DF-TAR: A Deep Fusion Network for Citywide Traffic Accident Risk Prediction with Dangerous Driving Behavior

Patara Trirat
School of Computing, KAIST
Daejeon, South Korea
patara.t@kaist.ac.kr

Jae-Gil Lee*
School of Computing, KAIST
Daejeon, South Korea
jaegil@kaist.ac.kr

ABSTRACT

Because traffic accidents cause huge social and economic losses, it is of prime importance to precisely predict the traffic accident risk for reducing future accidents. In this paper, we propose a **D**eep **F**usion network for citywide **T**raffic **A**ccident **R**isk prediction (*DF-TAR*) with *dangerous driving statistics* that contain the frequencies of various dangerous driving offences in each region. Our unique contribution is to exploit these statistics, obtained by processing the data from *in-vehicle sensors*, for modeling the traffic accident risk. Toward this goal, we first examine the correlation between dangerous driving offences and traffic accidents, and the analysis shows a strong correlation between them in terms of both location and time. Specifically, *quick start* (0.83), *rapid acceleration* (0.76), and *sharp turn* (0.76) are the top three offences that have the highest average correlation scores. We then train the *DF-TAR* model using the dangerous driving statistics as well as external environmental features. By extensive experiments on various frameworks, the *DF-TAR* model is shown to improve the accuracy of the baseline models by up to 54% by virtue of the integration of dangerous driving into the modeling of traffic accident risk.

CCS CONCEPTS

• **Information systems** → *Data mining*; • **Computing methodologies** → **Neural networks**.

KEYWORDS

Traffic accident risk prediction, Deep learning, Dangerous driving behavior, Correlation analysis, Intelligent transportation system

ACM Reference Format:
Patara Trirat and Jae-Gil Lee. 2021. DF-TAR: A Deep Fusion Network for Citywide Traffic Accident Risk Prediction with Dangerous Driving Behavior. In *Proceedings of the Web Conference 2021 (WWW '21), April 19–23, 2021, Ljubljana, Slovenia.* ACM, New York, NY, USA, 11 pages. https://doi.org/10.1145/3442381.3450003

*Jae-Gil Lee is the corresponding author.

WWW '21, April 19–23, 2021, Ljubljana, Slovenia

ACM ISBN 978-1-4503-8312-7/21/04.
https://doi.org/10.1145/3442381.3450003

1 INTRODUCTION

The globally significant increase in traffic accidents caused by fast urbanization has become a crucial socio-economic issue for humanity. The World Health Organization reported that traffic accidents annually killed around 1.24 million people and injured about 50 million people worldwide, most of whom were adolescents [20, 28]. In the year of 2012, *Traffic Safety Facts* reported that traffic accidents in the United States caused by approximately 55,000 buses, of which around 250 buses were associated with casualties; in addition to buses, passenger cars (18,000), small trucks (17,000), and large trucks (3,800) were also associated with fatal cases. In South Korea, car accidents caused about 5,300 deaths—260 cases from buses and 2,300 cases from private cars—in the same year [21].

Owing to the enormous suffering caused by traffic accidents, determining the cause of accidents is vital for designing a safer road environment and establishing an efficient policy. A practical and accurate forecast of traffic accident risk in different city districts over time is another necessity to prevent and mitigate the number of traffic accidents. For example, as noted by Zhou et al. [43], the deployment of the Tennessee accident prediction system reduced the death rate by about 8%. *Traffic accident risk prediction* is, however, still a challenging task because accidents are caused by various factors, including the number of vehicles as well as external environments such as weather, road condition, light ambient, and time of the day [6, 17, 37]. Hence, recent studies [19, 41, 43, 44] have attempted to combine these various factors using complex models to realize precise prediction.

Nevertheless, one crucial factor—*dangerous driving statistics*—has not yet been utilized for traffic accident risk prediction. In fact, dangerous driving is one of the most significant factors that cause traffic accidents [2, 18, 22, 35], along with unsafe road environments and inadequate protective equipment. Besides, diagnosing dangerous driving behavior becomes feasible because driving behavior data can be collected using various *in-vehicle sensors*. For instance, commercial vehicles in several countries [1, 4] are installed with a digital tachograph (DTG) device, and car insurance companies encourage customers to mount such a device for recognizing their driving behavior by offering them a discount on the premium in return [25, 29, 31, 36, 38]. Hence, this driving behavior data is readily available and useful for the prediction task.

In this paper, we explore the idea of using dangerous driving statistics in traffic accident risk prediction. More specifically, the research questions are two-fold, as follows:

- **RQ1.** Is the number of dangerous driving cases for each offence type correlated with the number of traffic accidents?
- **RQ2.** Do the statistics of dangerous driving cases improve the performance of traffic accident risk prediction?

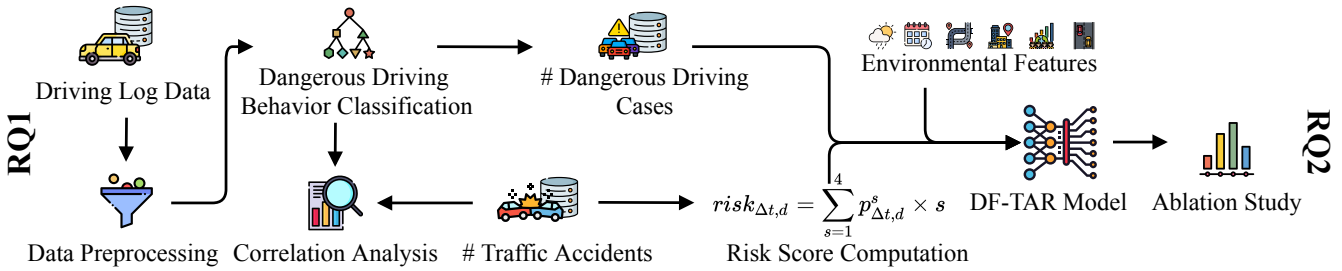


Figure 1: Overview of the research methodology in this paper.

For **RQ1**, we count the dangerous driving cases for each offence type (e.g., overspeeding) and each (sub)district; and then measure the geographical and temporal correlations between the number of dangerous driving cases and the number of actual accident records. For **RQ2**, we propose a model called *DF-TAR* (*Deep Fusion network for Traffic Accident Risk prediction*) and train it using various features including the dangerous driving statistics obtained from the previous stage as well as external environmental features. Since each feature set has different characteristics (e.g., dynamic or static features), *DF-TAR* is based on the fusion of different neural networks in learning the representation from each set to capture the essential pieces of the features and take advantage of the complementarity of each network by fusing them into a unified architecture. Finally, we conduct an ablation study to clarify the significance of dangerous driving statistics. Figure 1 illustrates the overall methodology employed to address the above research questions.

1.1 Key Contributions

Our key contributions are summarized in two perspectives corresponding to the two research questions, respectively.

Usefulness of Dangerous Driving Statistics: We examine the usefulness of dangerous driving statistics through correlation analysis. In our driving log datasets, which will be described in Section 2, *rapid acceleration* (0.88), *quick start* (0.84), and *sharp turn* (0.83) were shown to have a strong correlation to traffic accidents concerning the location of occurrences (in the subdistrict level). Similarly, regarding temporal correlation (in the hour-interval level), *quick start* (0.81), *sharp turn* (0.70), and *sudden u-turn* (0.67) were shown to have a strong correlation to traffic accidents. The above scores are reported with *p-value* < *.01*. Unexpectedly, the *overspeeding* has a very inconsistent correlation to traffic accidents. Therefore, we confirm that some types of dangerous driving are highly correlated with traffic accidents

Improvement of Prediction Performance: To effectively incorporate dangerous driving statistics, we propose a deep learning model for traffic accident risk prediction, *DF-TAR*, with a fusion of different architectures in an end-to-end fashion inspired by Sainath et al. [34]. It improves the prediction performance by exploiting each architecture's advantages in representing different types of data. In particular, injecting dangerous driving statistics into a model improves MAE by up to **32%** and RMSE by up to **5%**. As a result, *DF-TAR* significantly outperforms the baseline models with the improvement of up to **54%** in MAE and **18%** in RMSE. Overall, this result indeed demonstrates the usefulness of considering dangerous driving in traffic accident risk prediction.

Table 1: 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 m/s^2 (-100~100)	2.5
Y acceleration	Y-axis acceleration in 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°~360°)	168

1.2 Paper Structure

Section 2 explains the driving behavior dataset, the dangerous driving criteria, and the problem setting. Section 3 presents the result of correlation analysis to answer RQ1 and the detailed architecture of *DF-TAR*. Section 4 presents the evaluation results to answer RQ2. Then, Section 5 discusses the limitations of this study. Finally, Section 6 reviews related work, and Section 7 concludes this study.

