

ARTICLE

The Effectiveness of Socio-Economic Context in Short-Term Highway Traffic Forecasting

Ohsung Kwon¹ and Kyoung-Soub Lee^{2,*}

¹Department of Business Intelligence, Ajou University, Suwon, Republic of Korea

²School of Mobility Engineering, University of Ulsan, Ulsan, Republic of Korea

*Corresponding Author: Kyoung-Soub Lee. Email: lks0872@ulsan.ac.kr

Received: 12 April 2026; Accepted: 18 August 2026; Published: 22 October 2026

ABSTRACT: Traffic congestion incurs massive social and economic costs. Accordingly, the importance of transportation management and planning is gradually expanding. In response, the establishment of intelligent transportation systems utilizing advanced technology is increasing, as are global attempts to predict traffic. However, predicting traffic volume time series is challenging because they are influenced by various environmental, social, and economic factors, and thus exhibit high nonlinearity and stochasticity. Accurate prediction of highway traffic, which forms the backbone of a nation's transportation and logistics, is crucial. Therefore, this study examined whether meteorological, environmental, and economic covariates improve nationwide highway traffic forecasts in South Korea beyond what calendar information alone provides. Our dataset consisted of daily traffic volumes aggregated over all expressways, together with calendar variables and additional factors such as temperature, precipitation, particulate matter concentrations, and fuel prices. Using these inputs, we compared twelve models spanning naive, statistical, ensemble machine learning, recurrent, convolutional, and Transformer families. Across these models, all evaluated on an identical test period, calendar variables accounted for essentially all of the achievable accuracy, reducing the error of the best model from 4.40% to 3.67% mean absolute percentage error (MAPE), whereas single-station metropolitan weather, air-quality, and fuel-price series yielded no measurable improvement. Among the evaluated models, the temporal convolutional network (TCN) achieved the highest accuracy. This result indicates that, for daily policy-oriented forecasting, the choice of calendar inputs and an appropriate level of model complexity matter more than architectural sophistication alone.

KEYWORDS: Time series forecasting; deep sequence modeling; temporal convolutional network; holiday effects; ablation analysis; nationwide demand forecasting

1 Introduction

Traffic volume prediction has become a significant concern in modern society [1]. Road traffic congestion results in substantial economic and social costs, and accurate predictions are essential for effective transportation management and planning [2]. Traffic prediction technologies are critical components of intelligent transportation systems. These systems analyze historical and real-time data for forecasting future traffic states, enabling authorities to implement proactive strategies, such as adaptive control of traffic lights, ramp metering, and displaying messages on electronic signs [3–5]. Such measures allow for the efficient management of vehicle flow, alleviation of congestion, and enhancement of overall road safety. The economic impacts of traffic congestion are profound; for instance, in the United States alone, annual economic losses owing to traffic delays are estimated at approximately \$70.4 billion [6]. Consequently, developing and

enhancing accurate traffic prediction techniques are vital for ensuring smooth traffic operations, energy conservation, and environmental improvement.

However, traffic volume time series are notoriously complex and characterized by high nonlinearity and stochasticity. They exhibit clear temporal periodicities, such as daily and weekly cycles, but are also significantly influenced by non-recurrent and external factors, such as weather conditions, accidents, and special events.

Meanwhile, the unique transportation environment in South Korea is significantly impacted by special events such as public and traditional holidays. Traffic volumes surge dramatically during major holiday periods because of nationwide travel associated with homecoming and returning. For instance, a previous study documented that the holiday Chuseok, known as Korean Thanksgiving, induces large-scale population movement, with approximately 35 million individuals traveling nationwide during this period, leading to severe traffic congestion [7]. In addition to these cultural events, environmental factors such as weather and air quality play a crucial role. For example, precipitation is another critical variable that directly impacts traffic flow. Rainfall often leads to reduced vehicle speeds and lower road capacity, as drivers tend to become more cautious. For instance, studies have shown that road speeds can decrease by over 6% on weekdays owing to rain, whereas other studies have indicated that traffic flow can decrease by 4%–7% in light rain and up to 14% in heavy rain [8,9].

Severe air pollution caused by particulate matter (PM), often referred to as fine dust, has recently emerged as a critical factor. Vehicle emissions are a primary source of urban air pollutants, including PM, and they establish a direct link between higher traffic volumes and increased pollution. However, high concentrations of PM have been shown to influence public behavior, deterring nonessential travel and outdoor activities, which in turn can temporarily reduce traffic volume [10]. Effectively modeling these complex, bidirectional, and impactful fluctuations is a key challenge in developing a robust prediction system for a specific cultural context such as South Korea.

Studies documenting weather, air-quality, and holiday effects on traffic conclude that such covariates should be incorporated into forecasting models, but this conclusion rests on association at the segment or city level and has not been tested against the marginal predictive value of these covariates at the scale on which national demand forecasts are produced. This study tests it, using national-scale traffic volume data from South Korea together with calendar, meteorological, environmental, and economic covariates, and measuring the contribution of each covariate group through controlled removal under a protocol held identical across all evaluated architectures. While previous research primarily focused on individual cities or specific road segments, this study captured patterns on a broader spatial scale by examining nationwide highway traffic flows. Our dataset includes time-ordered traffic volume records along with various auxiliary variables such as the day of the week, holiday indicators, temperature, precipitation, and PM concentrations, enabling an integrated consideration of diverse factors influencing traffic volumes. Utilizing historical time-series data spanning the previous 30 to 60 days, we implemented representative deep learning architectures, including Long Short-Term Memory (LSTM), Gated Recurrent Units (GRU), and Temporal Convolutional Network (TCN), as well as three Transformer-based models—Temporal Fusion Transformer (TFT), Autoformer, and PatchTST—to predict future traffic volumes. These were compared against six reference models spanning naive, statistical, and ensemble machine learning families under the same evaluation protocol. By comparatively analyzing each model's predictive performance under consistent experimental settings, we evaluate the effectiveness and characteristics of recurrent, convolutional, and Transformer-based approaches using the same nationwide dataset.

In summary, the contributions of this study are as follows: (1) We test whether meteorological, environmental, and economic covariates carry marginal predictive value for day-ahead forecasts of nationwide aggregated expressway demand once deterministic calendar information is available, a question that prior work addresses only through association at the segment or city level. (2) We provide a controlled comparison of twelve forecasting models spanning naive, statistical, ensemble-machine-learning, recurrent, convolutional, and Transformer families, all evaluated on an identical test period under a common evaluation protocol. (3) We show that calendar variables account for essentially all of the achievable accuracy, reducing the error of the best model from 4.40% to 3.67% MAPE, while temperature, precipitation, particulate matter, and fuel price, as represented by single-station metropolitan proxies, yield no measurable improvement. (4) We show that the lowest error is obtained by a comparatively simple convolutional model at a training period of 546 days, and that none of the three Transformer architectures evaluated improves upon it. (5) Through a residual decomposition by day type, we localize the unexplained variance to a small number of holiday periods, where the error reaches 8.54% against 3.41% on normal days, indicating where further modeling effort would yield the largest return. For practice, these results indicate that nationwide daily demand forecasting did not benefit from the weather, air-quality, or fuel-price series examined here.

The remainder of this paper is organized as follows: [Section 2](#) provides a review of related studies on traffic volume prediction, covering conventional and state-of-the-art deep learning approaches. [Section 3](#) describes the dataset and the proposed methodology, including the architecture of the deep learning models used. [Section 4](#) presents the experimental results and a comparative analysis of the model performance. Finally, [Section 5](#) concludes the paper and discusses future research directions.

2 Related Works

Over the past few decades, various methodologies have been proposed for traffic volume prediction. These can be broadly categorized into traditional statistical models, conventional machine-learning techniques, and, more recently, deep-learning-based approaches.

2.1 Traditional Statistical and Machine Learning Approaches

Traditional statistical models, such as autoregressive integrated moving average (ARIMA), have been widely employed because of their capability to deliver stable short-term predictions even with relatively limited data [11]. ARIMA and its variants, such as seasonal ARIMA (SARIMA), are parametric models that assume a linear relationship within time series data. Although effective at capturing trends and seasonality, their performance degrades when faced with the high nonlinearity and complex patterns inherent in real-world traffic data.

To address these limitations, nonparametric machine learning techniques, including multilayer perceptron (MLP) [12] and support vector regression (SVR) [13], have also been utilized to capture complex patterns. These models offer greater flexibility in modeling nonlinear relationships. However, their predictive power is often heavily dependent on manual feature engineering, in which domain expertise is required to select and construct relevant input features. This dependency can be a significant bottleneck and may limit the model's ability to generalize to unseen patterns. In fact, a recent short-term traffic prediction study reported that an MLP-based model incorporating holiday variables achieved superior accuracy compared with Seasonal ARIMA and SVR, highlighting the advantages of artificial intelligence approaches over conventional methods [14,15].

2.2 Deep-Learning-Based Approaches

Recently, deep-learning-based methodologies have emerged prominently in the field of traffic volume prediction [16]. Advances in deep learning have enabled the automatic learning of complex features and effective modeling of time series patterns, significantly enhancing the prediction performance of traffic data. Unlike traditional methods, deep learning approaches have the capability to learn hierarchical feature representations automatically from raw input, thereby reducing dependency on manual feature engineering.

Deep neural networks [17] and recurrent neural networks (RNNs) [18] were among the first deep learning architectures applied to this problem. In particular, RNNs are well-suited for sequential data, as their internal structure allows them to maintain a memory of past information. However, standard RNN architectures are affected by the vanishing gradient problem, which constrains their effectiveness in learning long-range dependencies. To overcome this problem, advanced architectures such as LSTM networks and GRU were developed [19]. These networks incorporate gating mechanisms that allow the model to manage information retention or dismissal across extended durations, rendering them highly effective for time series predictions. Empirical studies have shown that while ARIMA models exhibit limited predictive capability, deep learning models, particularly LSTM and GRU architectures, markedly improve prediction performance [20,21]. These findings align with research, indicating the superior performance and generalization capabilities of deep learning techniques for large-scale complex datasets.

