

# MG-TAR: Multi-View Graph Convolutional Networks for Traffic Accident Risk Prediction

Patara Trirat<sup>ID</sup>, Susik Yoon<sup>ID</sup>, and Jae-Gil Lee<sup>ID</sup>, *Member, IEEE*

**Abstract**—Due to the continuing colossal socio-economic losses caused by traffic accidents, it is of prime importance to precisely forecast the traffic accident risk to reduce future accidents. In this paper, we use **dangerous driving statistics** from driving log data and **multi-graph learning** to enhance predictive performance. We first conduct geographical and temporal correlation analyses to quantify the relationship between **dangerous driving** and **actual accidents**. Then, to learn various dependencies between districts besides the traditional adjacency matrix, we simultaneously model both **static** and **dynamic** graphs representing the spatio-temporal contextual relationships with heterogeneous environmental data, including the dangerous driving behavior. A graph is generated for each type of the relationships. Ultimately, we propose an end-to-end framework, called MG-TAR, to effectively learn the association of multiple graphs for accident risk prediction by adopting **multi-view graph neural networks** with a **multi-attention module**. Thorough experiments on ten real-world datasets show that, compared with state-of-the-art methods, MG-TAR reduces the error of predicting the accident risk by up to 23% and improves the accuracy of predicting the most dangerous areas by up to 27%.

**Index Terms**—Traffic accident risk prediction, spatial-temporal data mining, dangerous driving behavior, graph neural networks, intelligent transportation system.

## I. INTRODUCTION

CONTINUOUS growth of urban development in many countries has led to huge losses of life by traffic accidents. According to the World Health Organization [1], traffic accidents killed nearly 1.35 million people each year, which has become a critical socioeconomic problem for humankind as most of them were children and young adults. Given the huge number of vehicles in the complex transportation system, we need not only an understanding of the causes of accidents but also a reliable and precise traffic accident prediction model to mitigate potential traffic accidents. With such knowledge, public safety stakeholders (e.g., police and traffic authorities) can proactively allocate public

Manuscript received 6 August 2021; revised 18 April 2022; accepted 29 December 2022. This work was supported by the Institute of Information and Communications Technology Planning and Evaluation (IITP) Grant funded by the Korea Government (MSIT) (No. 2020-0-00862, DB4DL: High-Usability and Performance In-Memory Distributed DBMS for Deep Learning and No. 2022-0-00157, Robust, Fair, Extensible Data-Centric Continual Learning). The Associate Editor for this article was G. Guo. (*Corresponding author: Jae-Gil Lee.*)

Patara Trirat and Jae-Gil Lee are with the School of Computing, KAIST, Daejeon 34141, South Korea (e-mail: patara.t@kaist.ac.kr; jaegil@kaist.ac.kr).

Susik Yoon is with the Department of Computer Science, UIUC, Urbana, IL 61801 USA (e-mail: susik@illinois.edu).

Digital Object Identifier 10.1109/TITS.2023.3237072

resources and design better transportation planning strategies to improve road safety. Location-based services such as a navigation application can also be informed of dynamic traffic accident risk to offer a safer route. As a practical example, the deployment of the Tennessee accident prediction system reduced the death rate from traffic accidents by about 8% [2].

Behind the enormous benefits to be expected, **traffic accident risk prediction** is still a challenging task due to the sparsity and complexity of traffic accidents, which exhibit a sporadic nature, inconspicuous patterns, and intricate spatio-temporal correlations. Thus, existing approaches for continuous spatio-temporal prediction (e.g., traffic volume prediction) [3], [4], [5], [6] do not suit traffic accident forecasting, as reported by Zhou et al. [7]. Recent studies proposed graph neural network (GNN)-based models with a simple single-view topology modeling—by either traffic condition [2], [7] or spatial adjacency [8] correlation—to handle the complexity problem and employed data augmentation techniques to mitigate the **sparsity problem**. However, a *single* view of the connections between nodes in a graph may not fully reflect the dynamic structure of traffic accidents caused by various factors, while other contextual information such as district categories and road types is also crucial for accurate prediction. The data augmentation approach is not a fundamental solution because it is highly dependent on the existing data and does not add informative but unrevealed underlying factors that may cause traffic accidents, e.g., **dangerous driving** that did not lead to actual traffic accidents.

### A. Main Ideas

In this study, we aim to exploit **dangerous driving statistics** for traffic accident risk prediction to mitigate the sparsity problem of the accident data and **multi-view graph learning** to tackle the complexity of the prediction by explicitly formulating different spatial-temporal relationships of transportation networks among geographic districts. Dangerous driving, in fact, is one of the most significant factors that cause traffic accidents [9], [10], [11], along with unsafe road conditions and inadequate protective equipment. That is, dangerous driving statistics are supplemental to traffic accidents and can be treated as near-miss events. Here, we use these statistics as new indicative features for training the predictive model. Also, driving behavior data can be obtained continuously using various in-vehicle sensors. For instance, commercial

vehicles in several countries [12], [13] are installed with a digital tachograph (DTG) device, and automobile insurance companies encourage customers to mount such a device for being aware of their driving behavior by offering them a discount on the premium in return [14], [15], [16]. Thus, the feasibility and practicality of our approach are confirmed.

As shown by recent studies [17], [18], [19], [20], [21], the information from not only spatially adjacent regions but also contextually correlated regions is crucial for precise prediction. Accordingly, we adopt *multi-view* graph learning to model the diverse correlations between regions represented by multimodal node features. The multi-view approach is preferable to the single-view approach in real-world settings as the former provides more comprehensive representations that contain complementary information. In the graph context, we explicitly model the different relationships with multiple graphs, each corresponding to a type of relationships, e.g., spatial closeness, functional similarity, and road specifications.

Properly incorporating the dangerous driving statistics into multi-view graph learning, however, comes with challenges. First, it is not trivial to derive the dangerous driving statistics from raw driving records and verify their spatial and temporal correlations with past accidents since dangerous driving characteristics can vary case by case. Second, modeling only *static* relationships among regions may not be sufficient to reflect the complex spatio-temporal correlations of traffic accidents and other environmental features including dangerous driving statistics, which can change over time. While the dynamic graphs have been used recently in relevant problems (e.g., traffic forecasting [22], [23]) to model such dynamic relationships among regions adaptively at each time interval, they have not been explored in the context of traffic accident risk prediction with dangerous driving statistics. Given these challenges, we propose Multi-view Graph convolutional networks for Traffic Accident Risk prediction (**MG-TAR**) by comprising the following main ideas.

1) *Understanding Dangerous Driving Behavior*: To solve the first challenge, we adopt a rule-based classification for extracting dangerous driving offenses based on official rules well-defined by domain experts. It is fast and generalizable as the rules can be customized for the region where the driving records are being collected. After classifying the driving records, we quantify the correlation between dangerous driving cases and past traffic accidents in geographical and temporal aspects using the Pearson correlation to verify its efficacy.

2) *Simultaneous Modeling of Static & Dynamic Graphs*: To solve the second challenge, as illustrated in Fig. 1, we simultaneously model static and dynamic contextual spatial-temporal relationships of different areas. For the static cases, the different types of relationships are encoded into multiple graphs by measuring the similarity between nodes with static features, e.g., point-of-interest (POI) information. Nodes will be linked if their similarity is above a threshold. Then, we construct the dynamic spatial-temporal dependencies with time-variant features, incorporating dangerous driving statistics, by computing the similarity between regions at each time step. Therefore, the model can jointly learn spatial and

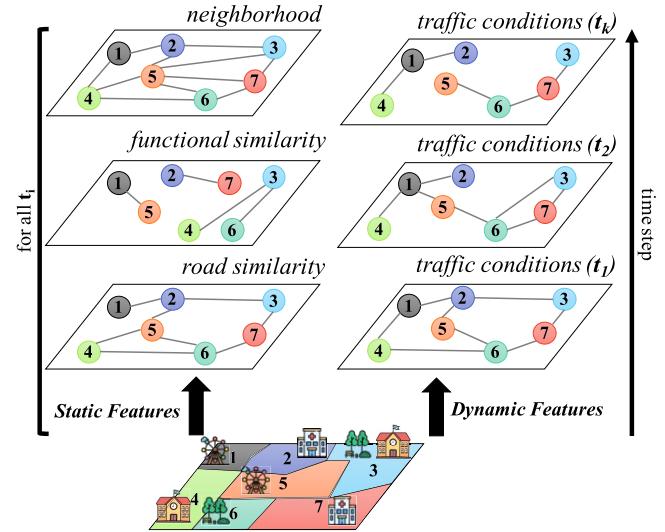


Fig. 1. An example of multi-view correlations among regions.

temporal relationships among regions to make more comprehensive representations of each node. To the best of our knowledge, none of the existing studies concurrently consider both the static and dynamic spatial-temporal correlations for traffic accident risk prediction.

3) *Learning With Multi-Attention Module*: We employ the *attention learning scheme* to adaptively prioritize the contributions from multiple views and temporal scopes. Specifically, MG-TAR is equipped with *two* attention modules to capture each view's contribution (i.e., inter-view attention) and the importance of each time step (i.e., temporal attention) on the predicted accident risk scores. Hence, MG-TAR is capable of learning the dynamic representations from multiple views by giving more weight to the essential information. It is worth noting that the attention mechanism used in this paper is different from the one used in graph attention networks. In particular, inter-view attention is designed to prioritize the importance of different views, while attention-based graph learning [19], [24], [25] is designed for weighting the spatial correlations between nodes in a graph.

## B. Summary

Our key contributions are summarized as follows.

- We prepare dangerous driving behavior cases and conduct a correlation analysis showing how dangerous driving behavior correlates with actual traffic accidents using real-world driving log data from five cities.
- We simultaneously model both the static and dynamic spatial-temporal relationships between districts using heterogeneous features into a multi-view graph scheme.
- We propose, MG-TAR, a novel multi-view graph convolutional network with the multi-attention module to simultaneously learn and emphasize the essential representations of different spatial-temporal dependencies for traffic accident risk prediction.
- To demonstrate the superiority of MG-TAR and gain more insightful results, we conduct extensive experiments

against various state-of-the-art methods. The results show that MG-TAR reduces the regression error by up to **23%** in MSE and increases the accuracy of predicting the most dangerous areas by up to **27%** in Recall@K.

The rest of the paper is organized as follows. Section II reviews related work. Section III explains preliminaries and problem statements. We present how dangerous driving cases correlate to traffic accidents in Section IV. The proposed MG-TAR is described in Section V. Then, Section VI presents the experiment results and Section VII discusses a few limitations. Finally, Section VIII concludes this study.

## II. LITERATURE REVIEW

### A. Traffic Forecasting

Initial attempts investigated traffic prediction using conventional methods (e.g., ARIMA, SVR, and k-NN) for specific road segments [26], [27], [28], [29], [30]. However, they have limitations in learning non-linear and dynamic spatio-temporal patterns. Recently, with the rise of deep learning that showed superiority in capturing non-linear spatial and temporal correlations in traffic conditions, researchers have proposed citywide traffic prediction models [3], [4], [31] where most of them use convolutional neural networks (CNNs) for modeling spatial correlations in grid-based Euclidean space and recurrent neural networks (RNNs) for modeling temporal patterns [32]. More sophisticatedly, Guo and Zhang [33] proposed a residual spatio-temporal network based on an augmented convolutional long short-term memory (ConvLSTM) and residual learning to jointly learn the spatial and temporal features for more accurate results. Besides, a BERT-based model [34] was designed for long-range traffic flow forecasting.

To incorporate spatial dependencies more effectively, more recent studies formulated the traffic prediction on graphs and utilized graph convolutional networks (GCNs) to model the **non-Euclidean** correlations in the traffic networks [5], [35], [36], [37], [38]. For example, Wu et al. [37] proposed **Graph Wavenet** with dilated convolution by adopting diffusion convolution in spatial modeling [35] to consider both connected and unconnected nodes in the modeling process. Moreover, many researchers additionally combined the attention mechanism [24], [39] to advance their models' performance with the Transformer-like encoder-decoder architecture [6], [22], [25], [40], [41], [42]. Another line of work focused more on data-related problems. For instance, Cui et al. [43] proposed a graph Markov network to deal with missing values. Hierarchical or multi-scale models [44], [45] were also proposed to enhance forecasting performance in different scales and effectively learn the hierarchical nature of traffic networks. For better input feature extraction, Zhu et al. [46] built an AST-GCN model specifically for learning dynamic and static external information, while Ali et al. [47] built a TEST-GCN model with similar motivation but for topological features of road networks. **Lastly, due to the insufficiency in the expressiveness of static graph structures, researchers have started to comprehensively model dynamic graphs to deal with complex spatio-temporal dependencies in road traffic** [6], [22], [23], [48].