2 PRELIMINARIES

2.1 Driving Behavior (Digital Tachograph) Data

A digital tachograph (DTG) is a record-keeping instrument for driving log data from moving vehicles. The DTG device stores time-stamped driving logs—mostly in intervals of a few seconds—including vehicle type, acceleration, speed, car location, brake signal, azimuth, and more, as shown in Table 1. The installation of this device enhances traffic safety, prevents extreme driving behavior, and promotes fair competition among the transportation companies [4]. The two monthly datasets in September 2016 and September 2018 were 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,

Table 2: Examples of criteria for dangerous driving behavior by KTSA.

Offence Type	Vehicle Type	Definition of Criteria
OS	All	Driving speed exceeds the road speed limit by more than **20km/h**.
RA	Taxi	Driving speed is greater than 6km/h and the acceleration is more than **8km/h** per second.
	Bus	Driving speed is greater than 6km/h and the acceleration is more than **6km/h** per second.
	Truck	Driving speed is greater than 6km/h and the acceleration is more than **5km/h** per second.
ST	Taxi	Driving speed is greater than **30km/h** and the cumulative angle within **3** seconds is between 60° and 120°.
	Bus	Driving speed is greater than **25km/h** and the cumulative angle within **4** seconds is between 60° and 120°.
	Truck	Driving speed is greater than **20km/h** and the cumulative angle within **4** seconds is between 60° and 120°.

Gwangju, and Daejeon) in Korea were selected as these cities have enough population as well as diverse types of roads and vehicles.

2.2 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. In this paper, we used the criteria defined by the KTSA [3], 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 offences, *(long-term) 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 **5km/h** per second for trucks, but **6km/h** per second for buses. Table 2 shows part of the criteria for each vehicle type used in this study.

2.3 Problem Statement

Definition 2.1. (**Driving Log Record**) A *driving log record* is a collection of measurements of the variables in Table 1 within specific time interval Δt (e.g., 8–9 AM). Here, $\mathbf{r}^i_{\Delta t}$ denotes driving log records of vehicle i within time interval Δt. In this study, the time interval is one hour.

Definition 2.2. (**City District**) A *city district* is an uneven local area based on the designated administrative division consisting of n two-dimensional points p of geographic polygons (i.e., the boundary of GPS coordinates). Thus, $\mathbf{d} = \{p_1, p_2, ..., p_n\} \in \mathbb{R}^n$.

Definition 2.3. (**Dangerous Driving Case**) A set of *dangerous driving cases* $\mathbf{c}_{\Delta t,d} = \{c^1_{\Delta t,d}, c^2_{\Delta t,d}, ..., c^b_{\Delta t,d}\}$. $\mathbf{c}^b_{\Delta t,d} = \sum_{i \in V} \mathbf{r}^{i,b}_{\Delta t,d}$, where Δt is the occurred timeframe in district d by vehicle i. $\mathbf{r}^{i,b}_{\Delta t,d} = 1$ if it satisfies the conditions of offense behavior b in Table 2, otherwise 0.

Definition 2.4. (**Environmental Feature**) A set of *environmental features* encompasses several aspects (e.g., weather and time of day) that affect traffic accidents.

The set of environmental features used in this paper is described in Section 4.1. These features are provided (or aggregated) for each district of a city at given period. All feature representations are concatenated to form a fixed-length vector $\mathbf{e}_{\Delta t,d}$ for district d during time interval Δt.

Definition 2.5. (**Traffic Accident Risk**) [12] The *traffic accident risk* indicates the number of injured people in traffic accidents and their severity in a district at a particular time interval (e.g., an hour),

$$risk_{\Delta t,d} = \sum_{s=1}^{4} p^s_{\Delta t,d} \times s, \tag{1}$$

where d is a district in a city (e.g., Gangnam district in Seoul) through time interval Δt, $p^s_{\Delta t,d}$ is the number of injured people of severity level s during time interval Δt in district d, and s indicates the severity level of the accident. Specifically, the severity levels in this study are (1) slight injury with consciousness, (2) small injury without consciousness, (3) serious injury, and (4) death.

Traffic Accident Risk Prediction: Given historical traffic accident risks $risk_{\Delta t,d}$ in Definition 2.5, historical environmental features $\mathbf{e}_{\Delta t,d}$ in Definition 2.4, and dangerous driving cases $\mathbf{c}_{\Delta t,d}$ in Definition 2.3, the objective is to predict the traffic accident risks $risk_{\Delta t',d}$ of each district for the future interval $\Delta t'$, where $\Delta t' > \Delta t$.

Table 3: DTG datasets statistics.

City	Period (Sept.)	Interval	# Districts (Subdistricts)	# Data Entries	# Trips	File Size
Seoul	2016	10 seconds	25	255,692,812	134,201	54GB
	2018	1 second	(467)	283,975,061	28,899	87GB
Busan	2016	10 seconds	16	216,811,769	101,085	46GB
	2018	1 second	(192)	134,652,851	20,955	42GB
Daegu	2016	10 seconds	8	169,865,843	91,803	36GB
	2018	1 second	(204)	76,632,815	12,086	24GB
Gwangju	2016	10 seconds	5	166,246,146	77,963	35GB
	2018	1 second	(202)	124,223,157	19,278	38GB
Daejeon	2016	10 seconds	5	122,343,848	87,933	26GB
	2018	1 second	(177)	117,580,035	18,068	36GB

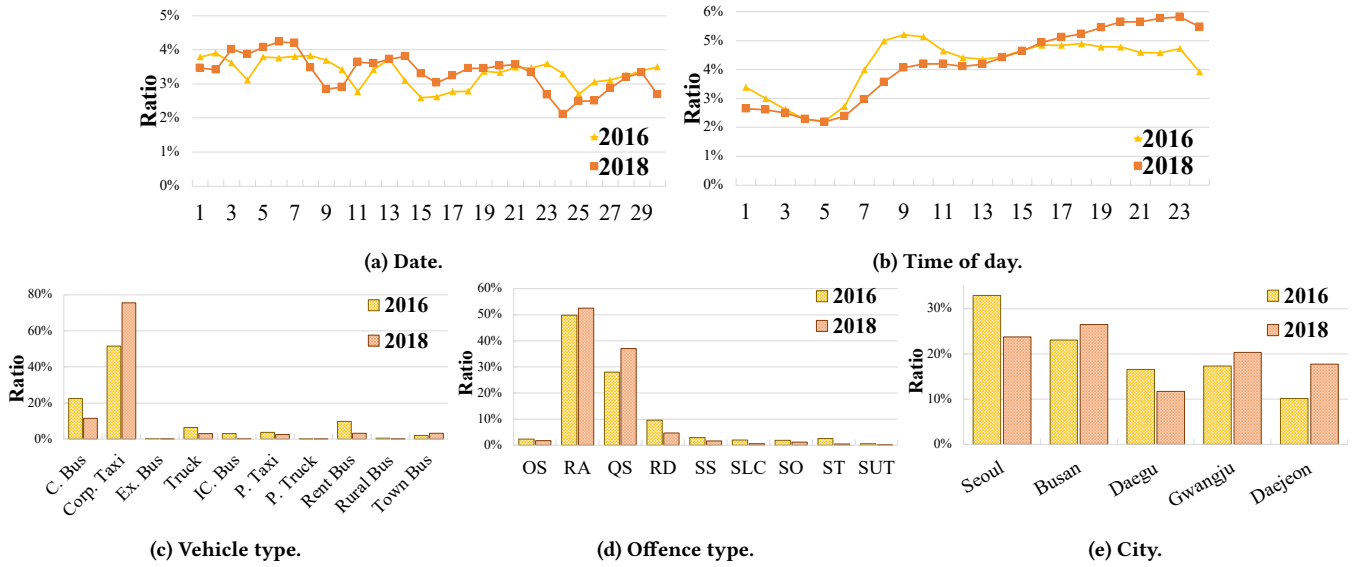
3 METHODOLOGY

3.1 Correlation Analysis

To answer **RQ1**, we measure the correlation between the number of dangerous driving cases and the number of actual accidents. The assumption is that if the dangerous driving behavior has a significantly high correlation with the past accident records, integrating the dangerous driving behavior is reasonable to improve the accuracy of traffic accident risk prediction. The dangerous driving behavior data are grouped according to the geographical and temporal aspects obtained from Definition 2.3—namely, a subdistrict level and an hour interval level. Then, the correlation scores are quantified using the Pearson correlation coefficient.