Furthermore, the diversity of deep-learning models for time series analysis is increasing. Beyond conventional sequential RNNs, such as LSTM and GRU, the TCN, based on convolutional neural networks, has recently garnered attention as an effective alternative [22]. A TCN employs a dilated convolutional structure, enabling the learning of long-term temporal dependencies and faster and more stable parallel processing capabilities compared with recurrent networks, which process data sequentially. Some studies have reported that a TCN achieves better prediction accuracy than LSTM in time series prediction tasks, positioning it as a promising alternative in terms of performance and computational efficiency [23,24]. Notably, the dilated convolutional structure of the TCN provides a fixed receptive field and parameter efficiency, which can be advantageous for stable forecasting when training data are limited. Such deep-learning-based approaches are becoming increasingly mainstream in traffic prediction and are applied in various scenarios, including intersection network traffic prediction, highway traffic prediction, and urban road traffic prediction [25–27].

In parallel with these developments, attention-based and Transformer architectures [28] have recently been introduced into the field of traffic forecasting. Inspired by advances in natural language processing, several studies have proposed spatio-temporal attention networks that selectively focus on informative road segments and time steps, demonstrating improved long-horizon prediction performance compared with conventional RNN-based models [29,30]. More recently, Temporal Fusion Transformers (TFT) have been applied to time series forecasting tasks because of their ability to handle heterogeneous static and time-varying covariates and to provide interpretable multi-horizon predictions [31]. In addition, long-horizon time series Transformer models such as Autoformer and PatchTST have shown state-of-the-art performance on benchmark datasets that include traffic-related series by leveraging series decomposition and patch-based representations, respectively [32,33]. However, most existing Transformer-based approaches either focus on short-term horizons measured in minutes or hours or evaluate generic traffic benchmarks in isolation, with relatively few studies investigating nationwide, day-ahead highway traffic forecasting that jointly accounts for socio-economic calendar effects and environmental factors. Whether these architectures are in fact advantageous remains contested. Zeng et al. reported that substantially simpler linear models can match or exceed Transformer-based forecasters on standard long-horizon benchmarks, and the conditions under which attention-based architectures confer an actual advantage are still debated [34]. Settings that differ from

the standard benchmarks in aggregation level, sample size, and covariate structure are informative for this question, and nationwide daily demand forecasting is one such setting.

2.3 Modeling of Holiday Effects

A key challenge in traffic prediction is accounting for special events such as public holidays. Traffic patterns during these periods often deviate significantly from the norm, making prediction by standard models difficult. Recent research underscores that holiday traffic patterns differ substantially from regular weekday/weekend patterns and thus need to be considered explicitly within prediction models [15,35]. Similarly, previous studies have indicated that conventional prediction models exhibit reduced accuracy during holiday periods, emphasizing the need for approaches that either handle holiday data separately or incorporate features derived from analogous historical holidays to improve prediction performance [36]. This reveals a clear opportunity for developing models that can dynamically learn from and adapt to irregular and culturally specific events to ensure year-round accuracy.

Consequently, incorporating recurrent traditional holiday effects, such as South Korea's Seollal (Lunar New Year) and Chuseok, is crucial for enhancing nationwide traffic prediction accuracy. Building on these findings, the present study explicitly incorporates public and traditional holiday indicators into nationwide highway traffic forecasting models, allowing for a systematic comparison of how different deep learning architectures exploit such socio-cultural calendar effects.

2.4 Impact of Environmental Factors on Traffic

In addition to event-based factors, environmental conditions are increasingly recognized as significant external variables in traffic prediction. Air quality, specifically fine dust (e.g., PM10 and PM2.5) concentration, has become a major public health concern in South Korea, directly influencing citizens' travel behavior [10].

The relationship between traffic and PM is complex and bidirectional. On one hand, traffic volume is a primary contributor to PM concentrations in urban areas, as vehicle emissions, secondary PM formation, and the resuspension of road dust all add to ambient particulate levels, and studies in Seoul have noted a measurable association between traffic volume and PM2.5 concentrations. On the other hand, high concentrations of PM may deter outdoor activities and non-essential travel, particularly discretionary trips such as leisure or recreation, as opposed to obligatory weekday commutes. This behavioral response suggests that citizens may refrain from traveling on days with high PM levels to avoid exposure to air pollution.

Whether air-quality data improve traffic forecasts, however, does not follow from these associations. Most existing models account for basic weather conditions such as temperature and rainfall, and few include PM concentrations; where they do, the stated justification is the documented behavioral association rather than a measured gain in forecast accuracy. The premise that these variables belong in forecasting models therefore remains untested, particularly at the nationwide daily scale. Accordingly, this study includes temperature, precipitation, particulate matter concentrations, and fuel prices as additional input variables in modern deep learning models and systematically examines whether these meteorological, environmental, and economic factors meaningfully improve short-term, day-ahead highway traffic forecasts beyond what can be achieved using traffic and calendar information alone.

2.5 Research Gap and Positioning of This Study

Table 1 summarizes the studies reviewed above along the dimensions relevant to the present work. Two patterns are apparent. First, the forecasting target is almost always a road segment, corridor, or single city;

nationwide aggregate demand, which is the quantity used for national capacity planning and holiday demand management, is rarely the target. Second, where meteorological or environmental covariates are included, they are included on the basis of documented association with traffic, and their marginal contribution to forecast accuracy is seldom isolated by controlled removal. The recent long-horizon Transformer models [31–33] are evaluated primarily on generic benchmark datasets rather than on tasks combining heterogeneous exogenous covariates with daily aggregation, so their behavior in this setting is not established either.

Table 1: Comparison of representative studies on traffic volume forecasting.

Study	Spatial Unit	Temporal Resolution	Covariates Beyond Traffic History	Covariate Contribution Isolated by Ablation?	Models Compared
[9]	Urban road	Sub-hourly	Rainfall	No	DL
[10]	City (Seoul)	Daily	PM	No (association study)	Statistical
[14,15]	Segment	Hourly	Calendar	No	MLP, SARIMA, SVR
[25–27]	Intersection /highway/urban road	Sub-hourly	None or limited	No	DL
[29,30]	Network segments	Sub-hourly	None	No	Attention-based
[31]	Benchmark series	Varies	Heterogeneous	Partial	TFT
[32,33]	Benchmark series	Varies	None	No	Autoformer, PatchTST
[35]	Region (Taiwan)	Daily	Calendar	No	DL
[36]	Segment	Daily	Multiple	Yes (input selection)	ML
This study	Nationwide aggregate	Daily	Calendar, weather, PM, fuel price	Yes (full factorial ablation)	Twelve models across six families

The contributions listed in Section 1 correspond to these gaps. Contribution (1) tests the untested premise regarding covariate value. Contribution (2) supplies the protocol-matched comparison across model families that is absent from the studies in Table 1. Contributions (3) through (5) report what that test yields, including the negative result on meteorological, environmental, and economic covariates.

3 Data and Methodology

3.1 Data and Variables

This study used highway traffic volume data provided by the Korea Expressway Corporation to develop a highway traffic prediction model for South Korea. The daily data consisted of the departure tollgate, arrival tollgate, departure traffic volume, and arrival traffic volume. This study aimed to predict the highway traffic

volume for all of South Korea for traffic policy utilization; accordingly, the traffic volume data were aggregated on a daily basis. Because the daily records contain both departure and arrival traffic volumes that may differ slightly, we used their average as the analysis target. This choice reflects the objective of this study, which is to forecast the total nationwide daily travel demand for policy and capacity-planning purposes, rather than directional flows on specific corridors. At the nationwide aggregation level, the total departure and arrival volumes are nearly identical, since a trip departing from one region arrives in another within the same network, so their average serves as a stable representative measure of overall daily traffic. We note, however, that this averaging removes the directional asymmetry that characterizes holiday travel, in which outbound trips dominate before a holiday and return trips dominate afterward.

The data period spans from 2022 to 2024—this was set considering that the period before 2021 was a social distancing period due to COVID-19 and the possibility that data from that period could distort the training model. For the deep sequence models, data from 1 January 2022, to 30 June 2023, were used for training; data from 1 July to 31 December 2023, for validation; and data from 1 January to 31 December 2024, for testing. The naive, statistical, and ensemble baselines require no validation partition and were therefore fitted on all data through 31 December 2023. Considering the characteristics of each model, the validation data were used for epoch-level early stopping in the LSTM, GRU, and TCN models, whereas in the Transformer-based models they were used to determine the number of training steps by examining the training and validation cost curves. A preliminary experiment (Fig. 1) confirmed that, for the Transformer-based models, the validation loss reached its minimum well before the predefined maximum number of training steps and increased thereafter, indicating overfitting to the validation period. We therefore adopted a fixed number of training steps rather than validation-based early stopping for these models, while keeping the data partition identical across all architectures. The selected value of 100 training steps was chosen as a conservative point close to the validation minimum, ensuring that the training loss had sufficiently converged; test performance was insensitive to small variations around this value.

