevention*, vol. 185, pp. 107022-107037, 2023, doi: 10.1016/J.AAP.2023.107022.
- [8] J. Lee and K. Jang, "A framework for evaluating aggressive driving behaviors based on in-vehicle driving records," *Transportation Research Part F Traffic Psychology Behaviour*, vol. 65, pp. 610–619, 2019, doi: 10.1016/J.TRF.2017.11.021.
- [9] L. Zhang, H. Wu and K. Cui, "Evaluation of Bus Driving Behavior Based on Vehicle Location Data," 8th International Conference on Image, Vision and Computing (ICIVC), pp. 825-830, 2023, doi: 10.1109/ICIVC58118.2023.10270432.
- [10] R. Wang, W. Zheng, M. Huang and G. Li, "Driving Behavior Evaluation Based on DBSCAN and Kmeans++ Clustering," 5th International Conference on Advanced Electronic Materials, Computers and Software Engineering (AEMCSE), pp. 188-193, 2022, doi: 10.1109/AEMCSE55572.2022.00046.
- [11] K. Yang, C. Al Haddad, G. Yannis and C. Antoniou, "Driving Behavior Safety Levels: Classification and Evaluation," 7th International Conference on Models and Technologies for Intelligent Transportation Systems (MT-ITS), pp. 1-6, 2021, doi: 10.1109/MT-ITS49943.2021.9529309.
- [12] H. Xiao, Y. Lu, R. Su, B. Wang, N. Zhao and Z. Hu, "Clustering and Analysis of the Driving Style in the Cut-in Process," *IEEE 26th International Conference on Intelligent Transportation Systems (ITSC)*, pp. 3613-3618, 2023, doi: 10.1109/ITSC57777.2023.10421969.
- [13] "Traffic Accident Analysis System(TAAS)" Accessed: Feb. 13, 2024. [Online]. Available: [https://taas.koroad.or.kr/web/shp/sbm/.initGisAnals.do?menuId=WEB\\_KMP\\_GIS\\_TAS](https://taas.koroad.or.kr/web/shp/sbm/.initGisAnals.do?menuId=WEB_KMP_GIS_TAS)
- [14] F. Pedregosa FABIANPEDREGOSA et al., "Scikit-learn: Machine Learning in Python," *Journal of Machine Learning Research*, Vol. 12, pp. 2825–2830, 2011, [Online]. Available: <http://jmlr.org/papers/v12/pedregosa11a.html>
- [15] L. Van Der Maaten and G. Hinton, "Visualizing Data using t-SNE," *Journal of Machine Learning Research*, Vol. 9, No. 11, pp. 2579–2605, 2008.

## Authors



Tae-Wook Kim is an undergraduate majoring in Division of Software, Yonsei University, Korea. Mr. Kim's research interests include Data Mining, Optimization, and applied artificial intelligence.



Ji-Woong Yang received the B.S. in Department of Computer & Telecommunications Engineering, Yonsei University, Korea, in 2024. Mr. Yang is currently pursuing the combined M.S./Ph.D.

degree in Artificial Intelligence Semiconductor, Hanyang University, Korea. His research interests include Machine Learning, Deep Learning, and Computer Vision.



Hyeon-Jin Jung received the B.S degree in Department of Computer & Telecommunications Engineering, Yonsei University, Korea, 2024. Mr. Jung is currently pursuing the combined M.S./Ph.D.

degree in Department of Computer Science, Yonsei University. His research interests include Artificial Intelligence (AI) and Lightweight Deep Learning.



Han-Jin Lee received the B.S degree in Department of Computer & Telecommunications Engineering, Yonsei University, Korea, 2024. Mr. Lee is currently pursuing the combined M.S./Ph.D.

degree in Department of Computer Science, Yonsei University. His research interests include Machine Learning, Deep Learning, and Computer Vision.



Ellen J. Hong received B.S. degree in electrical engineering from the Pusan National University (PNU), Busan, in 2005, and M.S., Ph.D. degrees in electrical engineering from KAIST, Daejeon, in 2007 and 2013.

Dr. Hong was a postdoctoral researcher in KAIST and PNU, from 2013 to 2016. From 2016 to 2017, she was a professor at the Division of Computing Engineering, Dongseo University, Busan. She was a research engineer at the Convergence Laboratory, KT, from 2018 to 2019. Since 2019, she has been with Yonsei University, Wonju, where she is currently a professor at the Dept. of Computer and Telecommunications Engineering. Her research interests include deep learning, digital twin, cyber physical system, and evolutionary computation.