Despite the advancement in traffic forecasting and the promising results, we argue that those intensive and continuous sequence prediction cannot be directly used for traffic accident prediction due to the sporadic spatial-temporal distribution, especially in the finer granularity, as well as the complex correlation with exogenous factors. Therefore, many attempts have been proposed to tackle these challenges in forecasting traffic accidents, as summarized in the next subsection.

### B. Traffic Accident (Risk) Prediction

Earlier approaches [49], [50], [51], [52], [53] tried to understand the impact of limited environmental factors and predict the accidents on specific road segments using traditional methods such as regression models and decision trees. Other earlier methods [54], [55], [56] analyzed traffic accidents in multiple highways with linear models. More recent studies [57], [58], [59] extended the study area from specific roads to citywide traffic zones. These studies applied classical data mining techniques on **small-scale traffic accident data** with limited features and did not efficiently address unique data characteristics such as time periodicity, spatial autocorrelation, and heterogeneity, thereby failing to achieve high accuracy.

Thanks to the power of deep neural networks in learning spatio-temporal patterns and meaningful representations of contextual data, numerous studies [60], [61], [62] started utilizing fully-connected networks to learn the hierarchical non-linear representations of traffic-related data. Ren et al. [63] and Moosavi et al. [64] built LSTM models to capture the temporal dynamics with other contextual information. More sophisticated methods [65], [66] attempted to jointly learn the spatial-temporal correlations with a ConvLSTM-based architecture. Similarly, TA-STAN [67] and DFN [68] employed spatial and temporal attention mechanisms to model the spatio-temporal patterns dynamically. For a higher resolution, He et al. [69] proposed a ResNet-based [70] model for risk map prediction on road segments trained with satellite image, road segments, and GPS data. With the same goal, Charandabi et al. [71] proposed a radial basis function-based neural network with self-organizing map to estimate the road accident risk map trained with heterogeneous features collected through crowdsourcing.

As graph-based networks have been proven efficient in traffic prediction, a few researchers [2], [7], [8] adopted a GCN for finer granularity prediction. **Recently, GSNet** [21] **was introduced with the use of semantic spatial correlations among regions. Although this work has the closest relationship with our work, it is worth noting that GSNet only considered static semantic graphs and separately learns the semantic and geographic features.** In contrast, we simultaneously model the multiple graphs from different views, including dynamic ones, and adaptively weigh the importance of different semantic features (i.e., views) through **the inter-view attention module to effectively incorporate multi-source spatial-temporal factors coupled with traditional attention in the temporal dimension.**

Even though these studies addressed the spatial-temporal patterns of traffic accident data and achieved better prediction results than the traditional approaches, as reported by these papers [8], [67], the artificial separation of the administrative

district (i.e., grid maps) breaks the whole pattern of the regions and causes deviation from the prediction. Therefore, to lessen the deviation of predicted results, we train the proposed MG-TAR based on the actual administrative districts, where each district ranges from about  $2\text{km}^2$  to  $400\text{km}^2$ , instead of the grid maps. This is also aligned with the actual distribution of the police traffic department and fire station in South Korea.

### III. PRELIMINARIES

#### A. Driving Behavior (Digital Tachograph) Data

A digital tachograph (DTG) is a driving log recorder in a vehicle, which collects time-stamped driving records—typically, every few seconds—including car location, speed, acceleration, vehicle type, and more. Table I presents the key features of the DTG records. Its installation promotes road safety, prevents severe driving behavior, and encourages fair competition among the transportation companies [13]. The DTG data which we use in this paper was collected in September 2016 and September 2018 and was provided by the Korea Transportation Safety Authority (KTSA). Ten commercial vehicle types, taxis (personal and corporate), buses (town, city, rural, intercity, express, and rent), and trucks (personal and general), are included. The records that appear in five big metropolitan cities (i.e., Seoul, Busan, Daegu, Gwangju, and Daejeon) in Korea are selected because these cities have enough population and diverse types of roads and vehicles.

#### B. Criteria for Dangerous Driving

There are various criteria or rules for judging dangerous driving in each country. Any criteria can be employed for our methodology. Here, we used the criteria defined by the KTSA [72], which are widely adopted in Korea. Each rule is specified by imposing a threshold condition on a variable such as speed, acceleration, and direction. There are nine types of dangerous driving offenses, *overspeed(OS)*, *rapid acceleration(RA)*, *quick start(QS)*, *rapid deceleration(RD)*, *sudden stop(SS)*, *sudden lane change(SLC)*, *sudden overtaking(SO)*, *sharp turn(ST)*, and *sudden u-turn(SUT)*. The threshold conditions are dependent on vehicle types; for example, the threshold for *RA* is  $5\text{km/h}$  per second for trucks, but  $6\text{km/h}$  per second for buses. Table II shows the criteria for each vehicle type which we consider.

Note that we apply the above criteria because of the source of our datasets. Also, it will be straightforward to either adopt these rules or apply a new set of rules when using our framework for other regions or countries. However, if specific criteria are not available, a data-driven approach such as clustering will also be feasible to extract underlying characteristics of driving log data.

#### C. Problem Definitions

We summarize all the necessary mathematical symbols used throughout this paper in Table III.

**Definition 1 (City Graph):** A **city graph**  $\mathcal{G} = (\mathcal{D}, \mathcal{E})$  is an undirected graph constructed from actual administrative division. Here,  $\mathcal{D} = \{d_1, d_2, \dots, d_n\}$  is the set of nodes (i.e.,

TABLE I  
KEY FEATURES IN THE DTG DATASETS

Name	Description	Example
Trip key	Identifier key of a trip	C5000123
Date	Recorded date (YYYYMMDD)	20180913
Time	Recorded time (HHMMSS)	140848
Vehicle type	Type of the vehicle	city bus
RPM	Engine revolution per minute (0~9999)	600
Speed	Current driving speed in km/h (0~255)	19
X acceleration	X-axis acceleration in $\text{m/s}^2$ (-100~100)	2.5
Y acceleration	Y-axis acceleration in $\text{m/s}^2$ (-100~100)	-1.1
Brake signal	Brake status: off (0) or on (1)	1
X coordinate	GPS longitude	127.124885
Y coordinate	GPS latitude	37.46692
Azimuth	Car direction in degree ( $0^\circ \sim 360^\circ$ )	168

districts of a city), where  $d_i$  denotes the  $i$ -th district specified by its geographic border. Given two nodes  $d_i, d_j \in \mathcal{D}$ , the edge  $e_{ij} \in \mathcal{E}$  of these two nodes indicates the connection between two adjacent districts.

**Definition 2 (Driving Log Record):** A **driving log record**  $r_d^t$  is a collection of measurements of the features in Table I at a district  $d$  within a specific time interval  $t$  (e.g., 10–11 AM). The time interval of this study is one hour, corresponding to the ground truth interval length.

**Definition 3 (Dangerous Driving Case):** A set of **dangerous driving cases** is  $c_d^t = \{c_{d,1}^t, c_{d,2}^t, \dots, c_{d,b}^t\}$ . Here,  $c_{d,b}^t = \sum_{r \in R} r_{d,b}^t$ , where  $t$  is the occurred timeframe in a district  $d$ , and  $R$  is the set of all driving records;  $r_{d,b}^t = 1$  if it satisfies the conditions of an offense  $b$  (e.g., in Table II), otherwise 0.

**Definition 4 (Environmental Features):** A set of **environmental features** encompasses both static and dynamic aspects (e.g., *road specifications and weather conditions*) that affect traffic accidents. For each  $d \in \mathcal{D}$ , all features are aggregated to form fixed-length vectors  $g_{d,\text{static}}^t$  and  $g_{d,\text{dynamic}}^t$  indicating the static features and dynamic features, respectively, for the district  $d$  during the time interval  $t$ .

**Definition 5 (Traffic Accident Risk [60]):** The **traffic accident risk** quantifies the degree of risk caused by the traffic accidents occurred in a district  $d$  at a time interval  $t$ ,

$$risk_d^t = \sum_{s=1}^4 p_{d,s}^t \times l_s, \quad (1)$$

where  $p_{d,s}^t$  is the number of injured people of an accident  $s$  with a severity level  $l_s$  during the time interval  $t$  in the district  $d$ . Specifically, the severity levels are one of {1: slight injury with consciousness, 2: small injury without consciousness, 3: serious injury, 4: fatality}.

**Definition 6 (Node Feature Matrix):** A **node feature matrix** or graph signal matrix  $\mathcal{F}^t = \{f_{d_1}^t, f_{d_2}^t, \dots, f_{d_n}^t\} \in \mathbb{R}^{N \times M}$ , in a given city graph, represents a feature matrix for all districts at a time interval  $t$ , where  $N = |\mathcal{D}|$  and  $M$  is the number of features. Here,  $f_d^t = \{g_d^t, c_d^t, risk_d^t\}$ .

**Definition 7 (Contextual Matrix):** A **contextual matrix**  $\mathcal{A}_f \in \mathbb{R}^{N \times N}$  represents the contextual similarity between  $d_i, d_j \in \mathcal{D}$  as an additional adjacency matrix, where  $f$  indicates a feature set used to compute the similarity scores between  $d_i$  and  $d_j$ . We model both static and dynamic contextual correlations between districts.

TABLE II  
CRITERIA FOR DANGEROUS DRIVING BEHAVIOR BY KTSA

Offense	Definition of Criteria for {Truck, Bus, Taxi}
OS	Speed exceeds the road speed limit by more than <b>20km/h</b> .
RA	Speed is greater than 6km/h and the acceleration is more than <b>{5, 6, 8}km/h</b> per second.
QS	Speed is lower than 5km/h and the acceleration is more than <b>{6, 8, 10}km/h</b> per second.
RD	Speed is greater than 6km/h and the deceleration is more than <b>{8, 9, 14}km/h</b> per second.
SS	Speed is lower than 5km/h and the deceleration is more than <b>{8, 9, 14}km/h</b> per second.
SLC	Speed is greater than 30km/h with the angle of more than <b>{6, 8, 10}°</b> per second (left or right) while the acceleration is lower than $\pm 2$ km/h per second and cumulative angle for 5 seconds is lower than $\pm 2$ ° per second.
SO	Speed is greater than 30km/h with the angle of more than <b>{6, 8, 10}°</b> per second (left or right) while the acceleration is greater than 3 km/h per second and cumulative angle for 5 seconds is lower than $\pm 2$ ° per second.
ST	Speed is greater than <b>{20, 25, 30}km/h</b> and the cumulative angle within <b>{4, 4, 3} seconds</b> is between 60° and 120° (left or right).
SUT	Speed is greater than <b>{15, 20, 25}km/h</b> and the cumulative angle within <b>{8, 8, 6} seconds</b> is between 160° and 180° (left or right).

TABLE III  
DESCRIPTION OF MATHEMATICAL SYMBOLS

Notation	Description
$N$	Number of districts in the city graph
$M$	Number of features or feature dimensions
$T$	Temporal length of historical features for training
$L$	Length of the forecasting horizon
$\mathcal{D} = \{d_i\}$	Spatial city graph node set of administrative districts
$\mathcal{E}$	Edges between connected node districts
$\mathcal{F}^t$	Feature matrix for all districts at time step $t$
$\mathcal{A}_f^t$	Adjacency (similarity) matrix for the contextual feature $f$ at time $t$
$\varepsilon_f$	Similarity threshold for contextual matrix $\mathcal{A}_f$
$h_l^t$	Hidden representations of GCN layer $l$ at time $t$
$H^t$	Final representations after the temporal attention module
$H_v^t$	Representations of GCN of a contextual view $v$ at time $t$
$H_{attn}^t$	Representations of the inter-view attention module at time $t$

The set of environmental features (Definition 4) is described in Section VI-A. In addition, the feature sets used for contextual matrices (Definition 7) are detailed in Section V-A.

**Traffic Accident Risk Prediction:** Given the city graph with multi-view contextual matrices  $\mathcal{G} = (\mathcal{D}, \mathcal{E}, \mathcal{A}_f)$  and historical node features  $\mathcal{F}^t$  ( $t = 1, 2, \dots, T$ ), we aim to predict the traffic accident risk scores  $risk_d^{t'}$  for  $d \in \mathcal{D}$  and  $t' = T+1, T+2, \dots, T+L$ , where  $L$  is the forecast horizon.

#### IV. CORRELATION ANALYSIS

Intuitively, if the dangerous driving behavior has a significantly high correlation with the actual accident records, incorporating the dangerous driving statistics into the prediction step is needed to increase the prediction accuracy. In this section, we study its correlation with the past traffic accidents and report the results according to the geographical and temporal aspects obtained by Def. 3—namely, a **subdistrict level** and an **hour interval level**. Specifically, we quantify the correlation scores using the Pearson correlation coefficient, defined by

$$Corr(G) = \frac{\sum_{g \in G} (c_g - \bar{c})(a_g - \bar{a})}{\sqrt{\sum_{g \in G} (c_g - \bar{c})^2} \sqrt{\sum_{g \in G} (a_g - \bar{a})^2}}, \quad (2)$$

where  $Corr(G)$  is the correlation coefficient between the frequency  $c$  of dangerous driving cases and the frequency  $a$  of past accidents of a given group  $G$  (geographical or temporal).