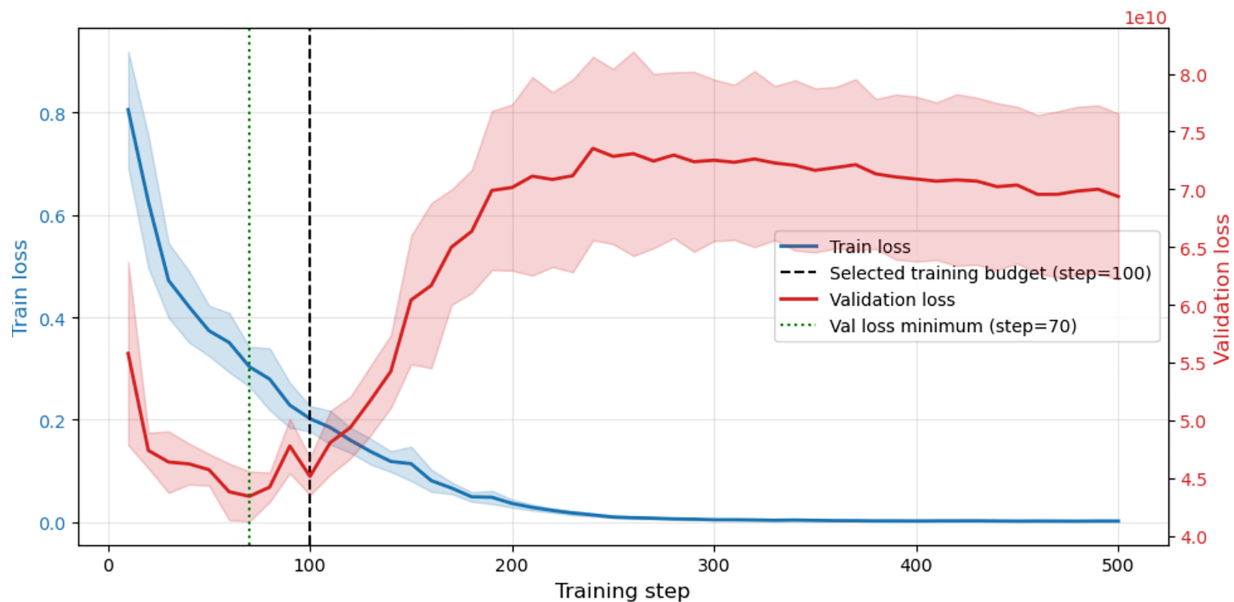


Figure 1: Training and validation loss for Transformer-based models.

Because traffic volume is influenced by previous traffic volume, a sliding window method was used to learn the traffic volume on specific dates. Consequently, for training and validation, periods of 30 and 60 days prior to a specific date were considered as sequence lengths. The calendar variables were encoded as follows. The day of the week was represented as seven binary one-hot variables (*day_Monday* through *day_Sunday*), each equal to 1 on the corresponding weekday and 0 otherwise. Two binary holiday indicators were used. The indicator *is_holiday* equals 1 on statutory public holidays, such as New Year's Day, Independence Movement Day, and Children's Day. The indicator *is_national_hol* equals 1 throughout the traditional Korean holiday periods of Seollal and Chuseok, spanning the full multi-day holiday period including any adjacent weekend days. These two indicators overlap but neither contains the other: weekend days within a traditional holiday period are marked by *is_national_hol* but not by *is_holiday*, whereas ordinary public holidays are marked by *is_holiday* but not by *is_national_hol*. To capture the pre- and post-holiday travel surges, two further binary indicators were derived from the traditional holiday periods: *hol_pre1* and *hol_post1* equal 1 on the days falling one day before and one day after each day of a traditional holiday period, respectively. All calendar variables are deterministic and known in advance for any future date.

Fig. 2 shows the daily highway traffic volume in South Korea from 2022 to 2024 and its monthly aggregated values. Traffic volume clearly exhibits seasonality, continuously increasing from January to May, decreasing in June, then increasing again from July to October, followed by a decrease until December. Fig. 3 illustrates the traffic volume by day of the week for both public and traditional holidays. Generally, the traffic volume from Monday to Thursday was similar, increasing on Friday and then significantly decreasing until Sunday. Furthermore, traffic volume on public holidays (excluding traditional holidays) is generally lower than that on normal days, whereas traffic volume tends to be higher during traditional holidays. Therefore, adding these weekday and situational variables can significantly impact the analysis.

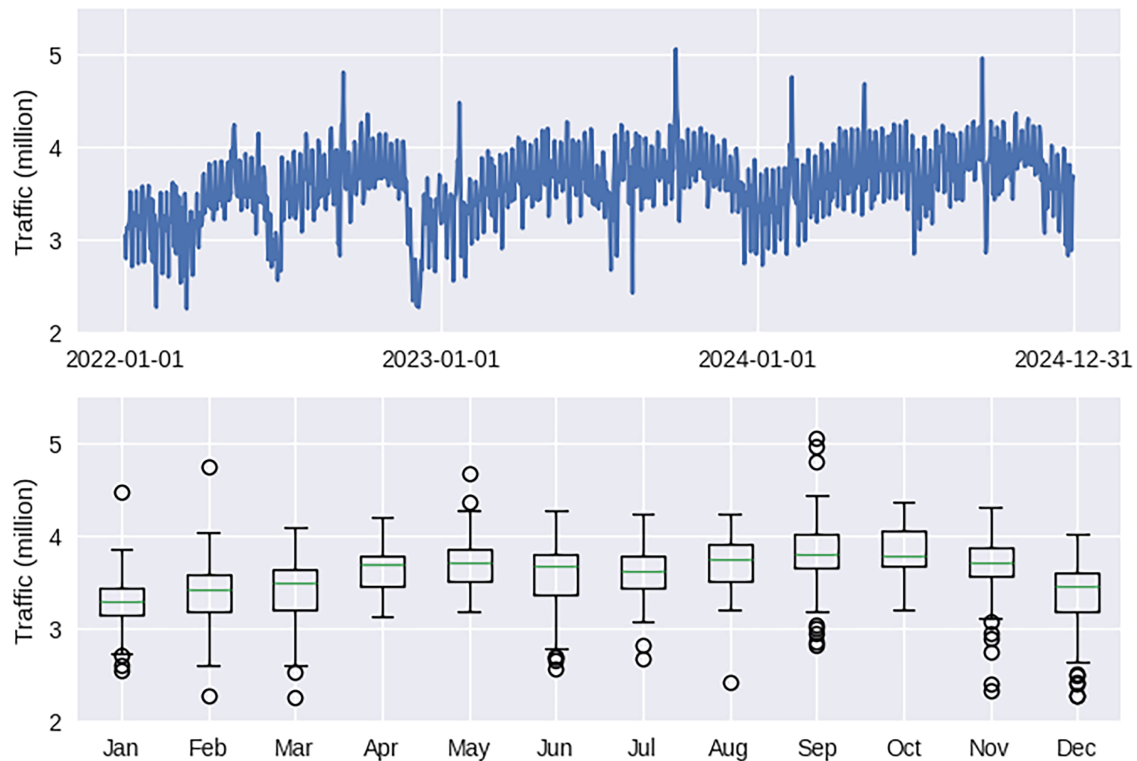


Figure 2: Overall and monthly highway traffic volume.

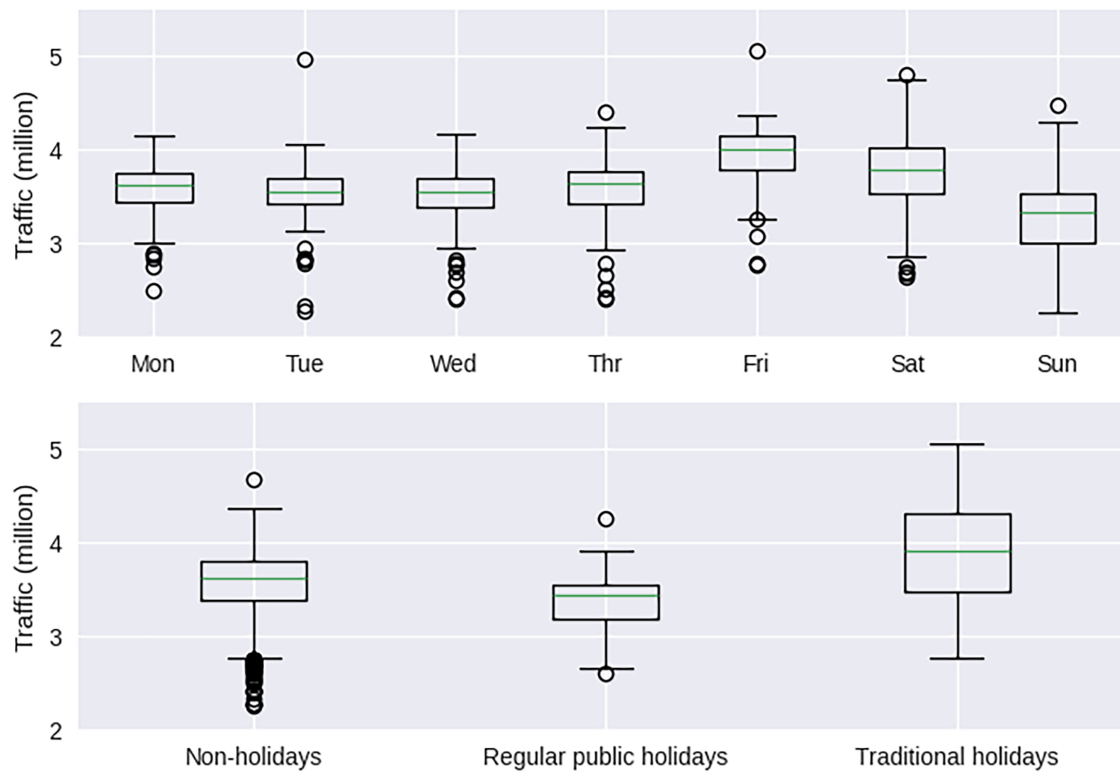


Figure 3: Situational highway traffic volume.

In this study, additional data on temperature, precipitation, PM, and oil prices were used to consider meteorological, environmental, and socio-economic factors. For temperature and precipitation, raw data from the Korea Meteorological Administration were used; for PM, data from the Korea Environment Corporation were used; and for oil prices, data from the Korea National Oil Corporation were used. For temperature and precipitation, data from Jongno-gu, a representative observation station in Seoul, were used; for PM, daily average measurements from Seoul were used, and for oil prices, the Seoul average value for regular gasoline was used as the standard. Seoul was adopted as a representative reference for these environmental variables, as it is the largest population and traffic origin center in South Korea and a substantial portion of nationwide expressway demand originates from or passes through the metropolitan area. Since weather and air quality vary across regions, these single-location measurements serve as an approximation of nationwide conditions, and the implications of this approximation are discussed further in [Sections 4.3](#) and [5](#). These covariates were selected on theoretical grounds rather than by a data-driven screening procedure. Each corresponds to a pathway to travel demand documented in [Sections 2.3](#) and [2.4](#): precipitation affects road capacity and driver caution, particulate matter affects discretionary travel decisions, fuel price affects the marginal cost of driving, temperature affects seasonal and discretionary travel, and the calendar variables encode the socio-cultural travel cycle. No pre-modeling feature selection filter was applied to this set. Because the question addressed in this study is whether these covariates carry predictive value beyond traffic history and calendar information, excluding any of them in advance on the basis of weak marginal association would make that question unanswerable by construction. We therefore retain the full candidate set and measure the contribution of each covariate group by controlled removal, comparing the traffic-only, traffic-plus-calendar, and full configurations reported in [Sections 4.2](#) and [4.3](#).