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

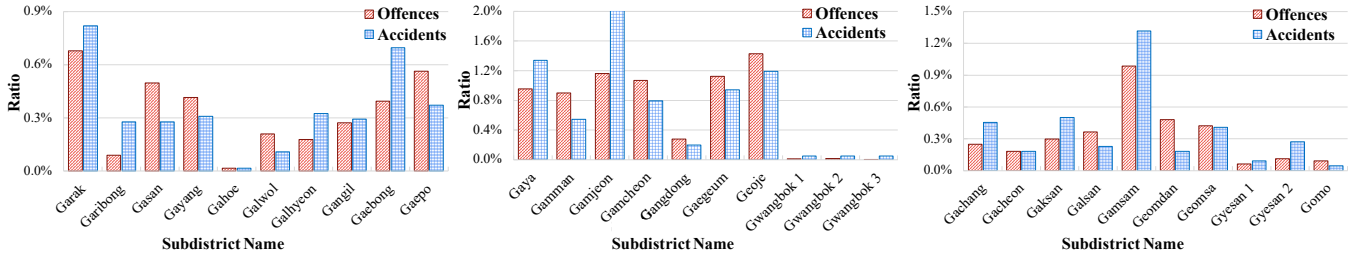
where $Corr(G)$ is the correlation coefficient between the frequency of dangerous driving cases c and the frequency of past accidents a



(a) Date. (b) Time of day.

(c) Vehicle type. (d) Offence type. (e) City.

Figure 2: Ratio of occurrences according to each aspect.



(a) Seoul. (b) Busan. (c) Daegu.

Figure 3: Comparison of ratio of the dangerous driving offences and actual accidents occurred in a subdistrict of each city.

of the given group G (geographical or temporal). That is, the scores are computed using the frequency of dangerous driving cases and past traffic accidents. The analysis results for each city are reported and visualized in the following subsections.

As a standard practice in data analysis, before conducting the above computation, we preprocess the given driving log data by removing outliers (e.g., RPM value is 9999 or the log occurred outside the selected cities' boundary) and imputing missing values. Trips were on the road for less than two minutes are also excluded. Additionally, we define the *idle state* (i.e., parked vehicle) to be deleted with the assumption that those records have a very minimal possibility to induce traffic accidents. A record is the idle state if (1) RPM is less than 1,000, (2) Speed is $0km/h$, (3) X and Y acceleration is between $-1m/s^2$ and $1m/s^2$, and (4) no brake signal. Table 3 shows the total number of records after the preprocessing.

3.2 Key Findings from Correlation Analysis

3.2.1 *Occurrences.* As visualized in Figure 2, we group our findings of the occurrence ratio of the dangerous driving behavior in the two months of different years as follows. **Date.** In general, the dangerous driving cases occurred most towards the end of the week, Friday (16%), Thursday (16%), and Saturday (15%). The lowest ratio of occurrence is during the holiday[1]. **Time of Day.** The period that the dangerous driving behavior occurred most is between 8 PM and 11 PM. **Vehicle Type.** Corporate Taxi (64%), City Bus (17%), and Rent Bus (7%) are the top three vehicle types that have the highest frequency of dangerous driving behavior. **Offence Type.** The top three of risky driving made by drivers are RA (51%), QS (33%), and RD (7%). **City.** Seoul (28%), Busan (25%), and Gwangju (19%) are the top three cities that dangerous driving behavior occurred most.

3.2.2 *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 ($0.3 \leq SD \leq 0.6$), but the computed scores are ***not statistically significant*** ($p\text{-value} > .10$) as well, in many cases regardless of the values. The possible reasons are: (1) the *OS* mostly occurred at very late night that the number of vehicles on the roads is extremely

[1]There is the Chuseok (Korean Thanksgiving) holiday during Sept 14-16, 2016 and Sept 23-25, 2018.

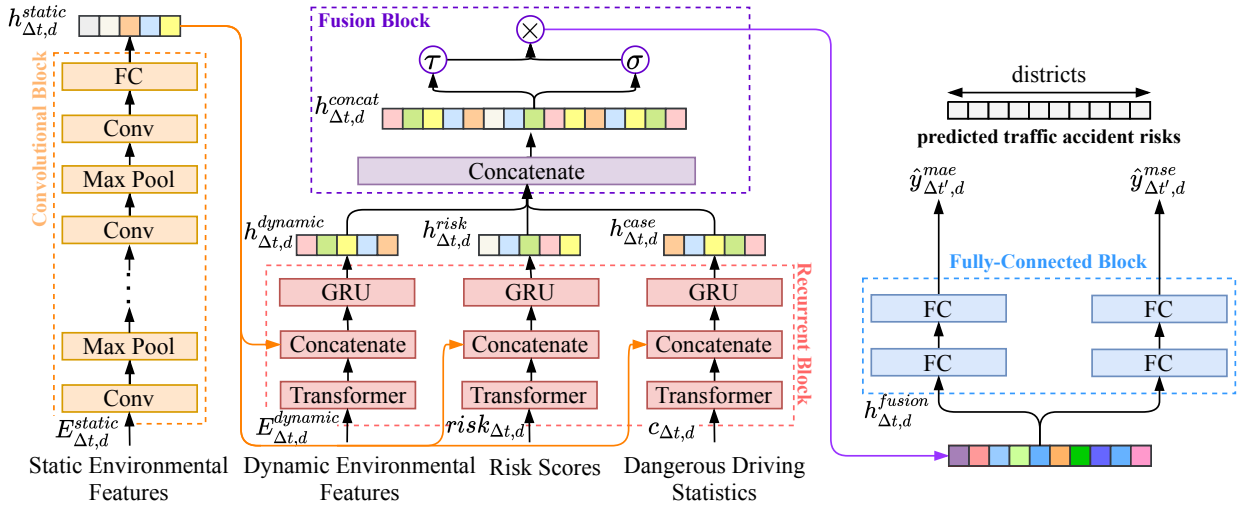


Figure 4: The overall architecture of the *DF-TAR* model.

Table 4: Overall correlation scores of each offence type.

Offence	OS	RA	QS	RD	SS	SLC	SO	ST	SUT	All	w/o OS
Subdistrict	0.35	0.88	0.84	0.71	0.64	0.68	0.79	0.83	0.82	0.88	0.88
Hour	0.01	0.65	0.81	0.64	0.62	0.66	0.40	0.70	0.67	0.73	0.75
Average	0.18	0.76	0.83	0.68	0.63	0.67	0.59	0.76	0.75	0.80	0.81

small, (2) the drivers of those vehicles may be high-skilled drivers that very confident to drive fast. Additionally, as a piece of evidence, the historical record from Korea's traffic accident analysis system (TAAS)[2] also confirms that, among other violations, the traffic accidents caused by *OS* are the smallest portion—about 0.1%.

3.2.3 *Geographic Correlation (Subdistrict Level).* Regarding the location of occurrence, within a subdistrict, the overall number of dangerous driving cases strongly correlates to the traffic accident (0.88). Specifically, *RA* (0.88), *QS* (0.84), and *ST* (0.83) are the top three dangerous driving cases that have a strong correlation to past accidents. Figure 3 presents the occurrence ratio of dangerous driving behavior and actual traffic accidents concerning location. We select the first ten subdistricts that are ordered alphabetically (in Korean) for the visualizations.

3.2.4 *Temporal Correlation (Hour-Interval Level).* In terms of temporal analysis, we compare the frequency of the dangerous driving behavior within an hour interval as it is the smallest unit in the given accident records. The analysis results show that, in general, there is also a strong correlation (0.73) regarding the time of occurrence. Specifically, *QS* (0.81), *ST* (0.70), and *SUT* (0.67) are the top three dangerous driving cases that have a strong correlation to past accidents.

To conclude, we present the overall correlation coefficient scores averaged from the five cities in Table 4. As the answer for RQ1, we ascertain that the dangerous driving offences are strongly correlated with the frequency of traffic accidents.

[2]http://taas.koroad.or.kr/

3.3 *DF-TAR*: A Deep Fusion Network

We propose the *DF-TAR* model that aims to predict the citywide traffic accident risk in each district. The architecture is motivated by work in speech recognition [34] called CLDNN (Convolutional, Long-short term memory, and fully connected Deep Neural Networks). The difference between the *DF-TAR* and the original CLDNN is that we extend the architecture by separating the input into four groups for four feature sets to effectively incorporate with dangerous driving behavior features while still work well with other environmental data. Moreover, we introduce a fusion block to the architecture as a gated activation unit and apply multiple loss functions to enhance the learning process. Thereby, we exploit the advantages of learning representations from different feature sets by the fusion of different neural networks.