TABLE IV  
DRIVING LOG DATASETS STATISTICS

City	Period (Sept.)	Interval (second)	# Districts (Sub-)	# Data Entries	# Trips	File Size
Seoul	2016	10	25	255M	134,201	54GB
	2018	1	(467)	375M	38,401	115GB
Busan	2016	10	16	216M	101,085	46GB
	2018	1	(192)	180M	27,915	56GB
Daegu	2016	10	8	169M	91,803	36GB
	2018	1	(204)	100M	16,150	31GB
Gwangju	2016	10	5	166M	77,963	35GB
	2018	1	(202)	166M	25,928	51GB
Daejeon	2016	10	5	122M	87,933	26GB
	2018	1	(177)	157M	24,115	49GB

As a preprocessing step, we assign an *idle state* (i.e., parked car) to the raw driving records to exclude the records that have a minimal possibility of inducing traffic accidents. We define the idle state as follows: (1) RPM is less than 1000, (2) Speed is  $0km/h$ , (3) X and Y accelerations are between  $-1m/s^2$  and  $1m/s^2$ , and (4) no brake signal. Table IV shows the statistics of the driving records datasets used in this analysis.

We present the correlation coefficient scores averaged from the five cities in Table V. Overall, the **dangerous driving offenses** are strongly correlated with the **frequency of traffic accidents** both in temporal and geographic aspects. Interesting findings and further analyses are provided as follows.

#### A. Counterintuitive Results About “Overspeeding”

Before giving further findings, it is worth noting the negative effect of *overspeed* (OS) on the correlation scores. The OS behavior not only shows the **inconsistent values** in both geographical and temporal aspects ( $0.3 \leq s.d. \leq 0.6$ ), but also the computed scores are **not statistically significant** ( $p\text{-value} \geq .10$ ) in many cases regardless of the values, possibly because (1) the OS mostly occurred at very late night that the number of vehicles on the roads is extremely small and (2) the drivers of those vehicles may be high-skilled drivers who are very confident in their driving skills. Additionally, as a piece of evidence, the historical record from Korea’s traffic

TABLE V  
OVERALL CORRELATION SCORES OF EACH OFFENCE TYPE AVERAGED FROM FIVE CITIES

Level	Dataset (Sept.)	Dangerous Driving Behavior										All	w/o OS
		OS	RA	QS	RD	SS	SLC	SO	ST	SUT			
Hour	2016	0.31	0.59	0.80	0.62	0.62	0.62	0.38	0.65	0.57	0.69	0.72	
	2018	-0.29	0.70	0.82	0.65	0.62	0.70	0.40	0.74	0.76	0.77	0.77	
	Average	0.01	0.65	0.81	0.64	0.62	0.66	0.39	0.70	0.67	0.73	0.75	
Subdistrict	2016	0.18	0.85	0.80	0.69	0.60	0.64	0.72	0.81	0.79	0.84	0.85	
	2018	0.53	0.91	0.88	0.74	0.69	0.72	0.86	0.84	0.86	0.91	0.91	
	Average	0.36	0.88	0.84	0.71	0.65	0.68	0.79	0.83	0.82	0.88	0.88	

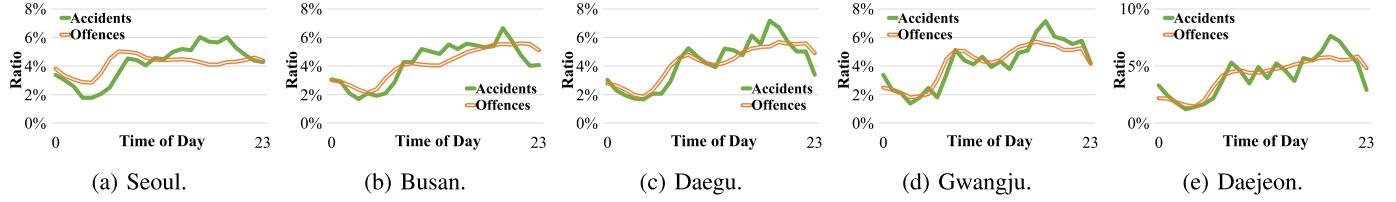


Fig. 2. Comparison between the occurrence ratio of all driving offences and traffic accidents of each city by time of day.

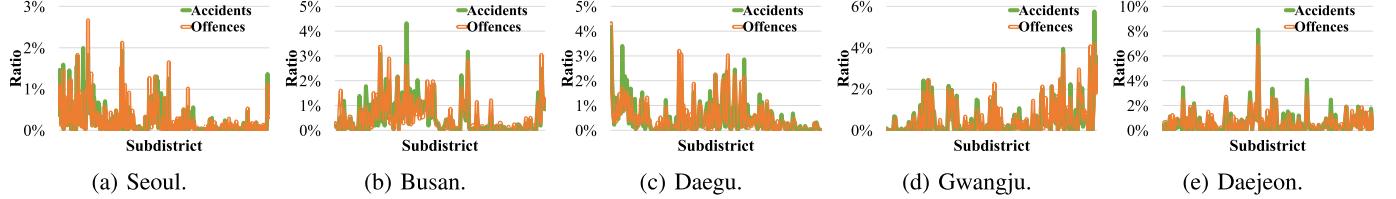


Fig. 3. Comparison between the occurrence ratio of all driving offences and traffic accidents of each city by location.

accident analysis system (TAAS)<sup>1</sup> also confirms that, among other violations, the traffic accidents caused by *OS* take the smallest portion—about 0.1%.

### B. Temporal Correlation (Hour-Interval Level)

In terms of temporal analysis, we compare the occurrence number of the dangerous driving behavior within an hour interval as it is the smallest unit in the given accident records. The analysis results exhibit a strong correlation in both datasets—0.69 in the year 2016 and 0.77 in the year 2018—regarding the time of an occurrence. Specifically, in general, *QS*, *ST*, and *SUT* are the top three dangerous driving cases that have a strong correlation to past accidents. Additionally, as illustrated in Fig. 2, the occurrence ratios of both dangerous driving offenses and past traffic accidents reveal similar patterns, indicating that they occurred more often at night.

### C. Geographic Correlation (Subdistrict Level)

Regarding the location of an occurrence in the subdistrict level, the overall number of dangerous driving cases strongly correlates to traffic accidents—0.84 in the year 2016 and 0.91 in the year 2018. Specifically, *RA*, *QS*, and *ST* are the top three dangerous driving cases that have a strong correlation to past accidents. Fig. 3 presents the occurrence ratios of dangerous driving cases and actual traffic accidents in the subdistricts.

<sup>1</sup><http://taas.koroad.or.kr/>

## V. THE MG-TAR MODEL

This section describes the detailed elements of the proposed multi-view graph convolutional networks for traffic accident risk prediction (MG-TAR) model. Fig. 4 illustrates the overall architecture of MG-TAR consisting of a multi-view graph construction module, graph convolutional networks, an interview attention module, and a temporal attention module, which are detailed in the following subsections.

### A. Multi-View Graph Construction

We build multiple graphs incorporating both the static and dynamic dependencies between nodes by computing the spatial and temporal contextual similarities between node features. Based on Def. 1 and Def. 7, the edges between districts in a city graph are formed if the similarity score exceeds a threshold  $\varepsilon_f$  for each  $\mathcal{A}_f$ , where  $f \in \{F, R, D, P\}$ . Intuitively, we set  $\varepsilon_f$  equal to the average of scores in the matrix to control its sparsity such that the matrix is not too dense (all districts are connected) or too sparse (no contextually adjacent district). Each optimal  $\varepsilon_f$  can be learned with a regression model of the predicted outputs, and we leave it as a topic of future work.

1) **Static Dependency Modeling:** We model the static contextual matrices similar to ST-MGCN [18] by encoding different types of non-Euclidean pair-wise correlations between districts into multiple graphs. The key idea is to explicitly leverage contextual information among regions for accurate

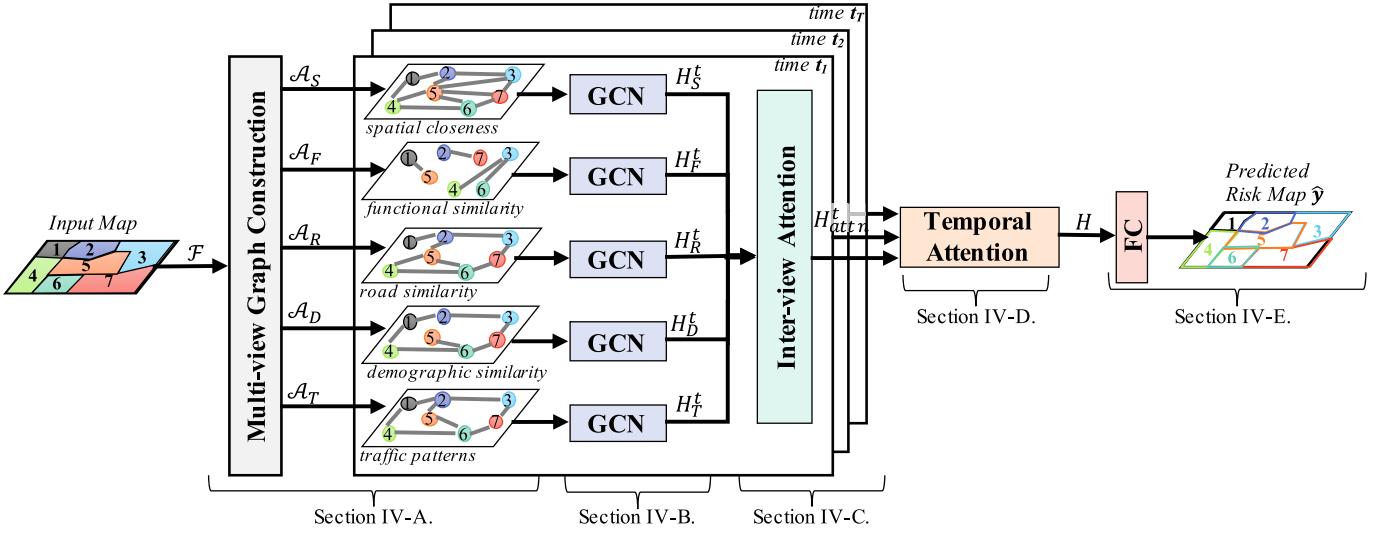


Fig. 4. The overall architecture of the MG-TAR model.

prediction because not only the spatial proximity between districts is correlated, but also the similarity of their additional features is. As illustrated in Fig. 1, a district may correlate to another distant region that shares similar functionality. In this study, the following four features are considered.

a) **Spatial (S) closeness:** The spatial closeness of a district is defined by the spatial proximity. We construct the graph by linking the districts whose boundaries are adjacent to each other, such that

$$\mathcal{A}_{S,ij} = \begin{cases} 1 & d_i, d_j \in \mathcal{D} \text{ are adjacent} \\ 0 & \text{otherwise.} \end{cases} \quad (3)$$

b) **Functional (F) similarity:** It is natural to refer to other districts similar to the current district in terms of functionality when making a prediction. We characterize the district functionality using point-of-interest(POI) data. The POI information exhibits the frequency distribution of POI categories in a district.

$$\mathcal{A}_{F,ij} = \begin{cases} 1 & sim(f_{poi,d_i}, f_{poi,d_j}) \geq \varepsilon_F \\ 0 & \text{otherwise,} \end{cases} \quad (4)$$

where  $f_{poi,d}$  represents the POI information of a district  $d$ .

c) **Road (R) similarity:** Two districts with similar road structures tend to have comparable accident patterns. In other words, the number of accidents relates to the road structures, so it is beneficial to aggregate the information from the districts with similar road structures. Following Eq. 4, we define  $\mathcal{A}_R$  by replacing  $f_{poi,d}$  with  $f_{road,d}$ , where  $f_{road,d}$  represents the road structure features.

d) **Demographic (D) similarity:** Furthermore, human behavior is also relevant to traffic accidents because different groups of people tend to have different driving characteristics. Therefore, aggregating the information from demographically similar regions will help understand the relationship between the demographic data and the accidents. Identically, we

compute  $\mathcal{A}_D$  with  $f_{demo,d}$ , where  $f_{demo,d}$  is the demographic data of a district  $d$ .

2) **Dynamic Dependency Modeling:** To encode the spatio-temporal contextual dependency, we additionally measure the similarity score with time-variant node features.