The variables used in the analysis were traffic volume (*Traffic*), minimum temperature (*MinTa*), maximum temperature (*MaxTa*), precipitation (*Rainfall*), *PM10*, *PM2.5*, and oil price (*OilPrice*). Table 2 presents the descriptive statistics of the variables. As confirmed in Fig. 2, *Traffic* shows seasonality. However, in terms of distribution, its standard deviation is not large compared with its average value of 3.59 million. *Rainfall* is often zero, but its standard deviation is quite large, indicating that it is a highly skewed variable. Fig. 4 shows the correlations between variables. While there was a strong positive correlation between *MinTa* and *MaxTa* and between *PM10* and *PM2.5*, none of the meteorological, environmental, or economic variables shows more than a weak association with traffic volume. Minimum and maximum temperature are weakly positively correlated with traffic, the particulate matter measures are weakly negatively correlated with it, and precipitation and fuel price are essentially uncorrelated with it. This provides an initial indication that the predictive contribution of these covariates may be limited, which the ablation analyses in Sections 4.2 and 4.3 examine directly.

Table 2: Descriptive statistics.

Variable	Mean	SD	Min	25%	50%	75%	Max
Traffic (million)	3.59	0.38	2.25	3.39	3.63	3.81	5.06
MinTa (°C)	10.10	10.98	−17.30	0.70	11.10	19.60	28.20
MaxTa (°C)	18.81	10.65	−8.60	9.38	21.15	27.72	36.40
Rainfall (mm)	4.28	14.95	0.00	0.00	0.00	0.50	176.20
PM10 (µg/m ³)	34.83	22.93	4.42	21.14	29.87	42.15	266.34
PM2.5 (µg/m ³)	18.37	11.73	1.96	10.19	15.85	23.07	91.61
OilPrice (KRW)	1772.80	134.14	1611.99	1680.77	1728.98	1811.65	2212.50

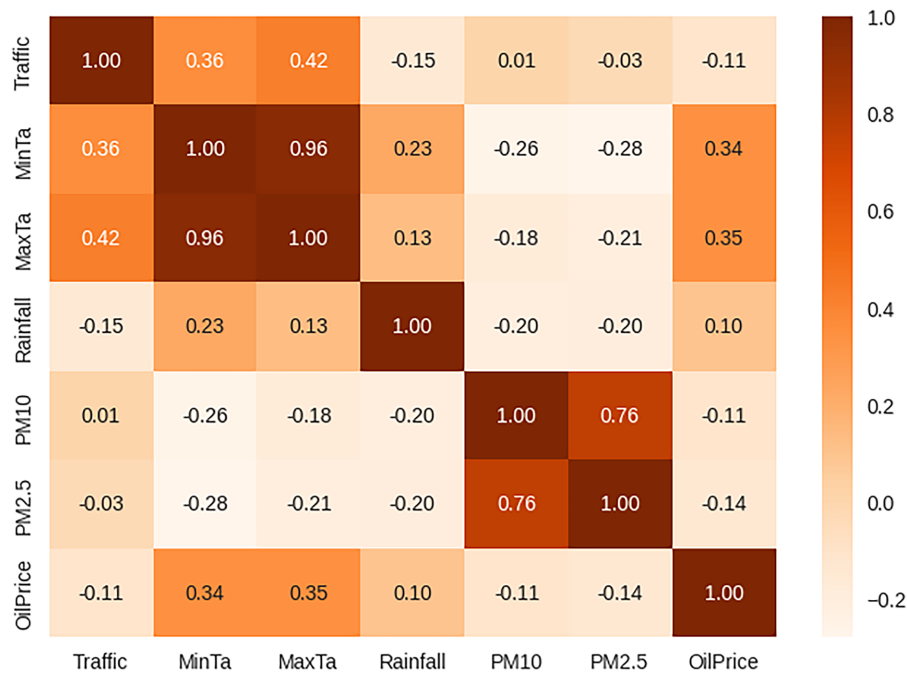


Figure 4: Correlation coefficients.

3.2 Methodology

In this study, RNN, CNN and Transformer-based models, which are core architectures in deep learning, were used for highway traffic prediction models. First, RNNs are suitable for models dealing with time-series data because they were developed considering sequential data processing. In RNNs, representative models such as LSTM and GRU were used. First, LSTM is developed to solve the long-term dependency problem, where information from the beginning of a long sequence is not sufficiently transmitted to the end in basic RNN models. LSTM primarily consists of a cell state, forget gate, input gate, and output gate. The cell state is updated through the results of each gate and plays the role of preserving information and passing it to the next time step. The forget gate determines the extent to which past information is remembered. It applies a sigmoid activation function after linearly transforming the output from the previous time step and current input value. The input gate determines how much of the current information to remember, obtaining values activated by sigmoid and hyperbolic tangent. Subsequently, the cell state is updated by adding the values from the forget and input gates. The output gate determines how much of the updated value from the cell state is passed to the next step. However, GRU is a simpler, improved version of LSTM that includes a reset gate and an update gate. GRU's reset gate combines the influence of the previous hidden state with new input values. The update gate acts similarly to combining LSTM's forget gate and input gate, determining how much information from the previous hidden state is maintained and how many new input values are accepted.

Among CNN models, TCN was designed to process sequential data. Each output value of a layer references only the input values from the previous time step and does not reference data from future time steps. Thus, as the layers progress, a wider range of input information is covered, and the final output layer includes values from all input sequences. In other words, unlike RNN models, TCN can process an entire sequence simultaneously, which has the advantage of shortening the training time. It also has the advantage of being relatively free from gradient, vanishing, exploding, and long-term dependency problems. However, unlike RNN-based models, TCNs do not rely on an explicit recurrent structure; instead, they capture temporal patterns through stacked dilated convolutions with a fixed receptive field. This design offers stable gradient flow and parameter efficiency, which can be advantageous when the available training data are limited.

Recently, Transformer-based models have been increasingly employed in the field of time series forecasting, particularly for long-horizon and large-scale benchmark tasks. Fundamentally, the Transformer architecture incorporates an attention mechanism based on query, key, and value vectors, distinct from the operations performed in RNNs or CNNs [28]. The originally proposed Transformer model consists of an encoder-decoder architecture, in which both components comprise multi-head self-attention mechanisms and feed-forward networks. The final output is processed through a softmax function to yield probability distributions.

This study compares the highway traffic forecasting performance of conventional models against three Transformer-based architectures: Temporal Fusion Transformer (TFT), Autoformer, and PatchTST. The TFT employs a gating mechanism that adaptively regulates network complexity based on the data and context [31]. It also features a variable selection network designed to identify relevant input features at each time step. This architecture enables the TFT to efficiently handle heterogeneous input types, including static covariates, known future inputs, and observed variables, while establishing a temporal processing framework that captures both long- and short-term temporal dependencies. The Autoformer was proposed to address the computational complexity associated with the self-attention mechanism in existing models by introducing an autocorrelation mechanism [32]. This approach allows the model to capture the periodicity inherent in time series data, making it specifically optimized for long-term forecasting tasks. Finally, PatchTST

is designed to process data at the patch level rather than individual points [33]. It utilizes subseries-level patches as input tokens and adopts a channel-independence strategy, where a single Transformer backbone learns each univariate time series independently. This patching design preserves local temporal and semantic information while mitigating noise interference between channels, thereby enhancing the model's generalization capabilities.

We distinguish input variables by their availability at prediction time. The calendar variables are deterministic and fully known in advance, so they are provided as known future inputs for the prediction day. In contrast, the meteorological, environmental, and economic variables (temperature, precipitation, PM, and oil price) are not known for the prediction day in real operational settings. We therefore treated them as past-only observed inputs (historical covariates) wherever the model architecture supports this distinction, as in the TFT, so that no future values of these variables are used. For architectures that do not separate observed and known-future covariates in the NeuralForecast implementation, namely Autoformer, these variables were supplied as future inputs; this gives such models access to additional information at prediction time. As reported in Section 4, even under this more favorable condition these models did not outperform the TCN, which strengthens our conclusion that environmental and economic variables contribute little to day-ahead forecasting. Consistent with this design, PatchTST was run as a univariate model using past traffic only.

The overall research framework is summarized in Fig. 5. The analysis is conducted following the data acquisition and preprocessing stages. However, since static data composed of dummy variables, such as days of the week and public holidays, are not effectively processed directly by RNN and CNN architectures, they were passed through a separate fully connected layer and subsequently concatenated with the sequential data. To evaluate the predictive performance, we utilized the mean absolute percentage error (*MAPE*), mean absolute error (*MAE*), root mean square error (*RMSE*), and the coefficient of determination (R^2). Given that each metric possesses distinct advantages and limitations, this study adopted this comprehensive set of evaluation criteria. The specific calculation formulas are presented below:

$$MAPE = \frac{100}{n} \sum_{i=1}^n \left| \frac{y_i - \hat{y}_i}{y_i} \right|, MAE = \frac{1}{n} \sum_{i=1}^n |y_i - \hat{y}_i|$$

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2}, \text{ and } R^2 = 1 - \frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{\sum_{i=1}^n (y_i - \bar{y})^2}$$

where n represents the total number of data points, y_i is the observed value (ground truth) of the i -th data point, $\hat{y}_i$ is the corresponding prediction value from the model, and $\bar{y}$ denotes the mean of the observed values.

The input sequence length was set to 30 days, with the models designed to predict the traffic volume for the subsequent day. To address the issue of varying scales among variables, we employed the StandardScaler from the scikit-learn library. To prevent any information leakage from the future, the data were partitioned strictly in chronological order, and all preprocessing statistics, including the StandardScaler parameters, were computed using the training period alone and then applied to the validation and test periods. Predictions on the test period were generated in a one-step-ahead manner. Except for Autoformer under the full configuration, each forecast relied only on information available up to the day before the target date. To ensure reproducibility, all deep sequence model experiments were repeated five times with fixed random seeds, and the results are reported as the mean and standard deviation. The specific hyperparameters applied to each model are detailed in Table 3.