As noted by Sainath et al. [34], CNNs, RNNs, and DNNs have specific limitations in their modeling capabilities, combining them in a jointly trained unified framework can improve the performance. The different networks capture information about the input at different scales and complement each other's modeling capabilities. Specifically, we use the convolutional block for learning the hidden representation of each district's spatial features, the recurrent block for the temporal modeling of time-related features, the fusion block for fusing learned representations from the recurrent block while filtering irrelevant ones. Finally, the fully-connected block is used to map aggregated latent representations to a more separable space and make the prediction. Figure 4 depicts the overview architecture of the *DF-TAR* model, which consists of the *convolutional block*, *recurrent block*, *fusion block*, and *fully-connected block*, showing how it fuses the learned representations of each block.

3.3.1 *Convolutional Block.* This block learns a hidden representation of each district's static environmental features and consists of sets of two-dimensional convolutional layers with max-pooling layers. The last layer of this block, the linear layer, is used for dimensionality reduction. The features are said static as they are the same

during all time intervals. The static environmental features are the concatenation of road features $e^{road}_{\Delta t,d}$, point-of-interest information $e^{poi}_{\Delta t,d}$, and demographic data $e^{demographic}_{\Delta t,d}$. Overall, this block is formulated by

$$h^{static}_{\Delta t,d} = FC(MaxPool(Conv(E^{static}_{\Delta t,d}, \Theta^{Conv}), \Theta^{Max}), \Theta^{FC}), \quad (3)$$

where $E^{static}_{\Delta t,d} = Concat([e^{road}_{\Delta t,d}, e^{poi}_{\Delta t,d}, e^{demographic}_{\Delta t,d}])$ is the static environmental feature set of district d occurred during time interval Δt, Θ^{FC},Θ^{Max}, and Θ^{Conv} are the sets of the parameters of the fully-connected layer, max pooling layers, and convolutional layers, respectively.

3.3.2 Recurrent Block. The recurrent block learns the latent representations of the three time-variant feature sets concerning each district. The first feature set is the concatenation of weather information $e^{weather}_{\Delta t,d}$, traffic volume $e^{volume}_{\Delta t,d}$, and calendar data $e^{calendar}_{\Delta t,d}$, hereafter dynamic environmental features. Risk scores $risk_{\Delta t,d}$ and dangerous driving cases $c_{\Delta t,d}$ are the remaining feature sets.

For each sub-block, it consists of the Transformer layer [39], the concatenate layer, and the gated recurrent unit (GRU) layer [14].

Transformer Layer: This layer highlights the significant period in the temporal input features through the attention mechanism. It is helpful because a specific period (e.g., rush hour or holiday) of time-variant information can significantly influence traffic accidents.

Concatenate Layer: It combines the hidden representation of spatial features $h^{static}_{\Delta t,d}$ and the time-variant features $F_{\Delta t,d}$ so that the highlighted significant period can be learned jointly in the next layer.

GRU Layer: This layer intends to capture the temporal patterns for specific geographical properties of aggregated representations. The GRU, an improved version of recurrent neural networks (RNNs), is chosen here because it addressed the vanishing gradient problem [19] and has achieved state-of-the-art performance in sequential modeling [14].

Specifically, each sub-block is formulated as

$$\begin{aligned} TM(F_{\Delta t,d}) &= Transformer(F_{\Delta t,d}, \Theta^{TM}) \\ Rec(F_{\Delta t,d}) &= GRU(Concat([TM(F_{\Delta t,d}), h^{static}_{\Delta t,d}]), \Theta^{GRU}), \end{aligned} \quad (4)$$

where $F_{\Delta t,d}$ represents the above feature set in district d during the time interval Δt. The Θ^{TM} and Θ^{GRU} are the sets of parameters of the Trasformer and GRU, respectively.

From Equation (4), the entire recurrent block is formulated by

$$\begin{aligned} h^{dynamic}_{\Delta t,d} &= Rec(E^{dynamic}_{\Delta t,d}) \\ h^{risk}_{\Delta t,d} &= Rec(risk_{\Delta t,d}) \\ h^{case}_{\Delta t,d} &= Rec(c_{\Delta t,d}), \end{aligned} \quad (5)$$

where $E^{dynamic}_{\Delta t,d} = Concat([e^{weather}_{\Delta t,d}, e^{calendar}_{\Delta t,d}, e^{volume}_{\Delta t,d}])$ is the dynamic environmental feature in district d occurred during time interval Δt. Accordingly, the outputs $h^{dynamic}_{\Delta t,d}$, $h^{risk}_{\Delta t,d}$, and $h^{case}_{\Delta t,d}$ are the latent representations of the dynamic environmental features, risk scores, and dangerous driving cases, respectively.

3.3.3 Fusion Block. The idea behind the fusion block is to fuse all related features while re-scale and selectively preserve the essential learned representations. To this end, the concatenate layer and the gated activation unit [27] are introduced.

Concatenate Layer: This layer merges the hidden representations of all sub-blocks in the recurrent block for the gated activation. Here, the global contexts of all time-variant features are jointly trained.

Gated Activation Unit: It consists of a hyperbolic tangent function and a sigmoid function that aim to re-calibrate the aggregated learned representations. After applying the activation functions, the outputs are multiplied to obtain the essential learned representations. Thus, the final output $h^{fusion}_{\Delta t,d}$ is the fusion block's hidden representation.

Overall, the fusion block is formulated by

$$h^{fusion}_{\Delta t,d} = \tau(h^{concat}_{\Delta t,d}) \odot \sigma(h^{concat}_{\Delta t,d}), \quad (6)$$

where $\odot$ denotes an element-wise multiplication, $\tau(\cdot)$ is a hyperbolic tangent activation function, $\sigma(\cdot)$ is a sigmoid activation function, and $h^{concat}_{\Delta t,d} = Concat([h^{dynamic}_{\Delta t,d}, h^{risk}_{\Delta t,d}, h^{case}_{\Delta t,d}])$ is the concatenation of hidden representations from the recurrent block.

3.3.4 Fully-Connected Block. To map the learned representations to a more separable feature space, we use fully-connected layers for estimating future risk scores of each district. It is added to learn the relationship from the impact of fused hidden representations of district d in the time interval Δt to the future time frame $\Delta t'$ of the same district. Overall, this block is formulated by

$$\begin{aligned} \hat{y}^{mae}_{\Delta t',d} &= FC^{mae}(h^{fusion}_{\Delta t,d}, \Theta^{FC}) \\ \hat{y}^{mse}_{\Delta t',d} &= FC^{mse}(h^{fusion}_{\Delta t,d}, \Theta^{FC}), \end{aligned} \quad (7)$$

where Θ^{FC} is the set of the parameters of the fully-connected layers, $\hat{y}^{mae}_{\Delta t',d}$ and $\hat{y}^{mse}_{\Delta t',d}$ are the predicted risk scores corresponding to mean absolute error (MAE) loss and mean squared error (MSE) loss functions, respectively, for district d during future time interval $\Delta t'$. FC^{mae} is a sub-block that connects to MAE loss, while FC^{mse} is a sub-block for MSE loss. Except the last layer, for each FC layer, the $ReLU(\cdot)$ activation function is used for nonlinear transformation.

3.3.5 Dual Loss Function. After the observation during model training, we find that the predicted results were affected by the loss function due to the scarcity and rarity of traffic accidents. Specifically, training with MAE, the model tried to produce only zeros to minimize the loss values. On the other hand, working with MSE, as it is sensitive to outliers (e.g., big crashes), the output risk scores were unreasonably high. Consequently, we decided to use both loss functions for model optimization to make the model perform consistently in both metrics and get the practical predicted results in general. Formally, the loss function is defined as

$$\mathcal{L} = \frac{1}{k}\sum_{j=1}^{k} |y_j - \hat{y}^{mae}_j| + (y_j - \hat{y}^{mse}_j)^2, \quad (8)$$

where y and $\hat{y}$ is the true and predicted risk scores for each loss function, respectively. j denotes the index of the value and k is the total number of the risk scores of a given district during a specific time interval.

4 EXPERIMENTS

To validate the efficiency of *DF-TAR* and the impact of dangerous driving behavior, we conduct extensive experiments on ten real-world datasets from five cities, each one-month long. The source code is available at https://github.com/kaist-dmlab/DF-TAR.

4.1 Data Description

As stated in Section 2.3, we predict citywide future traffic accident risks using historical accident-related data. Here, as in the recent study [44], the past and future time intervals are 12-hour (formally, we set $\Delta t = \Delta t' = 12$). However, one can train the proposed model with a different choice of Δt. The datasets used in this study are the real-world datasets from five South Korea metropolitan cities collected in September 2016 and September 2018. We categorize the datasets into the following four groups.