**Traffic Patterns (P):** As temporal patterns also influence traffic accidents [2], [7], we form the traffic pattern dependency matrix by combining traffic conditions and dangerous driving statistics of each district. Traffic conditions are usually represented by only traffic volume and speed, but they are insufficient to express the roads' complex spatio-temporal correlations. Hence, we incorporate the dangerous driving statistics computed in Sec. IV into the traffic patterns so that the model can jointly learn more complex and dynamic spatio-temporal patterns between districts.

For each time  $t$ , we build

$$\mathcal{A}_{P,ij}^t = \begin{cases} 1 & sim(f_{p,d_i}^t, f_{p,d_j}^t) \geq \varepsilon_P \\ 0 & \text{otherwise,} \end{cases} \quad (5)$$

where  $f_{p,d}^t$  is traffic patterns, including traffic conditions and dangerous driving offenses, of a district  $d$  at a time interval  $t$ .

Given all contextual matrices, we create five graphs corresponding to  $\mathcal{A}_S$ ,  $\mathcal{A}_F$ ,  $\mathcal{A}_R$ ,  $\mathcal{A}_D$ , and  $\mathcal{A}_P$  that respectively represent five types of correlations—(1) spatial proximity, (2) district functionality, (3) road structure, (4) population, and (5) traffic patterns of all time steps—among regions.

### B. Graph Convolutional Networks

To capture the spatial dependency between the districts of each constructed graph, we adopt a graph convolutional network(GCN) [73] which has been successfully employed in the previous studies [2], [7], [8], [21]. We first follow the renormalization trick by calculating  $\hat{\mathcal{A}} = \tilde{\mathcal{D}}^{-\frac{1}{2}} \tilde{\mathcal{A}} \tilde{\mathcal{D}}^{-\frac{1}{2}}$  in a preprocessing step for all matrices. Here,  $\tilde{\mathcal{A}} = \mathcal{A} + I_N$  is the adjacency or contextual matrix with self-connections, and  $\tilde{\mathcal{D}}_{ii} = \sum_j \tilde{\mathcal{A}}_{ij}$  is the degree matrix corresponding to  $\tilde{\mathcal{A}}$ .

Then, for each normalized contextual matrix  $\hat{A}$  at a time interval  $t$ , a layer of the GCN is modeled as

$$h_{l+1}^t = \text{ReLU}(\hat{A}^t h_l^t W_l^t), \quad (6)$$

where  $h_l^t$  indicates the  $l$ -th layer hidden representation of graph convolution,  $W_l^t$  denotes the weights of the corresponding convolution kernels in a layer  $l$ , and  $h_0^t = \mathcal{F}^t$ . Finally, for each time  $t$ , we get the representation  $H_v^t$  from all GCN layers, where  $v \in \{S, F, R, D, P\}$ .

### C. Inter-View Attention Module

We adopt the idea of channel-wise attention [74] and generalize it for multi-view dependency modeling to learn the significance of different views and emphasize them by giving higher weights. Its constituents are as follows.

1) *Pooling Layer*: At each time interval  $t$ , we first use global average pooling  $G_{pool}$  over each view  $v$  (i.e., contextual information) to summarize it into a scalar  $z$ ,

$$z_v^t = G_{pool}(H_v^t) = \frac{1}{N} \sum_{i=1}^N H_{v,i}^t. \quad (7)$$

2) *Attention Layer*: To get attention scores, we concatenate each summarized view into a single vector  $\mathbf{z}$ . As defined in Eq. (8), we apply an attention operation to the vector  $\mathbf{z}$ . Then, we apply  $s$  to scale the weights of each view.

$$\begin{aligned} s &= \sigma(W_2, \text{ReLU}(W_1 \mathbf{z})) \\ H_{attn}^t &= H^t \circ s^t, \end{aligned} \quad (8)$$

where  $W_1$  and  $W_2$  are the corresponding weights,  $\sigma$  is the sigmoid activation function, and  $s$  is the vector of view-wise attention scores.

### D. Temporal Attention Module

After we get the hidden states of the weighted view representations from each temporal aspect, in this step, we capture which time interval contributes most to the predicted results. Thereby, we employ the Transformer [39] and gated recurrent unit (GRU) [75] layers, which are known to learn temporal dependency well [76], [77].

1) *Transformer Layer*: This layer highlights the significant period in the temporal input features through the multi-head self-attention mechanism. It is indeed helpful because a specific period (e.g., rush hour or holiday) of the dynamic information from different views can more significantly influence traffic accidents.

2) *GRU Layer*: This layer intends to capture the temporal patterns of the aggregated representations from the highlighted views and timesteps. The GRU, an improved version of the recurrent neural networks (RNN), is chosen here because it addresses the vanishing gradient problem [68] and has achieved state-of-the-art performance in sequential modeling [75].

Overall, the temporal attention module is formulated by

$$H = \text{GRU}(\text{Transformer}([H_{attn}^1, \dots, H_{attn}^T])). \quad (9)$$

### E. Prediction and Optimization

1) *Fully Connected Layer*: To map the learned representations  $H$  to a more separable feature space, we use fully-connected layers (FCs) for estimating future risk scores of each district. This layer learns how the historical representations of a district  $d$  during a time frame  $t (= 1, 2, \dots, T)$  affects the future traffic accident risk of  $d$  at a future time frame  $t' (= T + 1, T + 2, \dots, T + L)$ . Overall, the fully connected layer block is formulated by

$$\hat{y} = \text{FC}(\text{ReLU}(\text{FC}(H))), \quad (10)$$

where  $\hat{y} \in \mathbb{R}^{N \times L}$  represents the predicted risk scores at the future time steps of a length  $L$ , and the *ReLU* activation function is used for nonlinear transformation.

2) *Loss Function*: The loss function for MG-TAR needs to be carefully designed owing to the sparsity and rarity of traffic accidents. For example, training with an absolute-error loss makes a model produce mostly zeros (i.e., no accident predicted at all) to minimize the loss values. On the other hand, training with a squared-error loss, the model becomes sensitive to outliers (e.g., big crashes), so the output risk scores become unreasonably high. Therefore, we adopt a robust loss function *Huber loss*—an elegant compromise between squared error and absolute error losses—as it has proved its robustness for optimization in regression tasks [20], [78]. Formally, the loss function  $\mathcal{L}(y, \hat{y})$  is defined by

$$\mathcal{L}(y, \hat{y}) = \begin{cases} \frac{1}{2}(y - \hat{y})^2 & |y - \hat{y}| \leq \delta \\ \delta|y - \hat{y}| - \frac{1}{2}\delta^2 & \text{otherwise}, \end{cases} \quad (11)$$

where  $y$  and  $\hat{y}$  are the true and predicted risk scores, respectively, and  $\delta$  is typically tuned to achieve the best result.

## VI. EMPIRICAL STUDIES

This section presents the results from extensive empirical studies, including the comparison with baselines, hyperparameter study, ablation study, and case study. The source code is available at <https://github.com/kaist-dmlab/MG-TAR>.

### A. Data Description

This study uses real-world datasets collected from five metropolitan cities in South Korea for September 2016 and September 2018. The datasets contain the following four groups of features used to generate the node feature matrix (Def. 6) and construct the contextual matrix (Def. 7).

1) *Traffic Accident Risk Scores (1 Feature)*: This feature represents the traffic accident risk scores (see Def. 5) in each city district at a time point. We collect the raw numbers of historical traffic accidents from the TAAS.<sup>2</sup>

2) *Dangerous Driving Behavior (9 Features)*: By Def. 3, we obtain the numbers of the nine dangerous driving offenses that occurred in a city district at a particular time interval and use them as the dangerous driving behavior features.

<sup>2</sup><http://taas.koroad.or.kr/gis/mcm/mcl/initMap.do?menuId=0>

TABLE VI  
DETAILS OF THE HYPERPARAMETER SEARCH SPACE IN EACH MODEL

Model	Search Space
ARIMA	$p: [0, 1, 2, 3, 4, 5, 6]$ , $d: [0, 1, 2, 3, 4, 5, 6]$ , $q: [0, 1, 2, 3, 4, 5, 6]$
XGB	<b>Estimators:</b> [10, 25, 50, 100, 200], <b>Max Depth:</b> [2, 3, 5, 7, 10]
SdAE	<b>FC Units:</b> [20, 40, 60, 80, 100], <b>FC Layers:</b> [1, 2, 3, 4]
TARPML	<b>LSTM Units:</b> [100, 200, 300], <b>FC Units:</b> [128, 256, 512], <b>LSTM Layers:</b> [2, 3, 4], <b>FC Layers:</b> [2, 3, 4]
Hetero-ConvLSTM	<b>CNN Kernels:</b> [3, 4, 6], <b>CNN Filters:</b> [64, 128, 256]
TA-STAN	<b>GRU Units:</b> [100, 200], <b>Embedding Size:</b> [8, 16], <b>FC Units:</b> [128, 256]
RiskOracle	<b>GCN Layers:</b> [7, 9, 11], <b>GCN Filters:</b> [256, 384, 512], <b>FC Units:</b> [128, 256]
RiskSeq	<b>GCN Layers:</b> [4, 6, 8], <b>GCN Filters:</b> [192, 256, 320], <b>LSTM Layers:</b> [2, 3], <b>LSTM Units:</b> [192, 256]
DF-TAR	<b>GRU Units:</b> [256, 512], <b>Transfomer Units:</b> [512, 1024, 2048], $\theta: [0.25, 0.35, 0.45, 0.5]$
MG-TAR	<b>GRU Units:</b> [256, 512], <b>GCN Filters:</b> [128, 256], <b>FC Units:</b> [512, 1024], <b>GCN Layers:</b> [2, 4, 6, 8, 10, 12], $\delta: [1, 3, 5, 7, 9, 11, 13]$

3) *Static Environmental Features (98 Features)*: This feature group is further divided into three sets, as follows.

*Demographic Data*: This feature set contains 15 features describing the population distribution of ages, genders, and other characteristics in a district obtained from the Korean Statistical Information Service.<sup>3</sup>

*Point-of-Interest (POI) Data*: We collect the POI information from the Korean Local Information Research and Development Institute.<sup>4</sup> In total, we have 41 features representing the numbers of businesses (e.g., restaurants and shops), health facilities, and institutions in a district.

*Road Network and Specification*: We obtain the road network information from the Standard Node Link dataset<sup>5</sup> provided by Korea's Transportation Information Center. It consists of 42 features representing the number of lanes, the number of connections, the speed limit, the road type (e.g., common road and bridge), the road rank (e.g., city road and expressway), and so on.

4) *Dynamic Environmental Features (25 Features)*: This feature group is further divided into three sets, as follows.

*Weather and Air Quality Data*: This feature set represents district-wide real-time observed meteorological data provided by the Korea Meteorological Administration.<sup>6</sup> In total, we have 13 features that describe weather-related information such as temperature, humidity, and air quality index (e.g., PM2.5).

*Traffic Volume and Speed*: Using the driving log datasets in Table IV, we estimate the traffic volume from the numbers of unique vehicles in a particular district at a time interval. Similarly to DSTGCN [8], we compute the traffic speed by averaging the taxi driving speed data.

*Calendar Data*: This set contains 10 features representing the date, time, day of the week, and holiday information.

## B. Experimental Settings

1) *Implementation Details*: In our experiments, we select 80% and 20% of the datasets for training and testing, respectively. Additionally, 10% of the training set are used for validation. We executed the proposed model and all baselines with Scikit-Learn 0.24 and Tensorflow 2.4 libraries on Ubuntu

<sup>3</sup><https://kosis.kr/index/index.do>

<sup>4</sup><http://localdata.kr/>

<sup>5</sup><https://its.go.kr/nodelink/intro>

<sup>6</sup><https://data.kma.go.kr/cmmn/main.do>

18.04 LTS equipped with an NVIDIA GeForce RTX 2080 Ti GPU. All input features are standardized by min-max normalization. During the training, we adopt the Adam optimizer [79] (initial  $\alpha = 0.001$ ,  $\beta_1 = 0.9$ ,  $\beta_2 = 0.999$ , and  $\epsilon = 10^{-7}$ ) to minimize the loss function and train the model using the backpropagation process with the batch size of 32. Last, to avoid overfitting, we employ the learning rate decay and early-stopping techniques. We run each model *ten* times to ensure reproducibility and avoid occasional results, and the average and standard deviation of the results are reported.

2) *Hyperparameter Configurations*: The hyperparameters of each model for each dataset are tuned by running a grid search with the validation set based on Rec@K (Def. 13) scores because correctly predicting high-risk regions is more vital in practice. Table VI presents the entire hyperparameter search space for each model.

3) *Evaluation Metrics*: As commonly employed in recent studies [2], [7], [21], we use two evaluation metrics from both regression and classification perspectives to comprehensively evaluate the predictive models' performance.