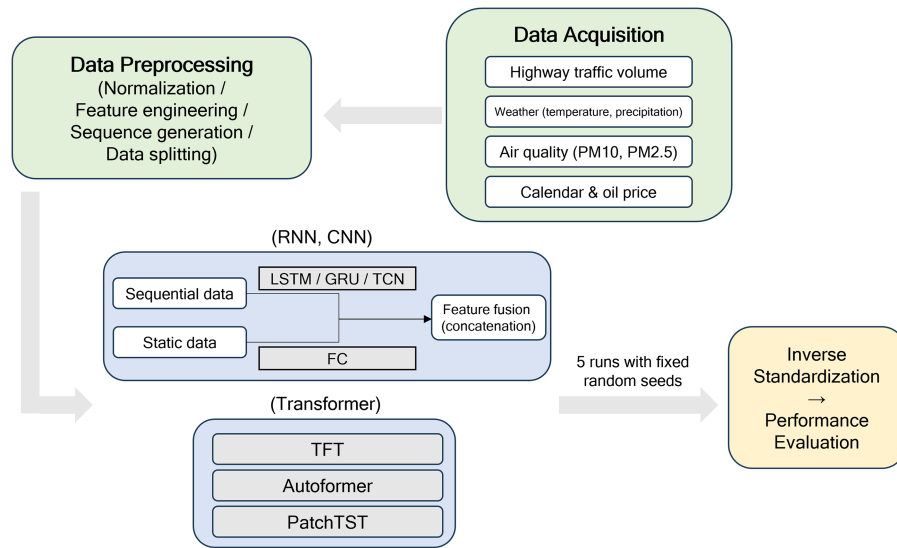


Figure 5: Overall research framework for nationwide highway traffic forecasting.

Table 3: Hyperparameters of the models.

Model	Hyperparameter	Value
Common configurations	Sequence length	30-day
	Batch size	32
	Optimizer	Adam (lr = 0.001)
RNN (LSTM, GRU)	Hidden size (sequential data)	64
	Layers (sequential data)	5
	Hidden size (static data)	32/16
	Dropout rate	0.2
	Epochs	100 (with 10 patience)
TCN	Filters (sequential data)	32
	Layers (sequential data)	4
	Kernel size	2
	Dilation factors	[1, 2, 4, 8]
	Hidden size (static data)	32/16
	Dropout rate	0.2
	Epochs	100 (with 10 patience)
TFT	Hidden size	128
	Attention heads	4
	LSTM layers (encoder/decoder)	1
	Dropout rate	0.1
	Max steps	100

(Continued)

Table 3 (continued)

Model	Hyperparameter	Value
Autoformer	Hidden size	128
	Attention heads	4
	Encoder layers	2
	Decoder layers	1
	Dropout rate	0.05
	Max steps	100
PatchTST	Patch length	16
	Stride	8
	Hidden size	128
	Attention heads	16
	Encoder layers	3
	Dropout rate	0.2
	Max steps	100

Note: Hyperparameters for the Transformer-based models retain the NeuralForecast implementation defaults. Those for the LSTM, GRU, and TCN were set to values commonly used in prior time series forecasting studies and applied uniformly across the three models.

The LSTM, GRU, and TCN models were implemented directly in PyTorch, and the TFT, Autoformer, and PatchTST models using the NeuralForecast library. For the Transformer-based models, the architectural hyperparameters retain the defaults of that implementation. For the LSTM, GRU, and TCN, they were set to values commonly used in prior time series forecasting studies and applied identically across the three models. No model-specific tuning was performed. The settings common to all architectures, namely the optimizer, learning rate, batch size, input window length, data partition, and scaling procedure, were held identical, so that no architecture received more tuning attention than another. No test-set information entered the specification of any model. For the Transformer-based models, the number of training steps was set with reference to the training and validation loss curves in [Fig. 1](#).

4 Results

4.1 Performance Comparison

To provide a single point of comparison, we evaluated twelve models spanning naive, statistical, ensemble-machine-learning, recurrent, convolutional, and Transformer families. All models were assessed on the same test period using the same four evaluation metrics, and all models that use calendar features received an identical calendar feature set. The naive and statistical baselines are a seasonal naive predictor with a lag of seven days, SARIMA, and Prophet; the ensemble machine learning baselines are random forest, XGBoost, and LightGBM; and the deep sequence models are the LSTM, GRU, TCN, TFT, Autoformer, and PatchTST. The baselines are intended as reference points against which the deep sequence models can be assessed, and were therefore run with commonly used settings rather than tuned to match the deep models; the input configuration supplied to each model is stated in [Table 4](#). Because the baselines require no validation partition, they were fitted on all data through the end of 2023, whereas the deep sequence models reserved the second half of 2023 for validation.

Table 4: Forecasting performance of all evaluated models on the 2024 test period.

Model	Family	Configuration	RMSE	MAE	MAPE (%)	R ²
LSTM	Recurrent	30-day/Traffic	193,650 ± 1407	143,010 ± 978	3.88 ± 0.03	0.653 ± 0.005
GRU	Recurrent	+ calendar	194,291 ± 2891	143,220 ± 2206	3.90 ± 0.06	0.651 ± 0.011
TCN	Convolutional	features/	188,665 ± 2929	135,284 ± 2895	3.67 ± 0.07	0.671 ± 0.010
TFT	Transformer	National holiday	194,338 ± 1443	137,260 ± 2407	3.77 ± 0.07	0.650 ± 0.005
Autoformer	Transformer	± 1-day	223,191 ± 4467	164,795 ± 4552	4.50 ± 0.14	0.538 ± 0.019
PatchTST	Transformer	30-day/Traffic only	240,292 ± 3572	168,183 ± 1250	4.58 ± 0.04	0.464 ± 0.016
XGBoost	Ensemble ML	Traffic (7 lags) + calendar features	203,800	150,809	4.09	0.615
LightGBM	Ensemble ML		215,156	151,835	4.13	0.571
Random Forest	Ensemble ML		206,614	152,494	4.14	0.604
Prophet	Statistical	Traffic + holidays	266,294	209,661	5.82	0.342
SARIMA (1,1,1)(1,0,1) ₇	Statistical	Traffic only	227,434	159,965	4.37	0.520
Seasonal Naive (lag-7)	Naive		287,036	189,260	5.18	0.236

Note: All models were evaluated on the same test period (1 January to 31 December 2024). Where “calendar” is indicated, the feature set is identical and comprises seven weekday indicators, statutory and traditional holiday indicators, and the holiday ± 1-day indicators. Prophet is the exception: it received the statutory and traditional holiday dates as holiday regressors and estimated weekly and yearly seasonality internally, so it did not use the weekday indicators or the holiday ± 1-day indicators. PatchTST does not accept exogenous covariates in the implementation used and is therefore reported on the traffic-only input set; its figure is not directly comparable to the calendar-informed configurations. The naive, statistical, and ensemble models serve as reference points and were run with commonly used settings rather than tuned to match the deep sequence models; the ensemble baselines therefore use seven traffic lags, while the deep sequence models use a 30-day input window. The deep sequence models were trained on data through 30 June 2023, with the remainder of 2023 used for validation; the baselines require no validation partition and were fitted on all data through 31 December 2023, and therefore received a longer training record. Deep sequence model results are the mean and standard deviation of five independent runs. The ensemble baselines were fitted with fixed random seeds; XGBoost and LightGBM are deterministic under the settings used, and repeated runs of the random forest varied by no more than 0.01 percentage points of MAPE. Prophet was fitted with multiplicative seasonality and a changepoint prior scale of 0.01. SARIMA forecasts were generated by one-step-ahead walk-forward re-estimation. Models are grouped by family rather than by accuracy; the best value on each metric is in bold.

Across the full set of twelve models, the TCN attained the lowest error on all four metrics, with an RMSE of 188,665 ± 2929, an MAE of 135,284 ± 2895, a MAPE of 3.67 ± 0.07%, and an R² of 0.671 ± 0.010 across five independent runs. Ordered by MAPE, the ranking is the TCN (3.67%), the TFT (3.77%), the LSTM (3.88%), the GRU (3.90%), XGBoost (4.09%), LightGBM (4.13%), the random forest (4.14%), SARIMA (4.37%), the Autoformer (4.50%), PatchTST (4.58%), the seasonal naive predictor (5.18%), and Prophet (5.82%). The margins separating the TCN from the LSTM, GRU, and TFT lie within approximately one to two standard deviations across five runs, and we therefore do not interpret them as evidence of an architectural advantage. None of the three Transformer architectures, all designed for long-horizon forecasting on larger datasets, improved upon a comparatively simple convolutional model at this sample size. The four strongest deep sequence models do, however, separate clearly from the naive and statistical baselines, and the ensemble regressors are competitive under commonly used settings.

4.2 Ablation Studies

An ablation study was conducted to validate the predictive performance of the proposed model configuration. This analysis focused on two primary aspects: input sequence length and the combination of input variables. Accordingly, in addition to the 30-day configuration, we evaluated a 60-day configuration, and for the input variables, we evaluated a configuration utilizing solely historical traffic data, excluding calendar features. Table 5 presents the results of the ablation study.

Table 5: Performance comparison across different input data configurations.