4.1.1 Traffic Accident Risk (1 feature). This dataset is the computed traffic accident risk score by Definition 2.5. It represents the numbers of traffic accident risk in a specific city district at a particular time interval. The original historical traffic accident record is collected from the TAAS[3].

4.1.2 Dangerous Driving Behavior (9 features). This dataset represents nine dangerous driving cases in a district at a particular time interval derived by Definition 2.3.

4.1.3 Static Environmental Features (98 features). We group the static environmental features as follows.

Demographic Data: This dataset is collected from the Korean Statistical Information Service (KOSIS)[4]. It consists of 15 features representing the number of population in a district and the distribution of gender, age, and more.

Point-of-Interest (POI) Data: The POI dataset is provided by the Korean Local Information Research and Development Institute (KLID)[5]. It contains the number of commercial buildings (e.g., restaurants, food stores, and entertainment venues), health facilities, and education institutes. We have 41 features that represent the number of each business in a district.

Road Network and Specification: The road data is collected and extracted from the Korea's Transportation Information Center Standard Node Link dataset[6]. This data represents the number of lanes, the road's rank (e.g., city road and expressway), the road's type (e.g., regular road and bridge), the number of connections, and the speed limit. We have 42 features in this dataset.

4.1.4 Dynamic Environmental Features (24 features). The following are the descriptions of dynamic environmental features.

Weather and Air Quality Data: The weather and air quality data is a district-wide real-time observed meteorological data provided by the Korea Meteorological Administration (KMA)[7]. It contains various weather-related information such as temperature, humidity, wind speed and direction, as well as amount of dust (e.g., PM10 and PM2.5). There are 13 features in this dataset.

Traffic Volume: Since the traffic volume data is not publicly available by authorities, the traffic volume here is represented by the number of unique vehicles from the given driving log dataset at a particular time interval in a district.

Calendar Data: The calendar data contains the date and time information. There are 10 features in total. It also represents the day of the week and holiday as a one-hot vector.

4.2 Experimental Settings

To construct training and testing sets, we select 70% of datasets as the training set—from September 1 to September 21—of which 5% are used for validation. The remaining 30% of datasets are the testing set. We predict the next 12-hour citywide traffic accident risk in each city's districts based on the past 12-hour feature sets. We also standardize all input features with min-max normalization.

Moreover, we implement the following baseline models using the Scikit-Learn 0.24 and Tensorflow 2.2 libraries. For classical models, the default hyperparameters are used, while for the deep learning models, we train them with hyperparameters presented in the original papers unless specified. All models are executed on the same platform with a single NVIDIA GeForce RTX 2080 Ti GPU. Concerning the hyperparameters during training, we train *DF-TAR* by minimizing the dual loss function (§3.3.5) using the Adam optimizer with an initial learning rate of 0.001. We also employ learning rate decay and early-stopping techniques. In our experiment, the batch size is set to 32.

Specifically, the hyperparameters of the model are as follows. In the *convolutional block*, there are four convolutional layers, each has 256 feature maps with a max-pooling layer whose filter size is (3,3) except for the last one. The filter sizes of the four convolutional layers are (9,9), (4,4), (3,3), and (3,3), respectively. For each sub-block of the *recurrent block*, the Transformer's embedding dimension is set to be the number of districts times the dimension of input features, and the number of attention heads is set to be the dimension of input features. The number of the hidden units of the GRU layers is set as 256. Finally, for each sub-block of the *fully-connected block*, there are three dense layers, each has 1024 hidden units, followed by the output layer with the number of neurons equals to the time step length times the number of districts.

4.2.1 Baseline Models. For comparison studies, three traditional and four state-of-the-art models are selected as baselines.

(1) Historical Average (HA). The average accident risk scores over the past 12 hours.

(2) Linear Regression (LR). A linear regressor that minimizes the residual sum of squares between the past 12-hour traffic accident risk scores and the next 12-hour predicted scores.

(3) Extreme Gradient Boosting (XGB). A distributed version of the tree-based gradient boosting technique trained with the past 12-hour traffic accident risk scores.

(4) SDAE [12]. A stacked denoising autoencoder model for real-time risk prediction with human mobility data. We use the 12-hour traffic volume as the human mobility data in this paper.

(5) TARPML [32]. A stacked long short-term memory (LSTM)-based approach for citywide traffic accident risk prediction. It consists of four LSTMs and three dense layers with the past 12-hour traffic accident risk scores as input.

[3] http://taas.koroad.or.kr/gis/mcm/mcl/initMap.do?menuId=GIS_GMP_STS_RSN
[4] https://kosis.kr/index/index.do
[5] http://localdata.kr/
[6] https://nodelink.its.go.kr/nodelink/intro
[7] https://data.kma.go.kr/cmmn/main.do

Table 5: Model performance scores for each city in the 2016-09 (September 2016) dataset with averaged scores.

Models	Seoul		Busan		Daegu		Daejeon		Gwangju		Average			
	MAE	RMSE	MAE	RMSE	MAE	RMSE	MAE	RMSE	MAE	RMSE	MAE	std.	RMSE	std.
HA	0.8820	1.6642	0.4373	1.1739	0.9064	1.6326	0.9343	1.6914	1.0025	1.9993	0.8325	±0.23	1.6323	±0.30
LR	1.9095	2.6014	0.8521	1.4728	1.0762	1.7200	1.0007	1.7071	1.1191	2.0907	1.1915	±0.41	1.9184	±0.44
XGB	0.9599	1.6302	0.5144	1.3830	0.9962	1.8437	1.0344	1.9227	1.2208	2.3850	0.9451	±0.26	1.8329	±0.37
SDAE	0.9033	1.6388	0.3932	1.1638	0.9110	1.5890	0.8929	1.6281	0.9862	1.9310	0.8173	±0.24	1.5901	±0.27
TARPML	0.8016	1.6630	0.3468	1.1501	0.8899	1.6378	0.7418	1.6728	1.0769	1.9938	0.7714	±0.27	1.6235	±0.30
Hetero-ConvLSTM	0.6514	1.6084	0.2958	1.2509	0.7354	1.6314	0.7261	1.6291	1.0883	1.9373	0.6994	±0.28	1.6114	±0.24
TA-STAN	0.7687	1.6582	0.3238	1.1462	0.7417	1.6171	0.8056	1.6548	0.9687	1.9290	0.7217	±0.24	1.6010	±0.28
DF-TAR	**0.5983**	**1.6050**	**0.2906**	**1.1090**	**0.5925**	**1.5992**	**0.5859**	**1.6279**	**0.6288**	**1.9252**	**0.5392**	±0.14	**1.5732**	±0.29

Table 6: Model performance scores for each city in the 2018-09 (September 2018) dataset with averaged scores.

Models	Seoul		Busan		Daegu		Daejeon		Gwangju		Average			
	MAE	RMSE	MAE	RMSE	MAE	RMSE	MAE	RMSE	MAE	RMSE	MAE	std.	RMSE	std.
HA	0.7220	1.5531	0.4002	1.1519	1.0172	1.8983	0.9522	1.8456	0.8170	1.6093	0.7817	±0.24	1.6116	±0.30
LR	1.7191	2.4215	0.8621	1.4713	1.2017	2.0179	1.0474	1.8918	0.9599	1.6302	1.1581	±0.34	1.8865	±0.37
XGB	0.8429	1.8028	0.4477	1.3174	1.1542	2.1564	1.1306	2.3309	0.9572	1.8007	0.9065	±0.29	1.8816	±0.39
SDAE	0.8097	1.5585	0.3746	1.1319	0.9744	1.8449	0.9389	1.8000	0.8766	1.5809	0.7948	±0.24	1.5833	±0.28
TARPML	0.7012	1.5504	0.3130	1.1217	0.8444	1.8954	0.9269	1.8210	0.8363	1.5950	0.7244	±0.24	1.5967	±0.30
Hetero-ConvLSTM	0.5448	1.5425	0.3010	1.2076	0.8698	1.8720	0.7236	1.8210	0.7542	1.5910	0.6387	±0.22	1.6068	±0.26
TA-STAN	0.7484	1.5473	0.3147	1.1962	0.8617	1.8809	0.8731	1.7946	0.8121	1.6017	0.7220	±0.23	1.6041	±0.27
DF-TAR	**0.4985**	**1.5406**	**0.2553**	**1.1200**	**0.7593**	**1.9157**	**0.5775**	**1.7432**	**0.6364**	**1.5344**	**0.5454**	±0.19	**1.5708**	±0.30

(6) Hetero-ConvLSTM [41]. A state-of-the-art deep learning framework for traffic accident prediction with heterogeneous spatial-temporal data implemented with ConvLSTM networks. In this paper, the SpatialGraph features are constructed with the above road network data. Except for the demographic data and dangerous driving behavior, we train the model with all features.