*Regression*: We use the *Mean Squared Error (MSE)* to evaluate the overall predictive performance.

*Classification*: To examine the models' practical applicability, we also adopt the *Recall of top K (Rec@K)* used in the relevant studies. Here, Rec@K evaluates the ratio of accurate predictions for a prediction list of length  $K$ . For time interval  $t'$ , if there are traffic accidents in  $k_{t'}$  districts, Rec@K is the percentage of the intersection of true top- $k$  and predicted top- $k$  regions with the highest risk scores. In each dataset, the  $K$  districts are the top 20% districts with the highest risk scores.

$$MSE = \frac{1}{T'} \sum_{t'=T+1}^{T'} (y_{t'} - \hat{y}_{t'})^2, \quad (12)$$

$$Recall = \frac{1}{T'} \sum_{t'=T+1}^{T'} \frac{|R_{t'} \cap \hat{R}_{t'}|}{|R_{t'}|}, \quad (13)$$

where  $y_{t'}$  and  $\hat{y}_{t'}$  are the true and predicted risk scores of all districts at time interval  $t'$ , respectively; and  $T' = T + L$  is the prediction length.  $R_{t'}$  and  $\hat{R}_{t'}$  are the sets of actual and predicted top  $K = |R_{t'}|$  highest risk districts, respectively. Lower MSE values mean the model predicts more accurate risks in all regions, while higher Rec@K values indicate the model performs better in figuring out more high-risk districts.

TABLE VII  
PERFORMANCE COMPARISON FOR EACH CITY WITH AVERAGED SCORES AND RANKS IN THE SEPT-2016 DATASET

Model / City	Seoul		Busan		Daegu		Daejeon		Gwangju		Average MSE↓	Rec@20%↑	Rank↓
	MSE↓	Rec@5↑	MSE↓	Rec@3↑	MSE↓	Rec@2↑	MSE↓	Rec@1↑	MSE↓	Rec@1↑			
HA	2.8567 (±0.00)	26.86% (±0.00)	1.7276 (±0.00)	26.28% (±0.00)	2.7823 (±0.00)	33.87% (±0.00)	3.1086 (±0.00)	19.77% (±0.00)	4.7827 (±0.00)	20.82% (±0.00)	3.0516 (±0.00)	25.52% (±0.00)	8
ARIMA	2.8567 (±0.00)	27.01% (±0.00)	1.7276 (±0.00)	27.51% (±0.00)	2.7823 (±0.00)	33.20% (±0.00)	3.1086 (±0.00)	19.77% (±0.00)	4.7827 (±0.00)	19.37% (±0.00)	3.0516 (±0.00)	25.37% (±0.00)	9.5
XGB	2.8944 (±0.00)	28.37% (±0.00)	1.7894 (±0.00)	27.86% (±0.00)	2.5793 (±0.00)	36.62% (±0.00)	2.9053 (±0.00)	24.10% (±0.00)	4.8132 (±0.00)	28.49% (±0.00)	2.9963 (±0.00)	29.09% (±0.00)	5
SdAE	2.7728 (±0.02)	28.29% (±0.02)	<b>1.3987</b> (±0.02)	26.38% (±0.03)	2.5395 (±0.06)	30.80% (±0.04)	3.1279 (±0.07)	25.93% (±0.04)	4.7446 (±0.04)	<b>31.11%</b> (±0.03)	2.9167 (±0.04)	28.50% (±0.03)	3.5
TARPML	2.7495 (±0.03)	26.48% (±0.02)	1.6489 (±0.01)	27.55% (±0.02)	2.6290 (±0.03)	34.63% (±0.05)	3.0387 (±0.05)	<b>26.30%</b> (±0.03)	4.8000 (±0.06)	26.52% (±0.03)	2.9732 (±0.04)	28.30% (±0.03)	5
Hetero-ConvLSTM	2.7423 (±0.04)	24.80% (±0.03)	2.0488 (±0.75)	27.99% (±0.06)	2.6271 (±0.06)	33.85% (±0.05)	2.9214 (±0.08)	22.39% (±0.05)	4.5299 (±0.08)	25.65% (±0.05)	2.9739 (±0.20)	26.94% (±0.05)	6
TA-STAN	2.6713 (±0.10)	31.03% (±0.02)	1.6483 (±0.05)	30.51% (±0.05)	2.6625 (±0.15)	34.31% (±0.06)	2.9750 (±0.19)	23.68% (±0.03)	5.3081 (±0.17)	30.24% (±0.04)	3.0530 (±0.13)	29.95% (±0.04)	6
RiskOracle	3.0949 (±0.03)	28.04% (±0.01)	1.7783 (±0.02)	21.73% (±0.02)	2.9705 (±0.05)	29.61% (±0.02)	3.1721 (±0.07)	17.66% (±0.03)	4.8582 (±0.10)	28.19% (±0.03)	3.1748 (±0.05)	25.04% (±0.02)	11
RiskSeq	2.9464 (±0.03)	28.34% (±0.02)	1.7069 (±0.01)	18.01% (±0.02)	2.7818 (±0.04)	31.55% (±0.02)	3.0057 (±0.03)	21.91% (±0.02)	4.5768 (±0.04)	27.23% (±0.03)	3.0035 (±0.03)	25.41% (±0.02)	8
DF-TAR	2.8247 (±0.03)	31.44% (±0.01)	1.6473 (±0.01)	32.96% (±0.02)	2.6608 (±0.04)	35.19% (±0.05)	2.9894 (±0.05)	22.66% (±0.03)	4.5592 (±0.04)	27.37% (±0.03)	2.9363 (±0.03)	29.92% (±0.02)	3
<b>MG-TAR</b>	<b>2.6051</b> (±0.00)	<b>32.45%</b> (±0.01)	1.6086 (±0.00)	<b>33.80%</b> (±0.02)	<b>2.4724</b> (±0.00)	<b>41.14%</b> (±0.00)	<b>2.8075</b> (±0.00)	24.78% (±0.00)	<b>4.4050</b> (±0.00)	27.39% (±0.00)	<b>2.7797</b> (±0.01)	<b>31.91%</b> (±0.01)	<b>1</b>

### C. Baselines

We compare MG-TAR with the following *ten* competitive baselines. For a fair comparison, all baselines predict the risk scores in the *next 6 time steps* using the *12 previous time steps* of all related input features as in the relevant study [7]. The best hyperparameters for each dataset are chosen using the grid search with the validation set.

- 1) **Historical Average (HA)** models the traffic accident risk as a seasonal process and uses the average of previous seasons as the prediction.
- 2) **ARIMA** [80] or autoregressive integrated moving average model is a classic time-series analysis method for predicting future values. It fits previous traffic accident risk data to predict the future risk scores.
- 3) **eXtreme Gradient Boosting (XGB)** [81] is an optimized distributed gradient boosting library based on a gradient boosting tree commonly used for prediction tasks. XGB and ARIMA use the same features as the input.
- 4) **SdAE** [60] uses stacked denoising autoencoders to learn hierarchical feature representations of the inputs. It utilizes human mobility data(i.e., traffic volume) for real-time risk prediction. Then, the learned features are fed to a linear model to predict the risk of traffic accidents.
- 5) **TARPML** [63] is a deep learning method that stacks LSTM networks for multi-step prediction. We train it with sequences of past risk scores as in the original paper.
- 6) **Hetero-ConvLSTM** [65] is a deep learning method for traffic accident prediction. It uses ConvLSTM networks to learn the spatio-temporal data simultaneously with newly introduced spatial graph features.
- 7) **TA-STAN** [67] is a deep neural network with spatial and temporal attention mechanisms to model the hierarchical impact of traffic-related features for citywide traffic accident risk prediction.

- 8) **RiskOracle** [2] is an advanced multi-task deep learning framework for traffic accident risk prediction based on differential time-varying GCNs.
- 9) **RiskSeq** [7] is an extended version of RiskOracle for multi-step traffic accident risk prediction with the context-guided LSTM decoder.
- 10) **DF-TAR** [76] is a recent deep learning model for traffic accident risk prediction with a fusion mechanism. Note that DF-TAR is our earlier work introducing the use of dangerous driving behavior for this problem.

### D. Experimental Results

1) *Performance Comparison:* Table VII and Table VIII show the comparison results of our MG-TAR with baseline models. Finding a good trade-off between regression error and classification accuracy by keeping the ranks of the risk scores is not trivial since accidents are sporadic events, and a simple predictive model can easily fall into the zero-inflated issue [2], [7]. Unlike the baselines, MG-TAR successfully balances the trade-off and achieves higher or comparable results in both the regression and classification tasks by adopting multi-view graphs learning from dangerous driving behavior. The specific results for each task are summarized as follows.

- **Traffic Accident Risk Prediction.** From the regression perspective, on the average for all datasets, MG-TAR achieves the best result in terms of the MSE with the improvement of up to 23% over the baselines. In general, we can observe that HA and ARIMA perform similarly. Expectedly, most of the deep learning baselines such as SdAE, TARPML, Hetero-ConvLSTM, and TA-STAN outperform the XGB model as they can generalize for the complex and heterogeneous data better than the traditional approaches. Unexpectedly, RiskOracle and RiskSeq underperform other deep learning models, probably because these models adopt data augmentation (i.e., converting zeros into negative values) and try to

TABLE VIII  
PERFORMANCE COMPARISON FOR EACH CITY WITH AVERAGED SCORES AND RANKS IN THE SEPT-2018 DATASET

Model / City	Seoul		Busan		Daegu		Daejeon		Gwangju		Average		
	MSE↓	Rec@5↑	MSE↓	Rec@3↑	MSE↓	Rec@2↑	MSE↓	Rec@1↑	MSE↓	Rec@1↑	MSE↓	Rec@20%↑	Rank↓
HA	2.7082 (±0.00)	22.44% (±0.00)	1.3523 (±0.00)	21.27% (±0.00)	3.4489 (±0.00)	33.94% (±0.00)	3.0096 (±0.00)	22.11% (±0.00)	2.8308 (±0.00)	35.89% (±0.00)	2.6700 (±0.00)	27.13% (±0.00)	7.5
	2.7082 (±0.00)	22.71% (±0.00)	1.3523 (±0.00)	20.97% (±0.00)	3.4489 (±0.00)	34.37% (±0.00)	3.0096 (±0.00)	21.63% (±0.00)	2.8308 (±0.00)	35.89% (±0.00)	2.6700 (±0.00)	27.11% (±0.00)	8.5
ARIMA	3.3325 (±0.00)	23.69% (±0.00)	1.4235 (±0.00)	24.10% (±0.00)	4.1775 (±0.00)	30.05% (±0.00)	4.3601 (±0.00)	22.63% (±0.00)	2.7533 (±0.00)	30.36% (±0.00)	3.2094 (±0.00)	26.16% (±0.00)	11
	2.6393 (±0.03)	25.71% (±0.02)	<b>1.2247</b> (±0.01)	22.96% (±0.03)	<b>3.1031</b> (±0.03)	34.31% (±0.03)	<b>2.5639</b> (±0.04)	23.18% (±0.03)	2.9992 (±0.03)	34.39% (±0.03)	2.5060 (±0.02)	28.11% (±0.03)	3.5
TARPML	2.6182 (±0.03)	25.57% (±0.01)	1.3307 (±0.02)	23.09% (±0.02)	3.3390 (±0.04)	33.82% (±0.03)	2.9558 (±0.05)	24.53% (±0.03)	2.8901 (±0.08)	30.97% (±0.03)	2.6268 (±0.04)	27.59% (±0.03)	6
	2.9831 (±0.48)	24.82% (±0.04)	1.2978 (±0.02)	23.68% (±0.05)	3.2645 (±0.08)	34.27% (±0.06)	2.9038 (±0.10)	21.64% (±0.08)	2.6880 (±0.07)	28.36% (±0.08)	2.6274 (±0.15)	26.55% (±0.06)	8
Hetero-ConvLSTM	2.6054 (±0.14)	27.65% (±0.02)	1.3270 (±0.13)	27.08% (±0.03)	4.2052 (±0.18)	30.86% (±0.03)	3.7687 (±0.25)	19.93% (±0.03)	2.8456 (±0.21)	35.30% (±0.18)	2.9504 (±0.03)	28.16% (±0.03)	6
	3.1049 (±0.07)	24.81% (±0.01)	1.4172 (±0.03)	21.49% (±0.02)	4.1286 (±0.17)	33.58% (±0.02)	3.2701 (±0.07)	21.79% (±0.03)	2.8558 (±0.14)	<b>38.97%</b> (±0.04)	2.9553 (±0.09)	28.13% (±0.03)	7
RiskOracle	2.7267 (±0.03)	25.41% (±0.01)	1.3182 (±0.01)	20.72% (±0.02)	3.3488 (±0.04)	37.74% (±0.02)	2.9745 (±0.03)	19.04% (±0.04)	2.6398 (±0.04)	36.17% (±0.04)	2.6016 (±0.03)	27.81% (±0.03)	5
	2.7267 (±0.03)	25.41% (±0.01)	1.3182 (±0.01)	20.72% (±0.02)	3.3488 (±0.04)	37.74% (±0.02)	2.9745 (±0.03)	19.04% (±0.04)	2.6398 (±0.04)	36.17% (±0.04)	2.6016 (±0.03)	27.81% (±0.03)	5
DF-TAR	2.6142 (±0.03)	27.65% (±0.01)	1.2574 (±0.00)	28.06% (±0.03)	3.1912 (±0.02)	40.03% (±0.01)	2.8973 (±0.04)	21.02% (±0.03)	<b>2.5872</b> (±0.03)	35.40% (±0.04)	2.5094 (±0.02)	30.43% (±0.02)	2.5
	<b>2.5029</b> (±0.01)	<b>28.58%</b> (±0.02)	1.2506 (±0.00)	<b>31.36%</b> (±0.02)	3.1457 (±0.00)	<b>40.24%</b> (±0.00)	2.7693 (±0.01)	<b>33.81%</b> (±0.01)	2.5984 (±0.00)	31.59% (±0.01)	<b>2.4534</b> (±0.01)	<b>33.11%</b> (±0.01)	1