Model	Configuration	MAPE (%)	R ²	Configuration	MAPE (%)	R ²
LSTM	30-day/Traffic only/National holiday $\pm$ 0-day	6.16 $\pm$ 0.13	0.137 $\pm$ 0.021	30-day/Traffic only/National holiday $\pm$ 1-day	6.16 $\pm$ 0.13	0.137 $\pm$ 0.021
GRU		6.01 $\pm$ 0.04	0.170 $\pm$ 0.004		6.01 $\pm$ 0.04	0.170 $\pm$ 0.004
TCN		4.32 $\pm$ 0.11	0.545 $\pm$ 0.014		4.40 $\pm$ 0.09	0.532 $\pm$ 0.014
TFT		5.72 $\pm$ 0.29	0.241 $\pm$ 0.062		5.72 $\pm$ 0.29	0.241 $\pm$ 0.062
Autoformer		5.29 $\pm$ 0.21	0.354 $\pm$ 0.049		5.27 $\pm$ 0.21	0.355 $\pm$ 0.049
PatchTST		4.58 $\pm$ 0.04	0.464 $\pm$ 0.016		4.58 $\pm$ 0.04	0.464 $\pm$ 0.016
LSTM	30-day/Traffic + calendar features/National holiday $\pm$ 0-day	3.87 $\pm$ 0.11	0.635 $\pm$ 0.014	30-day/Traffic + calendar features/National holiday $\pm$ 1-day	3.88 $\pm$ 0.03	0.653 $\pm$ 0.005
GRU		3.91 $\pm$ 0.11	0.629 $\pm$ 0.014		3.90 $\pm$ 0.06	0.651 $\pm$ 0.011
TCN		3.74 $\pm$ 0.06	0.645 $\pm$ 0.009		3.67 $\pm$ 0.07	0.671 $\pm$ 0.010
TFT		3.72 $\pm$ 0.15	0.641 $\pm$ 0.018		3.77 $\pm$ 0.07	0.650 $\pm$ 0.005
Autoformer		4.78 $\pm$ 0.14	0.475 $\pm$ 0.029		4.50 $\pm$ 0.14	0.538 $\pm$ 0.019
LSTM	60-day/Traffic only/National holiday $\pm$ 0-day	6.05 $\pm$ 0.05	0.156 $\pm$ 0.010	60-day/Traffic only/National holiday $\pm$ 1-day	6.05 $\pm$ 0.05	0.156 $\pm$ 0.010
GRU		6.01 $\pm$ 0.01	0.172 $\pm$ 0.003		6.01 $\pm$ 0.01	0.172 $\pm$ 0.003
TCN		4.30 $\pm$ 0.03	0.543 $\pm$ 0.005		4.36 $\pm$ 0.07	0.536 $\pm$ 0.010
TFT		6.20 $\pm$ 0.12	0.158 $\pm$ 0.032		6.20 $\pm$ 0.12	0.158 $\pm$ 0.032
Autoformer		5.50 $\pm$ 0.31	0.347 $\pm$ 0.063		5.58 $\pm$ 0.35	0.336 $\pm$ 0.068
PatchTST		4.70 $\pm$ 0.11	0.477 $\pm$ 0.012		4.70 $\pm$ 0.11	0.477 $\pm$ 0.012
LSTM	60-day/Traffic + calendar features/National holiday $\pm$ 0-day	3.95 $\pm$ 0.09	0.622 $\pm$ 0.009	60-day/Traffic + calendar features/National holiday $\pm$ 1-day	3.84 $\pm$ 0.10	0.661 $\pm$ 0.009
GRU		4.04 $\pm$ 0.12	0.606 $\pm$ 0.008		4.00 $\pm$ 0.18	0.634 $\pm$ 0.018
TCN		3.84 $\pm$ 0.13	0.631 $\pm$ 0.012		3.76 $\pm$ 0.11	0.653 $\pm$ 0.014
TFT		4.11 $\pm$ 0.14	0.586 $\pm$ 0.021		3.90 $\pm$ 0.18	0.623 $\pm$ 0.027
Autoformer		5.22 $\pm$ 0.25	0.405 $\pm$ 0.046		5.20 $\pm$ 0.41	0.410 $\pm$ 0.076

Note: The reported values represent the mean of five independent runs with standard deviations. The best results are highlighted in bold. PatchTST is reported only for the “traffic only” configuration, as it does not incorporate exogenous variables and therefore cannot use calendar features.

Regarding the input sequence length, extending the window from 30 to 60 days did not yield a consistent improvement. Under the “traffic + calendar features” configuration, the TCN achieved a MAPE of $3.67 \pm 0.07\%$ with the 30-day window, which was comparable to or slightly better than the $3.76 \pm 0.11\%$ obtained with the 60-day window. A similar pattern held for the other models, indicating that a 30-day window already captures the dominant weekly and holiday-driven periodicity of daily traffic, and that a longer window provides little additional benefit while increasing computational cost. We therefore adopt the 30-day window as the primary configuration.

Regarding the combination of input variables, substantial differences in performance were observed. Adding calendar features to historical traffic data improved the accuracy of every model markedly. For the 30-day configuration, the TCN improved from a MAPE of $4.40 \pm 0.09\%$ in the “traffic only” setting to $3.67 \pm 0.07\%$ once calendar features were added, while its R^2 rose from 0.532 to 0.671. The improvement was even

more pronounced for the TFT, which improved from 5.72% to 3.77%, and the Autoformer, from 5.27% to 4.50%; PatchTST could not be evaluated under this configuration, as it does not accept exogenous covariates in the implementation used. These results confirm that calendar-driven socio-cultural factors, rather than the length of the historical window, are the dominant determinant of nationwide daily traffic.

The ablation study reveals two main findings. First, the combination of input variables had a far greater impact on performance than the input sequence length, underscoring the central role of calendar information. Second, the TCN achieved the best or near-best performance across all configurations, including the “traffic only” setting (MAPE 4.30% to 4.40%). The behavior of the Transformer-based models in this setting differed substantially between architectures: PatchTST recorded a MAPE of 4.58% to 4.70%, the closest of the Transformer-based models to the TCN, whereas the TFT degraded sharply to 5.72% to 6.20%. The dependence on calendar information is therefore not a property of the Transformer family as a whole. This indicates that the TCN extracts useful temporal structure robustly even from raw traffic sequences, and that its advantage is maintained, rather than created, once calendar features are introduced. The TFT, by contrast, depends heavily on calendar features, suggesting that for this short-term, daily aggregated task, the convolutional inductive bias of the TCN is better suited than the attention-based architecture of the TFT.

Finally, we examined the effect of the holiday-adjacent indicators, which mark the days immediately before and after the traditional holiday periods. As shown in Table 5, including these indicators improved the best configuration (TCN, 30-day, traffic + calendar) from a MAPE of 3.74% to 3.67%, reflecting the tendency of holiday travel to peak before or after the official holiday dates. We therefore adopt the configuration including these indicators as the primary setting, while reporting both variants in Table 5.

4.3 Extended Analysis

In this section, we extended our analysis by incorporating meteorological and economic variables, specifically temperature, precipitation, particulate matter, and oil prices, alongside the existing historical traffic data and calendar features. Table 6 presents the results of this extended analysis, covering both 30-day and 60-day input sequence lengths.

Table 6: Experimental results of extended analysis.

Model	Configuration	MAPE (%)	R ²	Configuration	MAPE (%)	R ²
LSTM	30-day/ Full/National holiday ± 0-day	4.70 ± 0.13	0.505 ± 0.020	30-day/ Full/National holiday ± 1-day	4.74 ± 0.20	0.517 ± 0.034
GRU		4.51 ± 0.21	0.542 ± 0.027		4.41 ± 0.10	0.582 ± 0.019
TCN		3.83 ± 0.10	0.640 ± 0.009		3.77 ± 0.14	0.668 ± 0.017
TFT		4.85 ± 0.27	0.446 ± 0.055		4.43 ± 0.17	0.541 ± 0.045
Autoformer		4.75 ± 0.12	0.496 ± 0.029		4.47 ± 0.17	0.556 ± 0.036
LSTM	60-day/ Full/National holiday ± 0-day	4.88 ± 0.25	0.484 ± 0.043	60-day/ Full/National holiday ± 1-day	4.91 ± 0.32	0.484 ± 0.079
GRU		4.64 ± 0.22	0.530 ± 0.032		4.61 ± 0.07	0.554 ± 0.008
TCN		3.94 ± 0.08	0.631 ± 0.009		3.91 ± 0.07	0.652 ± 0.007
TFT		4.66 ± 0.11	0.469 ± 0.014		4.67 ± 0.45	0.493 ± 0.066
Autoformer		4.88 ± 0.26	0.480 ± 0.037		4.80 ± 0.14	0.486 ± 0.023

Note: The reported values represent the mean of five independent runs with standard deviations. The best results are highlighted in bold. PatchTST does not accept exogenous covariates in the implementation used and therefore cannot be evaluated under the full configuration; its “traffic-only” result is reported in Tables 4 and 5.

Within these configurations, the TCN again attained the lowest error across all evaluation metrics. For the 30-day full model, the TCN recorded a MAPE of $3.77 \pm 0.14\%$ and an R² of 0.668 ± 0.017 , outperforming

the TFT ($\text{MAPE } 4.43 \pm 0.17\%$, $R^2 \ 0.541 \pm 0.045$) and the Autoformer ($\text{MAPE } 4.47 \pm 0.17\%$, $R^2 \ 0.556 \pm 0.036$). A consistent pattern was observed for the 60-day configuration, where the TCN attained a MAPE of $3.91 \pm 0.07\%$ and an R^2 of 0.652 ± 0.007 , surpassing the other models evaluated. Notably, whereas the TCN maintained stable performance after the addition of environmental and economic variables, the Transformer-based models did not benefit from these additional inputs.

Notably, the “full model” incorporating all variables failed to outperform the configuration that utilized only traffic data and calendar features. For instance, the best TCN result in the full model (MAPE 3.77%) was essentially identical to, and did not improve upon, its traffic-plus-calendar result (MAPE 3.67%). These findings suggest that meteorological, environmental, and economic features, as represented by single-station metropolitan proxies, do not exert a significant influence on short-term highway traffic forecasting. While such external variables may be critical for long-term traffic volume prediction models, their contribution to short-term forecasts was not detectable with the covariate representation used here. From a modeling perspective, this demonstrates that a lighter model configuration attained the same accuracy as the full specification. Ultimately, the study indicates that high-quality forecasts can be generated using historical traffic data and calendar features alone, without gains from the external data sources examined here.