(7) TA-STAN [44]. An encoder-decoder framework with spatial and temporal attention modules. Also, we train this model with all features except for the demographic and dangerous driving data.

4.2.2 Evaluation Tasks. Through the experiments, we expect to answer the following three questions: (1) Is the result of *DF-TAR* better than those of the baselines and previous studies? (2) How each feature group impacts on the prediction performance: is the dangerous driving behavior most significant? (3) Is the predicted result of *DF-TAR* consistent with the ground truth?

4.2.3 Evaluation Metrics. The mean absolute error (MAE) and root mean squared error (RMSE) are adopted as the metrics to evaluate the prediction error. They are defined as

$$MAE = \frac{1}{k}\sum_{j=1}^{k}|y_j - \hat{y}_j| \text{ and } RMSE = \sqrt{\frac{1}{k}\sum_{j=1}^{k}(y_j - \hat{y}_j)^2}, \quad (9)$$

where y and $\hat{y}$ are the true and predicted risk scores, respectively. $\hat{y}$ is the average of $\hat{y}^{mae}$ and $\hat{y}^{mse}$.

4.3 Experimental Results

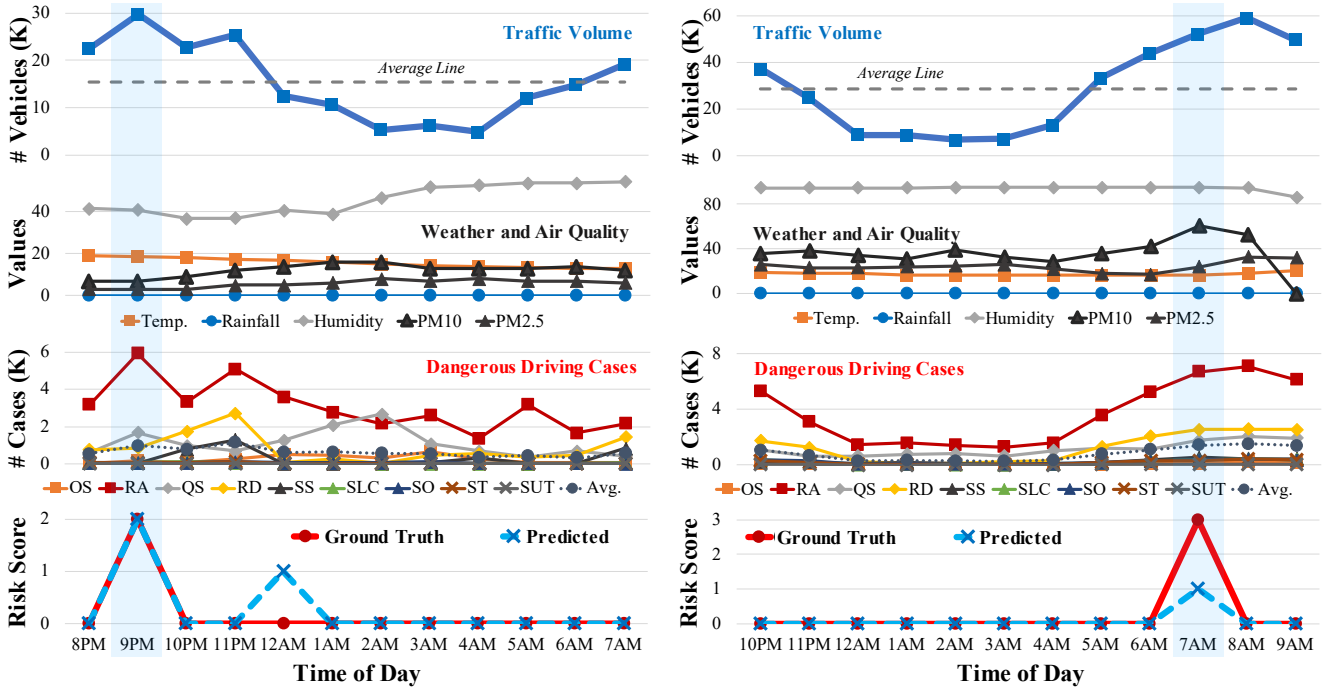
This section compares our model to the baseline models. In general, on average, we observe that HA is better than simple linear regression (LR) and extreme gradient boosting models, indicating that traffic accidents in each district have a certain periodicity and seasonality (when using only historical accident data) that can be a good estimator for general long-term prediction. SDAE, TARPML, and TA-STAN models are better than other machine learning models since the deep learning approaches can generalize the complex and heterogeneous input data better than the traditional approaches. Hetero-ConvLSTM, the second-best model, performs better than other deep learning models thanks to its combination of the Spatial-Graph feature and various environmental features. Similarly, our proposed *DF-TAR* model performs the best with the improvement of up to **54%** in MAE and **18%** in RMSE from the baseline when training with all features, including the dangerous driving statistics because it represents in which district the traffic accidents may occur. Table 5 and Table 6 present the MAE and RMSE scores for 2016-09 and 2018-09 datasets, respectively.

Table 7: Model performance with the different feature set.

Data	MAE	std.	RMSE	std.
Historical Traffic Accident Risks (HA)	0.8023	±0.26	1.6650	±0.28
HA + Static Features (SF)	0.7869	±0.29	1.6521	±0.26
HA + SF + Dynamic Features (DF)	0.7468	±0.26	1.6433	±0.26
HA + Dangerous Driving Cases (DC)	0.7304	±0.23	1.6428	±0.27
HA + SF + DF + DC	**0.5429**	±0.16	**1.5720**	±0.28

4.4 Ablation Study

To quantify the impact of each feature group, we perform a feature ablation study. In this paper, the feature ablation study is the same as the component ablation study since each feature group has a corresponding network to learn the representations. The results, presented in Table 7, indicate that even only with the dangerous driving behavior data can improve the performance by **9%** in MAE



(a) Seoul, Gangnam district. **(b) Daejeon, Daedeok district.**

Figure 5: Comparison between the predicted risk scores and the ground truth scores with time-variant features.

and **1%** in RMSE—comparable to using all environmental features. Also, as one might expect, combining more data sources—either external environmental features or the dangerous driving behavior—reduces the prediction errors by **32%** in MAE and **5%** in RMSE. Thereby, it means that traffic accident is actually related to other factors such as weather, road conditions, and the dangerous driving behavior of drivers. Given these results, as the answer to **RQ2**, we are confident that applying dangerous driving behavior can improve traffic accident risk prediction performance.

4.5 Case Study

To evaluate the proposed model's effectiveness, we select the various time intervals from the two cities for comparing the predicted risk scores and the actual traffic accident risks in the testing sets. Figure 5 gives the visualizations of the first district—in alphabetical order—of Seoul (capital city) and Daejeon. The visualizations reveal that our proposed model can predict the traffic accident risk close to the real risk values, even though the event is sporadic.

Finally, we investigate the effects of the time-variant features to determine whether the predicted values result from the model's learning capability. Taking Figure 5a as an example, it illustrates the results in the Gangnam district from 8 PM on September 24, 2018 to 7 AM on September 25, 2018 (12-hour interval). As there is no significant change (no rainfall with good air quality) in the weather and air quality data, the weather may not be the key factor for the accidents. Then, about the traffic volume and dangerous driving cases, they show high traffic volume (91% above average) and dangerous driving cases (56% above average) in that time interval. Specifically, the *RA* offense—one of the highest correlation with traffic accidents according to the findings in Section 3—occurred most during that time. The results from Daejeon also exhibit similar patterns, as shown in Figure 5b. Hence, this observation proves that the *DF-TAR* can learn to adjust inferences accordingly by capturing dynamic patterns of dangerous driving statistics, historical accident distributions, and environmental features.

5 LIMITATION

Due to the regulations in acquiring digital tachograph data, we can use only one-month data sets of five cities collected in 2016 and 2018 (i.e., ten months in total) to train and test 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 data sets to multiple months or years. Additionally, with more extensive data, especially for deep learning-based approaches, the model would become more generalizable.