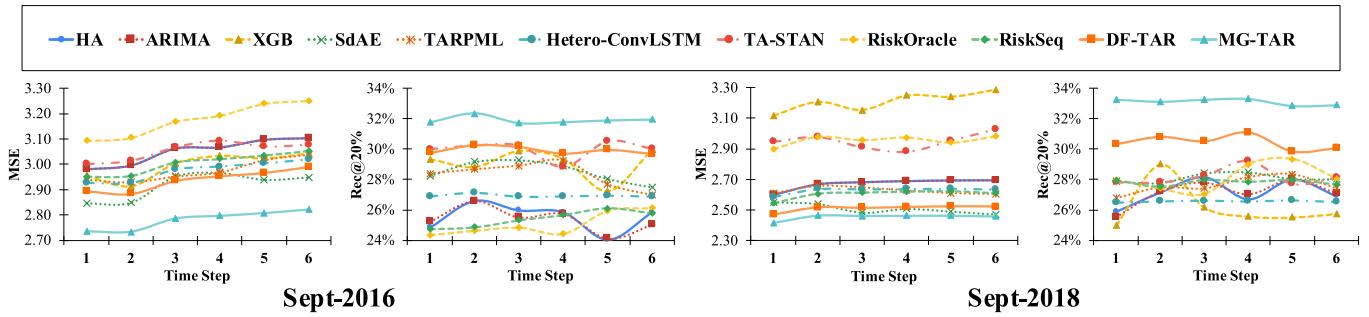


Fig. 5. Step-wise performance on MSE (lower is better) and Rec@K (higher is better) evaluation metrics.

preserve the rank of traffic accident risk scores. The RiskOracle is worse than RiskSeq as it was designed for a single-step prediction, whereas the RiskSeq is for a multi-step prediction.

- **Highest Risk Region Selection.** In the spatial classification task for determining the top- $k$  *highest risk* (i.e., most dangerous) districts, our MG-TAR achieves the highest results in Rec@K with the improvement of up to 27%, outperforming the other baselines due to the consideration of dangerous driving behavior. Note that the models such as SdAE, TARPML, and Hetero-ConvLSTM performing well in the regression tasks typically do not work well, probably because of the zero-inflated issue [2], [7], [66]. Here, HA, ARIMA, and Hetero-ConvLSTM perform similarly. We conjecture that the degradation of Hetero-ConvLSTM stems from the infeasibility of constructing spatial ensemble learning in irregular graph structures. TA-STAN performs moderately well due to its spatial and temporal attention modules. Although there are improvements of RiskOracle and RiskSeq, especially in Sept-2018, they still cannot achieve 30% of Rec@K, meaning that the data augmentation technique may not be

the right option here. On average, only DF-TAR and MG-TAR can reach the accuracy of at least 30% in Rec@K by incorporating the dangerous driving statistics.

- **Step-wise Performance.** To evaluate the long-term stability of the compared models, we visualize their step-wise error and accuracy for the upcoming six time steps in Fig. 5. Thanks to the effectiveness in learning the spatio-temporal dependency between districts and modeling sequential data, MG-TAR maintains its dominance, consistently showing the best results in all tasks during the six time steps.

2) *Hyperparameter Study:* To illustrate how different hyperparameters affect the performance of MG-TAR, we show the results of different settings on the following four aspects.

- **Number of GCN Layers.** We adjust the number of GCN blocks to observe the impact of model complexity. As in Tables IX and X, we can see that different cities with different timeframes require different complexity due to their size, heterogeneity, and temporal dynamics. Moreover, on average, there are only slight deviations in the model's performance—lower than 0.11% difference in MSE and 0.19% difference in Rec@K—to

TABLE IX  
VARYING THE NUMBER OF GCN LAYERS FOR THE SEPT-2016 DATASET

City # GCN	Seoul		Busan		Daegu		Daejeon		Gwangju	
	MSE↓	Rec@5↑	MSE↓	Rec@3↑	MSE↓	Rec@2↑	MSE↓	Rec@1↑	MSE↓	Rec@1↑
2	<b>2.6024</b>	32.30%	1.6079	32.90%	2.4720	41.14%	<b>2.8036</b>	<b>24.77%</b>	4.4103	27.23%
4	2.6030	32.69%	1.6081	33.58%	2.4728	41.14%	2.8071	24.77%	4.4032	27.24%
6	2.6032	32.58%	<b>1.6076</b>	<b>33.68%</b>	<b>2.4695</b>	<b>41.14%</b>	2.8061	24.77%	<b>4.4008</b>	<b>27.48%</b>
8	2.6093	32.29%	1.6076	33.65%	2.4758	41.14%	2.8073	24.77%	4.4037	26.99%
10	2.6078	32.03%	1.6085	34.25%	2.4709	41.14%	2.8037	24.77%	4.4054	27.24%
12	2.6086	<b>32.74%</b>	1.6083	33.65%	2.4727	41.14%	2.8036	24.77%	4.4047	27.48%

TABLE X  
VARYING THE NUMBER OF GCN LAYERS FOR THE SEPT-2018 DATASET

City # GCN	Seoul		Busan		Daegu		Daejeon		Gwangju	
	MSE↓	Rec@5↑	MSE↓	Rec@3↑	MSE↓	Rec@2↑	MSE↓	Rec@1↑	MSE↓	Rec@1↑
2	2.4989	28.52%	1.2494	32.44%	3.1466	40.24%	<b>2.7648</b>	<b>33.81%</b>	2.5987	31.57%
4	2.5012	28.77%	<b>1.2491</b>	<b>32.44%</b>	3.1560	40.24%	2.7680	33.81%	2.6004	31.57%
6	2.5022	29.23%	1.2506	32.44%	<b>3.1465</b>	<b>40.24%</b>	2.7726	31.11%	2.5981	31.57%
8	2.5020	<b>29.50%</b>	1.2513	30.97%	3.1485	40.24%	2.7656	33.81%	2.5997	<b>31.81%</b>
10	<b>2.4986</b>	29.44%	1.2495	32.44%	3.1470	40.24%	2.7690	31.11%	2.6030	31.57%
12	2.4995	27.76%	1.2503	31.95%	3.1504	40.24%	2.7653	33.81%	<b>2.5963</b>	31.57%

TABLE XI  
VARYING THE SIMILARITY FUNCTION

Function	MSE↓	Rec@20%↑
Cosine Similarity	2.6178	32.29%
Jaccard Similarity	<b>2.6140</b>	<b>32.83%</b>
Euclidean Distance	2.6192	31.53%
Cityblock Distance	2.6158	32.61%
Pearson Correlation	2.6166	32.66%

TABLE XII  
VARYING THE INPUT SEQUENCE LENGTH

Sequence Length	MSE↓	Rec@20%↑
6	2.6205	31.90%
12	<b>2.6149</b>	<b>32.48%</b>
18	2.6157	31.93%

hyperparameter changes. It indicates the robustness of the proposed model to hyperparameter settings.

- Similarity Metric.** Table XI shows the performance comparison on different similarity functions. MG-TAR consistently learns the dependency between regions well and achieves the comparable accuracy results regardless of the similarity function used. On average, the Jaccard similarity works best among all functions, and it is used as the default similarity function in our model.
- Input Sequence Length.** To examine how the number of historical observations affects the prediction performance, we vary the input sequence length. As shown in Table XII, to predict the following 6 hours using the previous 12 hours was found to be the optimal length in this work. A shorter length may not give enough information, while a longer length could introduce noises.
- Train-Test Ratios.** We also investigate the influences of the proportion of the training samples versus the testing

TABLE XIII  
VARYING THE TRAINING RATIO

Training Data Ratio	MSE↓	Rec@20%↑
0.8	<b>2.6131</b>	<b>32.91%</b>
0.7	2.6170	32.63%
0.6	2.6194	31.29%
0.5	2.6184	31.37%
0.4	2.6183	31.00%

samples. Our MG-TAR model performs best when the train ratio is 80% (i.e., 8:2), as presented in Table XIII. It is consistent with the fact that more training examples can better capture deep spatio-temporal correlation and lessen epistemic uncertainty.

3) *Ablation Study:* To assess the MG-TAR's elements, we perform the ablation studies with the following three aspects.

- Effect of Feature Set.** Table XIV shows the usefulness of various types of the *node features* sets, introduced in Sec. VI-A, in constructing a node feature matrix and a contextual matrix in MG-TAR. The performance of MG-TAR tends to improve as more features are added because the *multi-view graph* structure is able to embed complementary information well. Specifically, the dangerous driving behavior was the most effective for improving the performance of MG-TAR.
- Effect of Multi-view Graph Scheme.** The contribution of each view of the spatio-temporal graph is presented in Table XV. While the model with the whole multi-view graph results in the best performance, the *functional similarity* and *traffic patterns* views contribute the most to the learning process.
- Effect of Attention Mechanism.** Table XVI shows the evaluation result of the attention mechanism. Both the

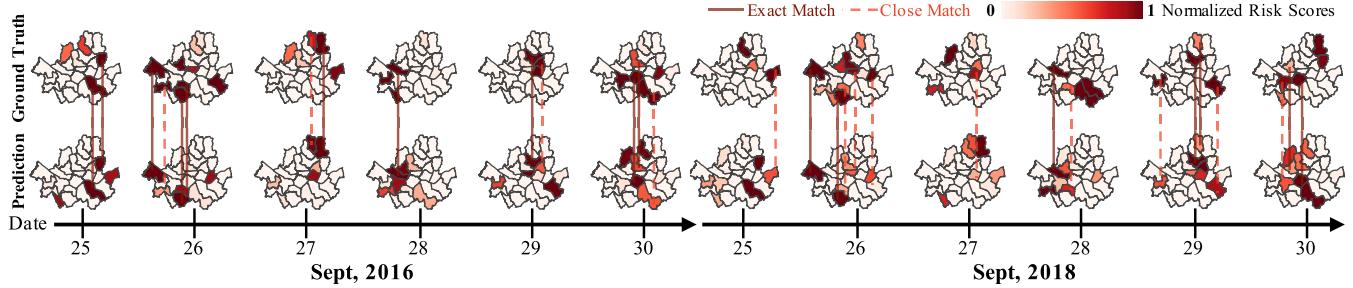


Fig. 6. Choropleth maps from the entire test datasets (six days) of Seoul.

TABLE XIV

EFFECT OF THE FEATURE SETS ON THE PERFORMANCE

Feature Set	MSE $\downarrow$	Rec@20% $\uparrow$
Historical Risk Scores (H)	2.6201	31.69%
H + Static Features (SF)	2.6151	31.93%
H + SF + Dynamic Features	2.6173	32.44%
H + Dangerous Driving Behavior	2.6162	32.58%
All Features	<b>2.6149</b>	<b>32.85%</b>

TABLE XV  
EFFECT OF THE VIEWS ON THE PERFORMANCE

Contextual View	MSE $\downarrow$	Rec@20% $\uparrow$
w/o Spatial Closeness	2.6164	32.46%
w/o Functional Similarity	2.6181	32.38%
w/o Road Similarity	2.6165	32.46%
w/o Demograph Similarity	2.6177	32.23%
w/o Traffic Patterns	2.6168	31.83%
Only Spatial Closeness	2.6188	31.88%
Full Multi-view Graphs	<b>2.6138</b>	<b>32.71%</b>

TABLE XVI  
EFFECT OF THE MULTI-ATTENTION ON THE PERFORMANCE

Module	MSE $\downarrow$	Rec@20% $\uparrow$
w/o Inter-view Attention	2.6344	31.78%
w/o Temporal Attention	2.6181	31.10%
All Attention Modules	<b>2.6168</b>	<b>32.98%</b>

inter-view attention and the temporal attention are proven to be necessary to balance the trade-off between classification accuracy and regression capacity; the *temporal attention* helps reducing error for high-risk district prediction, while the *inter-view attention* helps to capture the global contexts of accident risk distribution.