It should be noted that the environmental and meteorological inputs are derived from a single Seoul-based location, whereas the target is traffic aggregated across all national expressways. This spatial mismatch is a plausible structural explanation for why these variables yielded no measurable improvement, as a single-station signal cannot capture the regional weather heterogeneity that drives demand outside the metropolitan area. We therefore interpret the limited contribution of environmental covariates as a finding specific to nationwide daily aggregation with a metropolitan proxy, rather than as conclusive evidence that weather and air quality are irrelevant to expressway demand in general. Disentangling the effect of the model from that of the spatial aggregation of inputs is an important direction for future work.

Although the TCN provides the best forecasting accuracy, its convolutional structure does not offer direct interpretability of feature contributions. To examine which inputs drive the predictions, we therefore additionally analyzed the variable selection weights and temporal attention of the TFT, which provides built-in interpretability. As shown in Fig. 6, the encoder variable importance is dominated by past traffic volume, followed almost entirely by calendar variables such as the day of the week and national holiday indicators, while meteorological and environmental variables receive negligible weight. This pattern independently corroborates our ablation results, confirming from a model-internal perspective that calendar-based factors, rather than environmental covariates, are the primary drivers of nationwide daily traffic.

Temperature merits specific comment, since it varies substantially over the study period yet contributes nothing to the forecasts. Minimum daily temperature ranges from -17.3°C to 28.2°C and maximum daily temperature from -8.6°C to 36.4°C (Table 2), so the absence of a contribution does not reflect a lack of variation in the variable. A more plausible explanation is redundancy rather than irrelevance. The component of temperature that is informative for nationwide daily demand is its seasonal component, and that component is already carried by the calendar variables and by the 30-day window of past traffic, which itself embeds the seasonal cycle. The mechanisms through which temperature affects traffic operations more directly, such as reductions in speed and effective capacity, are largely intra-day and segment-specific, and these average out under nationwide daily aggregation. The negligible weight assigned to temperature in Fig. 6, together with its weak correlation with traffic volume in Fig. 4, is consistent with this interpretation.

Fig. 7 shows that temporal attention concentrates most strongly on the day two days before the prediction, with a secondary peak around nine days before. The seven-day interval between these two peaks reflects the weekly periodicity of traffic demand, indicating that the model attends both to the most recent days and to the corresponding day of the previous week.

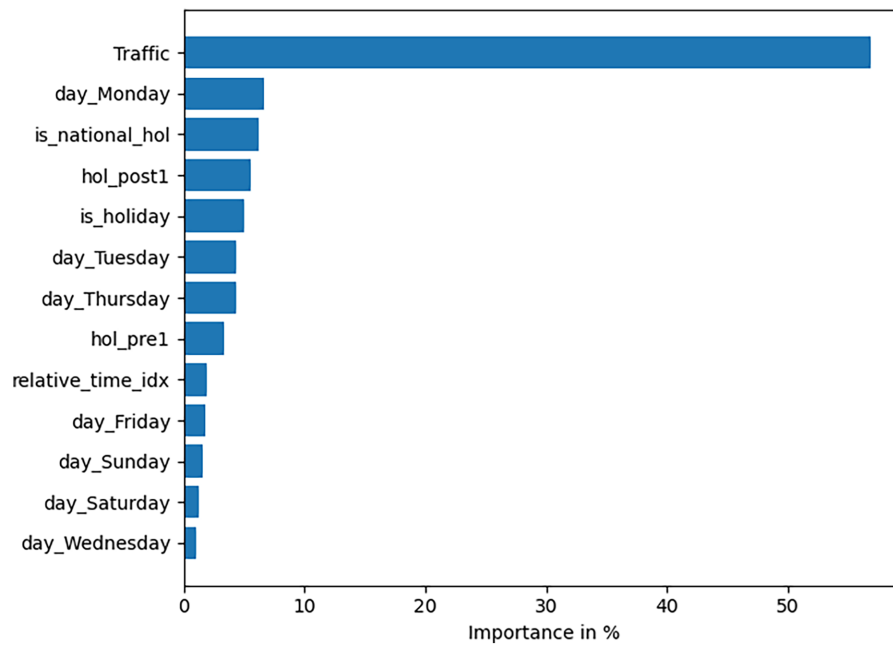


Figure 6: Encoder variables importance.

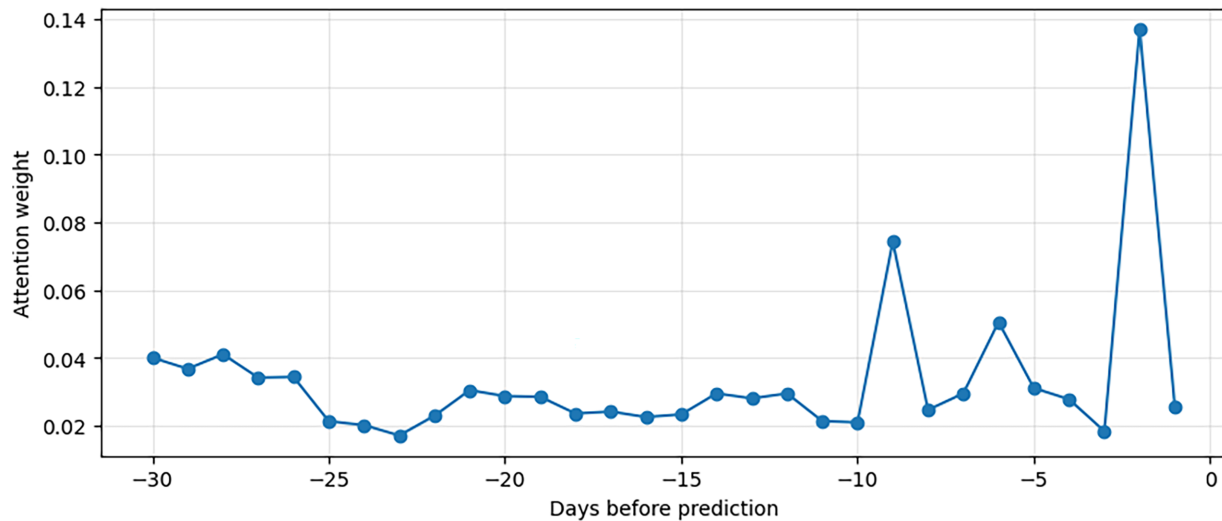


Figure 7: TFT temporal attention weights across the 30-day input window.

4.4 Residual Analysis

Although the best model achieves an R^2 of approximately 0.671, a portion of the daily variance remains unexplained. To identify the source of this residual error, we analyzed the prediction errors of the best model (TCN) on the test period by day type. As summarized in [Table 7](#), the error is highly concentrated around holidays: the mean absolute percentage error reaches 8.54% on national (traditional) holidays, 7.37% on public holidays, and 6.37% on holiday-adjacent days, compared with only 3.41% on normal days. The ten days with the largest errors are almost exclusively traditional holidays and public holidays, such as Chuseok, Seollal, Children's Day, and Christmas, together with their adjacent days. Consistent with this pattern, when

holiday-related days are excluded, the R^2 on normal days rises to approximately 0.72, higher than the value obtained over all days.

Table 7: MAPE of the best model (TCN) by day type.

Day Type	MAPE (%)
Normal day	3.41
Holiday-adjacent	6.37
Public holiday	7.37
National (traditional) holiday	8.54

Note: The R^2 on normal days alone is approximately 0.72, compared with 0.671 over all days.

These results indicate that the unexplained variance is not randomly distributed noise but is structurally concentrated in a small number of irregular, high-variance days. Holiday travel is inherently more volatile, as its exact timing and magnitude vary year to year and depend on factors such as the specific calendar placement of the holiday and multi-day travel behavior, which a daily aggregated model cannot fully resolve. This finding reinforces our central conclusion that calendar-driven events are the dominant determinant of nationwide daily traffic, and it clarifies that the remaining error originates primarily from the intrinsic unpredictability of these few high-demand periods rather than from a general limitation of the model.

5 Conclusion

Traffic prediction is an essential technology for modern transportation systems. However, it remains challenging because of the complexity and variability of traffic data. This study tested whether meteorological, environmental, and economic covariates carry marginal predictive value for day-ahead forecasts of nationwide aggregated expressway demand, using South Korea as a setting in which socio-cultural calendar effects are unusually pronounced. A comprehensive daily dataset was constructed by aggregating traffic volumes over all expressways and augmenting them with calendar variables and additional meteorological, environmental, and economic factors. Based on this dataset, the prediction performance of recurrent neural networks (LSTM and GRU), a TCN, and three Transformer-based models (TFT, Autoformer, and PatchTST) was systematically evaluated under multiple input configurations, together with six naive, statistical, and ensemble machine learning reference models.

To the best of our knowledge, this is the first study to evaluate day-ahead forecasting on nationwide-aggregated daily expressway traffic in South Korea, in contrast to the prevailing focus on individual cities, corridors, or sensor-level segments. This macroscopic scale isolates the influence of country-wide socio-cultural events that are invisible at the segment level. Through a controlled feature-ablation design under a unified train/validation/test protocol, we provide quantitative evidence that calendar-driven socio-cultural factors, namely weekday and Korean holiday cycles such as Seollal and Chuseok, are the dominant determinants of nationwide daily demand, whereas meteorological, environmental, and economic covariates yield no measurable improvement. This counterintuitive null result on environmental covariates is itself a finding, as it contradicts the common assumption that adding more contextual variables improves forecasting. Finally, the lowest forecasting error is achieved not by the most architecturally complex model but by a comparatively simple convolutional model using only traffic and calendar inputs, indicating that for daily, policy-oriented demand forecasting, a parsimonious input configuration and a moderate-complexity model are sufficient, and indeed preferable, relative to elaborate Transformer architectures.

Through our analysis, we identified several key findings. First, models that explicitly capture temporal dependencies achieved lower error than the naive and statistical baselines and than the ensemble regressors under commonly used settings, though the margin over the strongest ensemble baseline was modest. Among all architectures, the TCN, when trained with a 30-day input window using traffic and calendar features, achieved the best overall performance across all error metrics, with a mean absolute percentage error of approximately 3.67%. In contrast, the Transformer-based models, namely the TFT, Autoformer, and PatchTST, which are designed for long-horizon forecasting and emphasize periodic, patch-based, or attention-based representations, did not surpass the recurrent and convolutional baselines in the short-term, day-ahead forecasting setting considered in this study.