6 RELATED WORK

6.1 Classic Learning-based Studies

Eisenberg [17] investigated the association between rainfall and fatalities with a negative binomial regression model. Chang and Chen

[9] employed a tree-based model to predict accidents on specific highway road segments. Caliendo et al. [8] made several regression models to forecast the traffic accidents on given roads with a set of road-related attributes; for example, the annual average daily traffic (AADT), length, curve, sign distance, and presence of junction. Lv et al. [24] applied the k-nearest neighbor method to predict car accidents with Euclidean metric-based features selection. Xie and Yan [40] and Bíl et al. [7] studied traffic accidents by mainly focusing on hot spot identification. Bergel-Hayat et al. [6] examined the correlations between weather and accident damages by using an auto-regressive regression model. Sadeghi et al. [33] applied data envelopment analysis to suggest a novel approach in identifying road segments that are likely to have traffic accidents. Zhang et al. [42] studied on zonal safety evaluation and introduced a negative binomial regression model to determine critical factors for the hazardous zones. Lin et al. [23] created a decision tree model to predict traffic accidents with heterogeneous data (e.g., traffic volume, speed, weather, visibility, and occupancy information) from a section of the Virginia Interstate-64 in the United States.

To identify high-risk zones, Moradi et al. [26] conducted a spatial analysis of traffic accidents on casualties among pedestrians. Park et al. [30] designed a prediction workflow based on the k-means clustering technique and logistic regression with extensive traffic accident data of a city highway. More recently, Dunlop et al. [15] proposed a system that works with smartphones in cities to discover the unsafe segments of a road through a crowdsourcing approach for warning drivers with an alert. Tamerius et al. [37] analyzed the correlation between precipitation and car accidents over space and time by examining the relative accident rate. To classify the risky segments on an expressway, Cho et al. [13] employed negative binomial regression models that could establish the correlation between accidents and driving behavior by analyzing the driving log data, the accidents caused by drowsiness and road conditions. Chen et al. [11] applied the nonnegative matrix factorization method to predict the frequency of traffic accidents represented as a matrix.

Most previous studies adopted classical data mining techniques on small-scale traffic accident data with limited features. Additionally, they did not address unique data characteristics such as time periodicity, spatial autocorrelation, and heterogeneity; thus, they produced low accuracy. Finally, they could not meet the requirements of predicting real-time traffic accident risk in the nearby roads to suggest safer routes successfully. However, deep learning-based approaches can resolve these problems, as documented below.

6.2 Deep Learning-based Studies

Chen et al. [12] proposed a stack denoising autoencoder and a logistic regression model trained with a large amount of heterogeneous data on accidents and human mobility data, for traffic accident risk inference in Tokyo. Egilmez and McAvoy [16] obtained seven factors that impact road accidents from an optimized radial basis function neural network. Chen et al. [10] used a stacked denoising convolutional autoencoder model to predict the frequency of accidents in grid cells using traffic flow, past traffic accidents, and time data. Ren et al. [32] analyzed the spatial and temporal characteristics of traffic accident frequency and built an LSTM-based deep learning model trained with big traffic accident data in Beijing to predict the traffic accident risk.

More sophisticatedly, Yuan et al. [41] introduced a spatial ensemble model called Hetero-ConvLSTM, for accident frequency prediction within spatial grid cells of Iowa, using various environmental data such as satellite images, road conditions, traffic flow, weather, and precipitation. They evaluated their model with large-scale datasets of traffic accidents and explained the significance of addressing spatial heterogeneity and temporal trends to increase the prediction performance. Bao et al. [5] predicted the citywide short-term accident risk in different sizes of grid cells, with human mobility data inferred from taxis, using the spatiotemporal convolutional LSTM network (STCL-Net). Zhu et al. [44] introduced a deep neural network model that utilized spatial-temporal attention networks and various external datasets to forecast traffic accident risk in the traffic administrative divisions of New York City. Huang et al. [19] enhanced the ability of deep neural networks to model various external factors in a fully dynamic manner, focusing on abnormal events data. Recently, Zhou et al. [43] proposed a framework for minute-level citywide traffic accident risk prediction within grid maps, using a multi-task differential time-varying graph convolution network, called RiskOracle.

Although these studies addressed the spatial-temporal patterns of traffic accident data and achieved better prediction results than the traditional approaches, as mentioned by Zhu et al. [44], the artificial separation of the administrative district (i.e., grid maps) breaks the whole pattern of the regions and causes deviation from the prediction. Regarding the training datasets, most of the prior studies used only the environmental features; thus, they mainly relied on the external factors. However, as a rare event, a traffic accident may intrinsically be caused by driving behavior, which can be analyzed from the driving log data. Therefore, in this study, in addition to the environmental data, we harness the driving log data to extract the dangerous driving behavior and quantify its correlations with past road accidents. Besides, to lessen the deviation of predicted results, we train the proposed predictive model based on the actual administrative districts instead of the grid-based division.

7 CONCLUSION AND FUTURE WORK

This paper introduces a deep fusion network for citywide traffic accident risk prediction, which we call *DF-TAR*. Before training the model, we quantify the correlation scores between dangerous driving behavior and past accident records to substantiate their relationship. Our findings reveal that dangerous driving behavior strongly correlates to traffic accidents in both geographical and temporal aspects. Therefore, to forecast each district's future traffic accident risk in a city at a specific time interval, we train the proposed *DF-TAR* model with extracted dangerous driving behavior and various environmental data. The evaluation results show that the proposed model outperforms the baselines, with the improvement of up to 54% in MAE and 18% in RMSE.

For future work, with an appropriate data acquisition mechanism, dangerous driving behavior can be utilized with real-time traffic accident prediction systems to help individuals avoid potential risks associated with each vehicle on the road. Moreover, this approach may also be adopted with autonomous vehicles to alert each other on the roads collectively. Ultimately, we expect that our findings and the proposed model will prove beneficial in promoting safe driving and preventing future traffic accidents.

ACKNOWLEDGMENTS

This work was partly supported by Institute of Information & Communications Technology Planning & 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 the National Research Foundation of Korea (NRF) grant funded by the Korea government (Ministry of Science and ICT) (No. 2020R1A2B5B03095947).

REFERENCES

[1] Korean National Police Agency. 2014. *Traffic Safety Act of the Republic of Korea.* Chapter Article 55 (Equipment with An Operation Recording Device and Utilization of the Operation Recording, etc.).

[2] IM Almeida, Regina Célia P Leal-Toledo, Elson Magalhães Toledo, Diego Carrico Cacau, and GVP Magalhães. 2018. Drivers' Behavior Effects in the Occurrence of Dangerous Situations Which May Lead to Accidents. In *Proceedings of the 13th International Conference on Cellular Automata.* 441–450.

[3] Korea Transportation Safety Authority. 2015. Dangerous Driving Behavior Criteria. https://etas.ts2020.kr/etas/frtl0401/pop/goList.do

[4] Gianmarco Baldini, Luigi Sportiello, Michel Chiaramello, and Vincent Mahieu. 2018. Regulated Applications for the Road Transportation Infrastructure: The Case Study of the Smart Tachograph in the European Union. *International Journal of Critical Infrastructure Protection* 21 (2018), 3–21.

[5] Jie Bao, Pan Liu, and Satish V Ukkusuri. 2019. A Spatiotemporal Deep Learning Approach for Citywide Short-term Crash Risk Prediction with Multi-source Data. *Accident Analysis & Prevention* 122 (2019), 239–254.

[6] Ruth Bergel-Hayat, Mohammed Debbarh, Constantinos Antoniou, and George Yannis. 2013. Explaining the Road Accident Risk: Weather Effects. *Accident Analysis & Prevention* 60 (2013), 456–465.

[7] Michal Bíl, Richard Andrášik, and Zbyněk Janoška. 2013. Identification of Hazardous Road Locations of Traffic Accidents by Means of Kernel Density Estimation and Cluster Significance Evaluation. *Accident Analysis & Prevention* 55 (2013), 265–273.

[8] Ciro Caliendo, Maurizio Guida, and Alessandra Parisi. 2007. A Crash-Prediction Model for Multilane Roads. *Accident Analysis & Prevention* 39, 4 (2007), 657–670.

[9] Li-Yen Chang and Wen-Chieh Chen. 2005. Data Mining of Tree-based Models to Analyze Freeway Accident Frequency. *Journal of Safety Research* 36, 4 (2005), 365–375.

[10] Chao Chen, Xiaoliang Fan, Chuanpan Zheng, Lujing Xiao, Ming Cheng, and Cheng Wang. 2018. SDCAE: Stack Denoising Convolutional Autoencoder Model for Accident Risk Prediction via Traffic Big Data. In *Proceedings of the 6th International Conference on Advanced Cloud and Big Data.* 328–333.

[11] Quanjun Chen, Xuan Song, Zipei Fan, Tianqi Xia, Harutoshi Yamada, and Ryosuke Shibasaki. 2018. A Context-Aware Nonnegative Matrix Factorization Framework for Traffic Accident Risk Estimation via Heterogeneous Data. In *Proceedings of the 2018 IEEE Conference on Multimedia Information Processing and Retrieval.* 346–351.