### E. Efficiency of MG-TAR

We further examine the efficiency of MG-TAR measured by the computation time. All models are run with the settings described in Section VI-B. Table XVII presents the time cost of training and inferencing for our model and other end-to-end deep learning baselines averaged over the ten datasets. Even though our model is not the most efficient one with only slight differences, given the higher model complexity and

TABLE XVII  
COMPUTATION TIME COMPARISON

Model	Computation Time	
	Training (s/epoch)	Inference (s/sample)
TARPM	0.216	0.005
Hetero-ConvLSTM	10.912	0.147
TA-STAN	0.600	0.004
RiskOracle	0.224	0.005
RiskSeq	1.624	0.025
DF-TAR	0.704	0.010
MG-TAR	2.552	0.039

larger input feature sets, MG-TAR is still fast enough to be applied in a real-time application. Specifically, it takes merely 39ms to predict a sample containing the 12 previous time steps observations for the multi-step citywide prediction. Note that we can train each graph asynchronously in parallel to enhance the model efficiency.

### F. Case Study

To evaluate the proposed model's practicality, in Fig. 6, we visualize choropleth maps of the *entire* six-day testing period by aggregating the hourly predicted results into daily-level predictions to make the results more practical in real-world scenarios. We compare the predicted traffic accident risk scores with the ground truth scores of both Sept-2016 and Sept-2018 datasets in Seoul consisting of 25 districts. Since the range of risk scores varies day by day, we normalize the scores so that they are easy to be visually compared to each other.

In addition to the zero-risk and low-risk cases that are correctly predicted by MG-TAR, it also precisely predicts the districts that have the high-risk scores each day, as shown in Fig. 6. Accurately predicting even only *one* very high-risk region is crucial because we can potentially prevent incalculable losses caused by an accident in that area. In particular, MG-TAR exactly matches at least one of the highest risk districts ten days out of twelve days. However, there are some missed cases where the predicted results are usually the adjacent areas. We conjecture that this incorrect prediction stems from some dangerous driving data probably collected near two regions' borders while moving from one area to another. Consequently, the accident would happen in the adjacent district instead after those offenses had occurred.

As presented in this case study, we are confident that MG-TAR can successfully predict the risk scores by learning dynamic patterns of dangerous driving statistics, historical traffic accident risks, and environmental features accordingly.

## VII. LIMITATIONS

Due to the regulations in acquiring digital tachograph data, we used only one-month datasets of five cities collected in 2016 and 2018 (i.e., ten months in total) to train (24 days) and test (6 days) the predictive models. Although this data is fairly large and the results are promising, if more massive data were available, we could make the correlation analysis at a higher level (e.g., between months or years) and obtain more comprehensive findings. The predictive models would be more precise to forecast future traffic accidents in longer prediction lengths (e.g., several days) by extending the length of datasets to several months or years. By doing so, the relationships between heterogeneous features will be easier to capture, particularly dangerous driving statistics and historical accidents. Additionally, with more extensive data, especially for deep learning-based approaches, the model would become more generalizable.

## VIII. CONCLUSION

This paper introduces a deep learning-based framework for citywide traffic accident risk prediction, called MG-TAR. First, we show the meaningful relationship between dangerous driving behavior and traffic accidents by temporal and geographical correlation studies. Then, we develop a novel multi-view graph convolutional network with multi-attention module for accident risk prediction with dangerous driving statistics, heterogeneous features, and multi-view contextual matrices. These contextual matrices, both static and dynamic, are designed to learn important information from various aspects between districts in a city with the inter-view attention mechanism. Thorough experiments demonstrate the superiority of MG-TAR over baseline models with the improvement of up to 23% in MSE for predicting the accident risk scores, as well as with the improvement of up to 27% in Rec@K for predicting the most dangerous districts. Overall, we believe that the proposed framework can be practically used for various intelligent transportation systems.

## REFERENCES

- [1] *Global Status Report on Road Safety 2018: Summary*, World Health Organization, Geneva, Switzerland, 2018.
- [2] Z. Zhou, Y. Wang, X. Xie, L. Chen, and H. Liu, "RiskOracle: A minute-level citywide traffic accident forecasting framework," in *Proc. 34th AAAI Conf. Artif. Intell.*, 2020, pp. 1258–1265.
- [3] H. Yao, X. Tang, H. Wei, G. Zheng, and Z. Li, "Revisiting spatial-temporal similarity: A deep learning framework for traffic prediction," in *Proc. 33rd AAAI Conf. Artif. Intell.*, 2019, pp. 5668–5675.
- [4] H. Yao, Y. Liu, Y. Wei, X. Tang, and Z. Li, "Learning from multiple cities: A meta-learning approach for spatial-temporal prediction," in *Proc. World Wide Web Conf. (WWW)*, 2019, pp. 2181–2191.
- [5] C. Song, Y. Lin, S. Guo, and H. Wan, "Spatial-temporal synchronous graph convolutional networks: A new framework for spatial-temporal network data forecasting," in *Proc. 34th AAAI Conf. Artif. Intell.*, 2020, pp. 914–921.
- [6] S. Guo, Y. Lin, H. Wan, X. Li, and G. Cong, "Learning dynamics and heterogeneity of spatial-temporal graph data for traffic forecasting," *IEEE Trans. Knowl. Data Eng.*, vol. 34, no. 11, pp. 5415–5428, Nov. 2022.
- [7] Z. Zhou, Y. Wang, X. Xie, L. Chen, and C. Zhu, "Foresee urban sparse traffic accidents: A spatiotemporal multi-granularity perspective," *IEEE Trans. Knowl. Data Eng.*, vol. 34, no. 8, pp. 3786–3799, Aug. 2022.
- [8] L. Yu, B. Du, X. Hu, L. Sun, L. Han, and W. Lv, "Deep spatio-temporal graph convolutional network for traffic accident prediction," *Neurocomputing*, vol. 423, pp. 135–147, Jan. 2021.
- [9] I. Almeida, R. C. P. Leal-Toledo, E. M. Toledo, D. C. Cacau, and G. Magalhães, "Drivers' behavior effects in the occurrence of dangerous situations which may lead to accidents," in *Proc. 13th Int. Conf. Cellular Automata Res. Ind.* Cham, Switzerland: Springer, 2018, pp. 441–450.
- [10] J. M. Scanlon, R. Sherony, and H. C. Gabler, "Models of driver acceleration behavior prior to real-world intersection crashes," *IEEE Trans. Intell. Transp. Syst.*, vol. 19, no. 3, pp. 774–786, Mar. 2018.
- [11] Korea Road Traffic Authority. (2020). *What Causes Traffic Accidents*. [Online]. Available: [https://www.koroad.or.kr/en\\_web/view/trfEnv4.do](https://www.koroad.or.kr/en_web/view/trfEnv4.do)
- [12] *Traffic Safety Act of the Republic of Korea*, Korean National Police Agency, Seoul South Korea, 2014.
- [13] G. Baldini, L. Sportiello, M. Chiaramello, and V. Mahieu, "Regulated applications for the road transportation infrastructure: The case study of the smart tachograph in the European union," *Int. J. Crit. Infrastruct. Protection*, vol. 21, pp. 3–21, Jun. 2018.
- [14] J. M. Vincent and C. Threewitt. (Feb. 2018). *How do Those Car Insurance Tracking Devices Work?* U.S. News & World Report. [Online]. Available: <https://cars.usnews.com/cars-trucks/car-insurance/how-do-those-car-insurance-tracking-devices-work>
- [15] P. Stenquist. (Jul. 2020). *Letting Your Insurer Ride Shotgun, for a Discounted Rate*. the New York Times Company. [Online]. Available: <https://www.nytimes.com/2020/07/16/business/car-insurance-app-discounts.html>
- [16] P. G. Young. (Mar. 2014). *What if You Could Give a Driver With a Good Driving Habit a Discount on His Car Insurance?* Korea Road Traffic Authority. [Online]. Available: [https://www.koroad.or.kr/kp\\_web/krPrView.do?board\\_code=GABBS\\_050&board\\_num=126843](https://www.koroad.or.kr/kp_web/krPrView.do?board_code=GABBS_050&board_num=126843)
- [17] M. R. Khan and J. E. Blumenstock, "Multi-GCN: Graph convolutional networks for multi-view networks, with applications to global poverty," in *Proc. 33rd AAAI Conf. Artif. Intell.*, 2019, pp. 606–613.
- [18] X. Geng et al., "Spatiotemporal multi-graph convolution network for ride-hailing demand forecasting," in *Proc. 33rd AAAI Conf. Artif. Intell.*, 2019, pp. 3656–3663.
- [19] Y. Xie, Y. Zhang, M. Gong, Z. Tang, and C. Han, "MGAT: Multi-view graph attention networks," *Neural Netw.*, vol. 132, pp. 180–189, Dec. 2020.
- [20] J. Sun, J. Zhang, Q. Li, X. Yi, Y. Liang, and Y. Zheng, "Predicting citywide crowd flows in irregular regions using multi-view graph convolutional networks," *IEEE Trans. Knowl. Data Eng.*, vol. 34, no. 5, pp. 2348–2359, May 2022.
- [21] B. Wang, Y. Lin, S. Guo, and H. Wan, "GSNet: Learning spatial-temporal correlations from geographical and semantic aspects for traffic accident risk forecasting," in *Proc. 35th AAAI Conf. Artif. Intell.*, 2021, pp. 4402–4409.
- [22] H. Yan, X. Ma, and Z. Pu, "Learning dynamic and hierarchical traffic spatiotemporal features with transformer," *IEEE Trans. Intell. Transp. Syst.*, vol. 23, no. 11, pp. 22386–22399, Nov. 2022.
- [23] Y. Wang, S. Fang, C. Zhang, S. Xiang, and C. Pan, "TVGCN: Time-variant graph convolutional network for traffic forecasting," *Neurocomputing*, vol. 471, pp. 118–129, Jan. 2022.
- [24] T. Zhang and G. Guo, "Graph attention LSTM: A spatiotemporal approach for traffic flow forecasting," *IEEE Intell. Transp. Syst. Mag.*, vol. 14, no. 2, pp. 190–196, Mar. 2022.
- [25] Guo, Ge, and Wei Yuan, "Short-term traffic speed forecasting based on graph attention temporal convolutional networks," *Neurocomputing*, vol. 410, no. 14, pp. 387–393, Oct. 2020.
- [26] C.-H. Wu, J.-M. Ho, and D. T. Lee, "Travel-time prediction with support vector regression," *IEEE Trans. Intell. Transp. Syst.*, vol. 5, no. 4, pp. 276–281, Dec. 2004.
- [27] L. Moreira-Matias, J. Gama, M. Ferreira, J. Mendes-Moreira, and L. Damas, "Predicting taxi-passenger demand using streaming data," *IEEE Trans. Intell. Transp. Syst.*, vol. 14, no. 3, pp. 1393–1402, Sep. 2013.

- [28] Z. Zheng and D. Su, "Short-term traffic volume forecasting: A k-nearest neighbor approach enhanced by constrained linearly sewing principle component algorithm," *Transp. Res. C, Emerg. Technol.*, vol. 43, no. 1, pp. 143–157, Jun. 2014.
- [29] Y. Xu, Q.-J. Kong, R. Klette, and Y. Liu, "Accurate and interpretable Bayesian MARS for traffic flow prediction," *IEEE Trans. Intell. Transp. Syst.*, vol. 15, no. 6, pp. 2457–2469, Dec. 2014.
- [30] P.-T. Chen, F. Chen, and Z. Qian, "Road traffic congestion monitoring in social media with hinge-loss Markov random fields," in *Proc. IEEE Int. Conf. Data Mining*, Dec. 2014, pp. 80–89.
- [31] H. Yao et al., "Deep multi-view spatial-temporal network for taxi demand prediction," in *Proc. 32nd AAAI Conf. Artif. Intell.*, 2018, pp. 2588–2595.
- [32] S. Wang, J. Cao, and P. S. Yu, "Deep learning for spatio-temporal data mining: A survey," *IEEE Trans. Knowl. Data Eng.*, vol. 34, no. 8, pp. 3681–3700, Aug. 2022.
- [33] G. Guo and T. Zhang, "A residual spatio-temporal architecture for travel demand forecasting," *Transp. Res. C, Emerg. Technol.*, vol. 115, Jun. 2020, Art. no. 102639.
- [34] K. Jin, J. Wi, E. Lee, S. Kang, S. Kim, and Y. Kim, "TrafficBERT: Pre-trained model with large-scale data for long-range traffic flow forecasting," *Expert Syst. Appl.*, vol. 186, Dec. 2021, Art. no. 115738.
- [35] Y. Li, R. Yu, C. Shahabi, and Y. Liu, "Diffusion convolutional recurrent neural network: Data-driven traffic forecasting," in *Proc. 6th Int. Conf. Learn. Represent.*, 2018.
- [36] B. Yu, H. Yin, and Z. Zhu, "Spatio-temporal graph convolutional networks: A deep learning framework for traffic forecasting," in *Proc. 27th Int. Joint Conf. Artif. Intell.*, Jul. 2018, pp. 3634–3640.
- [37] Z. Wu, S. Pan, G. Long, J. Jiang, and C. Zhang, "Graph WaveNet for deep spatial-temporal graph modeling," in *Proc. 28th Int. Joint Conf. Artif. Intell.*, Aug. 2019, pp. 1907–1913.
- [38] M. Li and Z. Zhu, "Spatial-temporal fusion graph neural networks for traffic flow forecasting," in *Proc. 35th AAAI Conf. Artif. Intell.*, 2021, pp. 4189–4196.
- [39] A. Vaswani et al., "Attention is all you need," in *Proc. 30th Annu. Conf. Neural Inf. Process. Syst.*, 2017, pp. 5998–6008.
- [40] X. Wang et al., "Traffic flow prediction via spatial temporal graph neural network," in *Proc. Web Conf.*, Apr. 2020, pp. 1082–1092.
- [41] C. Zheng, X. Fan, C. Wang, and J. Qi, "GMAN: A graph multi-attention network for traffic prediction," in *Proc. 34th AAAI Conf. Artif. Intell.*, 2020, pp. 1234–1241.
- [42] C. Park et al., "ST-GRAT: A novel spatio-temporal graph attention networks for accurately forecasting dynamically changing road speed," in *Proc. 29th ACM Int. Conf. Inf. Knowl. Manage.*, Oct. 2020, pp. 1215–1224.
- [43] Z. Cui, L. Lin, Z. Pu, and Y. Wang, "Graph Markov network for traffic forecasting with missing data," *Transp. Res. C, Emerg. Technol.*, vol. 117, Aug. 2020, Art. no. 102671.
- [44] K. Guo, Y. Hu, Y. Sun, S. Qian, J. Gao, and B. Yin, "Hierarchical graph convolution network for traffic forecasting," in *Proc. 35th AAAI Conf. Artif. Intell.*, 2021, pp. 151–159.
- [45] X. Zhang et al., "Traffic flow forecasting with spatial-temporal graph diffusion network," in *Proc. 35th AAAI Conf. Artif. Intell.*, 2021, pp. 15008–15015.
- [46] J. Zhu, Q. Wang, C. Tao, H. Deng, L. Zhao, and H. Li, "AST-GCN: Attribute-augmented spatiotemporal graph convolutional network for traffic forecasting," *IEEE Access*, vol. 9, pp. 35973–35983, 2021.
- [47] M. A. Ali, S. Venkatesan, V. Liang, and H. Kruppa, "TEST-GCN: Topologically enhanced spatial-temporal graph convolutional networks for traffic forecasting," in *Proc. IEEE Int. Conf. Data Mining (ICDM)*, Dec. 2021, pp. 982–987.
- [48] L. Han, B. Du, L. Sun, Y. Fu, Y. Lv, and H. Xiong, "Dynamic and multi-faceted spatio-temporal deep learning for traffic speed forecasting," in *Proc. 27th ACM SIGKDD Conf. Knowl. Discovery Data Mining*, Aug. 2021, pp. 547–555.
- [49] R. Bergel-Hayat, M. Debbah, C. Antoniou, and G. Yannis, "Explaining the road accident risk: Weather effects," *Accident Anal. Prevention*, vol. 60, pp. 456–465, Nov. 2013.
- [50] D. Eisenberg, "The mixed effects of precipitation on traffic crashes," *Accident Anal. Prevention*, vol. 36, no. 4, pp. 637–647, Jul. 2004.
- [51] J. D. Tamerius, X. Zhou, R. Mantilla, and T. Greenfield-Huitt, "Precipitation effects on motor vehicle crashes vary by space, time, and environmental conditions," *Weather, Climate, Soc.*, vol. 8, no. 4, pp. 399–407, Oct. 2016.
- [52] C. Caliendo, M. Guida, and A. Parisi, "A crash-prediction model for multilane roads," *Accident Anal. Prevention*, vol. 39, no. 4, pp. 657–670, Jul. 2007.
- [53] L. Lin, Q. Wang, and A. W. Sadek, "A novel variable selection method based on frequent pattern tree for real-time traffic accident risk prediction," *Transp. Res. C, Emerg. Technol.*, vol. 55, pp. 444–459, Jun. 2015.
- [54] L. Chang and W. Chen, "Data mining of tree-based models to analyze freeway accident frequency," *J. Safety Res.*, vol. 36, no. 4, pp. 365–375, 2005.
- [55] S.-H. Park, S.-M. Kim, and Y.-G. Ha, "Highway traffic accident prediction using VDS big data analysis," *J. Supercomput.*, vol. 72, no. 7, pp. 2815–2831, Jul. 2016.
- [56] J. Cho, H. Lee, J. Lee, and D. Kim, "The hazardous expressway sections for drowsy driving using digital tachograph in truck," *J. Korean Soc. Transp.*, vol. 35, no. 2, pp. 160–168, Apr. 2017.
- [57] C. Zhang, X. Yan, L. Ma, and M. An, "Crash prediction and risk evaluation based on traffic analysis zones," *Math. Problems Eng.*, vol. 2014, pp. 1–9, Jan. 2014.
- [58] A. Moradi, H. Soori, A. Kavousi, F. Eshghabadi, E. Jamshidi, and S. Zeini, "Spatial analysis to identify high risk areas for traffic crashes resulting in death of pedestrians in Tehran," *Med. J. Islamic Republic Iran*, vol. 30, p. 450, 2016.
- [59] Q. Chen, X. Song, Z. Fan, T. Xia, H. Yamada, and R. Shibasaki, "A context-aware nonnegative matrix factorization framework for traffic accident risk estimation via heterogeneous data," in *Proc. IEEE Conf. Multimedia Inf. Process. Retr. (MIPR)*, Apr. 2018, pp. 346–351.
- [60] Q. Chen, X. Song, H. Yamada, and R. Shibasaki, "Learning deep representation from big and heterogeneous data for traffic accident inference," in *Proc. 30th AAAI Conf. Artif. Intell.*, 2016, pp. 338–344.
- [61] G. Egilmez and D. McAvoy, "Predicting nationwide road fatalities in the us: A neural network approach," *Int. J. Metaheuristics*, vol. 6, no. 4, pp. 257–278, 2017.
- [62] C. Chen, X. Fan, C. Zheng, L. Xiao, M. Cheng, and C. Wang, "SDCAE: Stack denoising convolutional autoencoder model for accident risk prediction via traffic big data," in *Proc. 6th Int. Conf. Adv. Cloud Big Data (CBD)*, Aug. 2018, pp. 328–333.
- [63] H. Ren, Y. Song, J. Wang, Y. Hu, and J. Lei, "A deep learning approach to the citywide traffic accident risk prediction," in *Proc. 21st Int. Conf. Intell. Transp. Syst. (ITSC)*, Nov. 2018, pp. 3346–3351.
- [64] S. Moosavi, M. H. Samavatian, S. Parthasarathy, R. Teodorescu, and R. Ramnath, "Accident risk prediction based on heterogeneous sparse data: New dataset and insights," in *Proc. 27th ACM SIGSPATIAL Int. Conf. Adv. Geographic Inf. Syst.*, Nov. 2019, pp. 33–42.
- [65] Z. Yuan, X. Zhou, and T. Yang, "Hetero-ConvLSTM: A deep learning approach to traffic accident prediction on heterogeneous spatio-temporal data," in *Proc. 24th ACM SIGKDD Int. Conf. Knowl. Discovery Data Mining*, Jul. 2018, pp. 984–992.
- [66] J. Bao, P. Liu, and S. V. Ukkusuri, "A spatiotemporal deep learning approach for citywide short-term crash risk prediction with multi-source data," *Accident Anal. Prevention*, vol. 122, pp. 239–254, Jan. 2019.
- [67] L. Zhu, T. Li, and S. Du, "TA-STAN: A deep spatial-temporal attention learning framework for regional traffic accident risk prediction," in *Proc. Int. Joint Conf. Neural Netw. (IJCNN)*, Jul. 2019, pp. 1–8.
- [68] C. Huang, C. Zhang, P. Dai, and L. Bo, "Deep dynamic fusion network for traffic accident forecasting," in *Proc. 28th ACM Int. Conf. Inf. Knowl. Manage.*, Nov. 2019, pp. 2673–2681.
- [69] S. He, M. A. Sadeghi, S. Chawla, M. Alizadeh, H. Balakrishnan, and S. Madden, "Inferring high-resolution traffic accident risk maps based on satellite imagery and GPS trajectories," in *Proc. IEEE/CVF Int. Conf. Comput. Vis. (ICCV)*, Oct. 2021, pp. 11977–11985.
- [70] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2016, pp. 770–778.
- [71] N. K. Charandabi, A. Gholami, and A. A. Bina, "Road accident risk prediction using generalized regression neural network optimized with self-organizing map," *Neural Comput. Appl.*, vol. 34, no. 11, pp. 8511–8524, Jun. 2022.
- [72] Korea Transportation Safety Authority. (2015). *Dangerous Driving Behavior Criteria*. [Online]. Available: <https://etas.ts2020.kr/etas/frtl0401/pop/goList.do>
- [73] T. N. Kipf and M. Welling, "Semi-supervised classification with graph convolutional networks," in *Proc. 5th Int. Conf. Learn. Represent.*, 2017.
- [74] L. Chen et al., "SCA-CNN: Spatial and channel-wise attention in convolutional networks for image captioning," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jul. 2017, pp. 5659–5667.
- [75] J. Chung, C. Gulcehre, K. Cho, and Y. Bengio, "Empirical evaluation of gated recurrent neural networks on sequence modeling," 2014, *arXiv:1412.3555*.

- [76] P. Trirat and J.-G. Lee, "DF-TAR: A deep fusion network for citywide traffic accident risk prediction with dangerous driving behavior," in *Proc. Web Conf.*, Apr. 2021, pp. 1146–1156.
- [77] M. Kim et al., "Hi-COVIDNet: Deep learning approach to predict inbound COVID-19 patients and case study in South Korea," in *Proc. 26th ACM SIGKDD Int. Conf. Knowl. Discovery Data Mining*, Aug. 2020, pp. 3466–3473.
- [78] P. J. Huber, "Robust estimation of a location parameter," in *Breakthroughs in Statistics*. Cham, Switzerland: Springer, 1992, pp. 492–518.
- [79] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," in *Proc. 3rd Int. Conf. Learn. Represent.*, 2015.
- [80] G. E. P. Box and D. A. Pierce, "Distribution of residual autocorrelations in autoregressive-integrated moving average time series models," *J. Amer. Statist. Assoc.*, vol. 65, no. 332, pp. 1509–1526, Apr. 1970.
- [81] T. Chen and C. Guestrin, "XGBoost: A scalable tree boosting system," in *Proc. 22nd ACM SIGKDD Int. Conf. Knowl. Discovery Data Mining*, Aug. 2016, pp. 785–794.



**Patara Trirat** received the B.S. degree in computer science from Kasetsart University in 2016 and the M.S. degree in knowledge service engineering from KAIST in 2020, where he is currently pursuing the Ph.D. degree with the School of Computing. His research interests include data mining and analysis, automated machine learning, and applied artificial intelligence.



**Susik Yoon** received the B.S. degree in industrial and systems engineering from KAIST in 2015 and the M.S. and Ph.D. degrees from the Graduate School of Knowledge Service Engineering, KAIST, in 2017 and 2021, respectively. He was a Research Assistant Professor at KAIST. He is a Post-Doctoral Researcher with the Data Mining Laboratory, UIUC. His research interests include data stream mining, anomaly detection, and complex event processing.



**Jae-Gil Lee** (Member, IEEE) received the B.S., M.S., and Ph.D. degrees in computer science from KAIST. He was a Post-Doctoral Researcher at the IBM Almaden Research Center and a Post-Doctoral Research Associate at the University of Illinois Urbana-Champaign. He is a Professor with KAIST and leading the Data Mining Laboratory. His research interests include data-centric AI, deep learning-based big data analysis, mobility and stream data mining, and large-scale distributed deep learning.