[12] Quanjun Chen, Xuan Song, Harutoshi Yamada, and Ryosuke Shibasaki. 2016. Learning Deep Representation from Big and Heterogeneous Data for Traffic Accident Inference. In *Proceedings of the 30th AAAI Conference on Artificial Intelligence.* 338–344.

[13] Jongseok Cho, Hyunsuk Lee, Jaeyoung Lee, and Ducknyung Kim. 2017. The Hazardous Expressway Sections for Drowsy Driving Using Digital Tachograph in Truck. *Journal of Korean Society of Transportation* 35, 2 (2017), 160–168.

[14] Junyoung Chung, Caglar Gulcehre, KyungHyun Cho, and Yoshua Bengio. 2014. Empirical Evaluation of Gated Recurrent Neural Networks on Sequence Modeling. *arXiv preprint arXiv:1412.3555* (2014).

[15] Mark D Dunlop, Marc Roper, Mark Elliot, Rebecca McCartan, and Bruce McGregor. 2016. Using Smartphones in Cities to Crowdsource Dangerous Road Sections and Give Effective In-car Warnings. In *Proceedings of the 2016 on Smart Cities for Better Living with HCI and UX.* 14–18.

[16] Gokhan Egilmez and Deborah McAvoy. 2017. Predicting Nationwide Road Fatalities in the US: A Neural Network Approach. *International Journal of Metaheuristics* 6, 4 (2017), 257–278.

[17] Daniel Eisenberg. 2004. TheThe Mixed Effects of Precipitation on Traffic Crashes. *Accident Analysis & Prevention* 36, 4 (2004), 637–647.

[18] Veronika Harantová, Simona Kubíková, and Luboš Rumanovskỳ. 2019. Traffic Accident Occurrence, Its Prediction and Causes. In *Proceedings of the 19th International Conference on Transport Systems Telematics.* 123–136.

[19] Chao Huang, Chuxu Zhang, Peng Dai, and Liefeng Bo. 2019. Deep Dynamic Fusion Network for Traffic Accident Forecasting. In *Proceedings of the 28th ACM International Conference on Information and Knowledge Management.* 2673–2681.

[20] Joon-Gyu Kang, Yoo-Won Kim, and Moon-Seog Jun. 2015. Real-time Dangerous Driving Behavior Analysis Utilizing the Digital Tachograph and Smartphone. *Journal of the Korea Society of Computer and Information* 20, 12 (2015), 37–44.

[21] Do-Gyeong Kim, Chungwon Lee, and Byung-Jung Park. 2016. Use of Digital Tachograph Data to Provide Traffic Safety Education and Evaluate Effects on Bus Driver Behavior. *Transportation Research Record* 2585, 1 (2016), 77–84.

[22] Korea Road Traffic Authority. 2020. What Causes Traffic Accidents? https://www.koroad.or.kr/en_web/view/trfEnv4.do

[23] Lei Lin, Qian Wang, and Adel W Sadek. 2015. A Novel Variable Selection Method based on Frequent Pattern Tree for Real-time Traffic Accident Risk Prediction. *Transportation Research Part C: Emerging Technologies* 55 (2015), 444–459.

[24] Yisheng Lv, Shuming Tang, and Hongxia Zhao. 2009. Real-Time Highway Traffic Accident Prediction Based on the k-Nearest Neighbor Method. In *Proceedings of the 2009 International Conference on Measuring Technology and Mechatronics Automation*, Vol. 3. 547–550.

[25] Jason Metz. 2018. *Your Car's OBD-II: What Is It?* EverQuote, Inc. https://www.everquote.com/blog/car-insurance/your-car-obd/

[26] Ali Moradi, Hamid Soori, Amir Kavousi, Farshid Eshghabadi, Ensiyeh Jamshidi, and Salahdien Zeini. 2016. Spatial Analysis to Identify High Risk Areas for Traffic Crashes Resulting in Death of Pedestrians in Tehran. *Medical Journal of the Islamic Republic of Iran* 30 (2016), 450.

[27] Aaron van den Oord, Sander Dieleman, Heiga Zen, Karen Simonyan, Oriol Vinyals, Alex Graves, Nal Kalchbrenner, Andrew Senior, and Koray Kavukcuoglu. 2016. WaveNet: A Generative Model for Raw Audio. *arXiv preprint arXiv:1609.03499* (2016).

[28] World Health Organization. 2015. *Global Status Report on Road Safety 2015.*

[29] Geun Young Park. 2014. *What If You Could Give a Driver with a Good Driving Habit a Discount on His Car Insurance?* Korea Road Traffic Authority. https://www.koroad.or.kr/kp_web/krPrView.do?board_code=GABBS_050&board_num=126843

[30] Seong-hun Park, Sung-min Kim, and Young-guk Ha. 2016. Highway Traffic Accident Prediction Using VDS Big Data Analysis. *The Journal of Supercomputing* 72, 7 (2016), 2815–2831.

[31] Sang Young Park. 2020. *Automatically Record Business Vehicles of Speeding, Sudden Stop, and Reckless Driving through Smartphones.* The Kyunghyang Shinmun. http://news.khan.co.kr/kh_news/khan_art_view.html?art_id=202003302130005

[32] Honglei Ren, You Song, Jingwen Wang, Yucheng Hu, and Jinzhi Lei. 2018. A Deep Learning Approach to the Citywide Traffic Accident Risk Prediction. In *Proceedings of the 21st International Conference on Intelligent Transportation Systems.* 3346–3351.

[33] Aliasghar Sadeghi, Esmaeel Ayati, and Mohammadali Pirayesh Neghab. 2013. Identification and Prioritization of Hazardous Road Locations by Segmentation and Data Envelopment Analysis Approach. *PROMET-Traffic&Transportation* 25, 2 (2013), 127–136.

[34] Tara N Sainath, Oriol Vinyals, Andrew Senior, and Haşim Sak. 2015. Convolutional, Long Short-Term Memory, Fully Connected Deep Neural Networks. In *Proceedings of the 2015 IEEE International Conference on Acoustics, Speech and Signal Processing.* 4580–4584.

[35] John M Scanlon, Rini Sherony, and Hampton C Gabler. 2017. Models of Driver Acceleration Behavior Prior to Real-World Intersection Crashes. *IEEE Transactions on Intelligent Transportation Systems* 19, 3 (2017), 774–786.

[36] Paul Stenquist. 2020. *Letting Your Insurer Ride Shotgun, for a Discounted Rate.* The New York Times Company. https://www.nytimes.com/2020/07/16/business/car-insurance-app-discounts.html

[37] JD Tamerius, X Zhou, R Mantilla, and T Greenfield-Huitt. 2016. Precipitation Effects on Motor Vehicle Crashes Vary by Space, Time, and Environmental Conditions. *Weather, Climate, and Society* 8, 4 (2016), 399–407.

[38] Cherise Threewitt and John M Vincent. 2018. *How Do Those Car Insurance Tracking Devices Work?* US News & World Report. https://cars.usnews.com/cars-trucks/car-insurance/how-do-those-car-insurance-tracking-devices-work

[39] Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N Gomez, Lukasz Kaiser, and Illia Polosukhin. 2017. Attention is All you Need. *arXiv preprint arXiv:1706.03762* (2017).

[40] Zhixiao Xie and Jun Yan. 2013. Detecting Traffic Accident Clusters with Network Kernel Density Estimation and Local Spatial Statistics: An Integrated Approach. *Journal of Transport Geography* 31 (2013), 64–71.

[41] Zhuoning Yuan, Xun Zhou, and Tianbao Yang. 2018. Hetero-ConvLSTM: A Deep Learning Approach to Traffic Accident Prediction on Heterogeneous Spatio-Temporal Data. In *Proceedings of the 24th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining.* 984–992.

[42] Cuiping Zhang, Xuedong Yan, Lu Ma, and Meiwu An. 2014. Crash Prediction and Risk Evaluation Based on Traffic Analysis Zones. *Mathematical Problems in Engineering* (2014).

[43] Zhengyang Zhou, Yang Wang, Xike Xie, Lianliang Chen, and Hengchang Liu. 2020. RiskOracle: A Minute-Level Citywide Traffic Accident Forecasting Framework. In *Proceedings of the 34th AAAI Conference on Artificial Intelligence.* 1258–1265.

[44] Lei Zhu, Tianrui Li, and Shengdong Du. 2019. TA-STAN: A Deep Spatial-Temporal Attention Learning Framework for Regional Traffic Accident Risk Prediction. In *Proceedings of the 2019 International Joint Conference on Neural Networks.* 1–8.