Second, using calendar variables alongside traffic volumes markedly improved prediction accuracy compared to relying on traffic volumes alone. However, incorporating meteorological, environmental, and economic variables, as represented by single-station metropolitan proxies, did not meaningfully improve the results. This indicates that calendar-related social patterns are the dominant determinant of nationwide daily traffic volumes, at least relative to the covariate representation available in this study.

Third, the TCN attained the lowest error across the evaluated metrics and configurations, with the recurrent models (LSTM and GRU) close behind, while the more complex Transformer-based models did not improve upon them. A particularly notable observation is that the TCN remained robust even in the “traffic only” setting and after the addition of environmental variables, whereas the TFT degraded sharply without calendar features. These results suggest that model architecture must be carefully aligned with the forecasting task. Since Autoformer and PatchTST are designed primarily for long-horizon sequences, they may be less suited to the short-term, day-ahead predictions required in this study. More broadly, this indicates that architectural complexity is neither necessary nor sufficient for daily aggregated forecasting, and that a parsimonious convolutional model such as the TCN, or even the computationally efficient LSTM and GRU, remains a highly competitive choice for nationwide highway traffic forecasting.

Future research can extend this work by addressing its current methodological boundaries. This study is based on nationwide daily aggregated data, and its forecasts are therefore intended to support medium-term traffic planning and policy decisions, such as anticipating nationwide demand around holidays, rather than real-time or operational traffic control. Real-time applications, such as signal control or ramp metering, require data at an hourly or finer temporal resolution and at the level of individual road segments, which are beyond the scope of the present daily, nationwide setting. Accordingly, an important next step is to evaluate how the meteorological, environmental, and economic factors examined here behave on hourly or sub-daily data, and to incorporate graph-based models that capture regional spatial dependencies. Extending the framework in these directions would be necessary before the approach could inform real-time traffic management. This study also does not evaluate the most recent generation of forecasting architectures, such as linear-decomposition models, inverted Transformers, or pretrained time-series foundation models, and their behavior at this aggregation level and sample size remains an open question. Hyperparameters were set to recommended or default values rather than optimized, so the ordering among the closely spaced models may be sensitive to a systematic search.

In conclusion, this study shows that simply using advanced deep learning models is insufficient. The key to building smarter and more reliable transportation systems is adopting an approach that carefully considers a region's calendar structure and the exceptional periods within it.

Acknowledgement: Not applicable.

Funding Statement: This research was supported by the 2024 Research Fund of University of Ulsan.

Author Contributions: The authors confirm contribution to the paper as follows: Study conceptualization and design: Ohsung Kwon and Kyoung-Soub Lee; data collection: Ohsung Kwon; analysis and interpretation of results: Ohsung Kwon; draft manuscript preparation: Ohsung Kwon and Kyoung-Soub Lee; review & editing: Ohsung Kwon and Kyoung-Soub Lee; supervision: Kyoung-Soub Lee. All authors reviewed and approved the final version of the manuscript.

Availability of Data and Materials: The data that support the findings of this study are available from the corresponding author upon reasonable request at: lks0872@ulsan.ac.kr.

Ethics Approval: Not applicable.

Conflicts of Interest: The authors declare no conflicts of interest.

Abbreviations

The following abbreviations are used in this manuscript:

ARIMA	Autoregressive integrated moving average
CNN	Convolutional neural network
GRU	Gated recurrent units
LSTM	Long short-term memory
MAE	Mean absolute error
MAPE	Mean absolute percentage error
MLP	Multilayer perceptron
PatchTST	Patch time-series transformer
PM	Particulate matter
RMSE	Root mean square error
RNN	Recurrent neural network
SARIMA	Seasonal autoregressive integrated moving average
SVR	Support vector regression
TCN	Temporal convolutional network
TFT	Temporal fusion transformer

References

1. Vlahogianni EI, Karlaftis MG, Golias JC. Short-term traffic forecasting: where we are and where we're going. *Transp Res Part C Emerg Technol.* 2014;43(5):3–19. doi:10.1016/j.trc.2014.01.005.
2. Peled I, Lee K, Jiang Y, Dauwels J, Pereira FC. On the quality requirements of demand prediction for dynamic public transport. *Commun Transp Res.* 2021;1:100008. doi:10.1016/j.commtr.2021.100008.
3. Kumar N, Rahman SS, Dhakad N. Fuzzy inference enabled deep reinforcement learning-based traffic light control for intelligent transportation system. *IEEE Trans Intell Transp Syst.* 2021;22(8):4919–28. doi:10.1109/TITS.2020.2984033.
4. Tajdari F, Roncoli C, Papageorgiou M. Feedback-based ramp metering and lane-changing control with connected and automated vehicles. *IEEE Trans Intell Transp Syst.* 2022;23(2):939–51. doi:10.1109/TITS.2020.3018873.
5. Chatterjee K, McDonald M. Effectiveness of using variable message signs to disseminate dynamic traffic information: evidence from field trials in European cities. *Transp Rev.* 2004;24(5):559–85. doi:10.1080/0144164042000196080.
6. Pishue B. 2023 INRIX global traffic scorecard. 2024 [cited 2026 Aug 20]. Available from: <https://inrix.com/scorecard-2023/>.
7. Ahn YJ, Lee SK, Baek U, Kim JY. Determinants of homebound travel during Korean thanksgiving. *Tour Hosp Res.* 2020;20(3):288–99. doi:10.1177/1467358419848132.
8. Yao Y, Wu D, Hong Y, Chen D, Liang Z, Guan Q, et al. Analyzing the effects of rainfall on urban traffic-congestion bottlenecks. *IEEE J Sel Top Appl Earth Obs Remote Sens.* 2020;13:504–12. doi:10.1109/JSTARS.2020.2966591.

9. Jia Y, Wu J, Xu M. Traffic flow prediction with rainfall impact using a deep learning method. *J Adv Transp*. 2017;2017(722):6575947. doi:10.1155/2017/6575947.
10. Kim SN, Jung S, Joo Y, Kim H. Air pollution hindering a transit-oriented city: examining the association of particulate matter concentration with public transit ridership and road traffic in Seoul, South Korea. *J Public Transp*. 2024;26:100111. doi:10.1016/j.jpubtr.2024.100111.
11. Ahmed MS, Cook AR. Analysis of freeway traffic time-series data by using Box-Jenkins techniques. *Transp Res Rec*. 1979;722:1–9.
12. Mozolin M, Thill JC, Lynn Usery E. Trip distribution forecasting with multilayer perceptron neural networks: a critical evaluation. *Transp Res Part B Methodol*. 2000;34(1):53–73. doi:10.1016/S0191-2615(99)00014-4.
13. Castro-Neto M, Jeong Y, Jeong MK, Han LD. AADT prediction using support vector regression with data-dependent parameters. *Expert Syst Appl*. 2009;36(2):2979–86. doi:10.1016/j.eswa.2008.01.073.
14. Rajalakshmi V, Vaidyanathan SG. Hybrid time-series forecasting models for traffic flow prediction. *Promet Traffic Transp*. 2022;34(4):537–49. doi:10.7307/ptt.v34i4.3998.
15. Luo X, Li D, Zhang S. Traffic flow prediction during the holidays based on DFT and SVR. *J Sens*. 2019;2019(10):6461450. doi:10.1155/2019/6461450.
16. Kashyap AA, Raviraj S, Devarakonda A, Nayak KSR, KV S, Bhat SJ. Traffic flow prediction models—a review of deep learning techniques. *Cogent Eng*. 2022;9(1):2010510. doi:10.1080/23311916.2021.2010510.
17. Nguyen H, Kieu LM, Wen T, Cai C. Deep learning methods in transportation domain: a review. *IET Intell Transp Syst*. 2018;12(9):998–1004. doi:10.1049/iet-its.2018.0064.
18. Osipov V, Nikiforov V, Zhukova N, Miloserdov D. Urban traffic flows forecasting by recurrent neural networks with spiral structures of layers. *Neural Comput Appl*. 2020;32(18):14885–97. doi:10.1007/s00521-020-04843-5.
19. Zhao Z, Chen W, Wu X, Chen PCY, Liu J. LSTM network: a deep learning approach for short-term traffic forecast. *IET Intell Transp Syst*. 2017;11(2):68–75. doi:10.1049/iet-its.2016.0208.
20. Siami-Namini S, Tavakoli N, Siami Namin A. A comparison of ARIMA and LSTM in forecasting time series. In: *Proceedings of the 2018 17th IEEE International Conference on Machine Learning and Applications (ICMLA)*; 2018 Dec 17–20; Orlando, FL, USA. p. 1394–401. doi:10.1109/icmla.2018.00227.
21. Lu S, Zhang Q, Chen G, Seng D. A combined method for short-term traffic flow prediction based on recurrent neural network. *Alex Eng J*. 2021;60(1):87–94. doi:10.1016/j.aej.2020.06.008.
22. Lea C, Flynn MD, Vidal R, Reiter A, Hager GD. Temporal convolutional networks for action segmentation and detection. In: *Proceedings of the 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*; 2017 Jul 21–26; Honolulu, HI, USA. p. 1003–12. doi:10.1109/CVPR.2017.113.
23. Gopali S, Abri F, Siami-Namini S, Namin AS. A comparison of TCN and LSTM models in detecting anomalies in time series data. In: *Proceedings of the 2021 IEEE International Conference on Big Data (Big Data)*; 2021 Dec 15–18; Orlando, FL, USA. p. 2415–20. doi:10.1109/bigdata52589.2021.9671488.
24. Hewage P, Behera A, Trovati M, Pereira E, Ghahremani M, Palmieri F, et al. Temporal convolutional neural (TCN) network for an effective weather forecasting using time-series data from the local weather station. *Soft Comput*. 2020;24(21):16453–82. doi:10.1007/s00500-020-04954-0.
25. Zhao L, Song Y, Zhang C, Liu Y, Wang P, Lin T, et al. T-GCN: a temporal graph convolutional network for traffic prediction. *IEEE Trans Intell Transp Syst*. 2020;21(9):3848–58. doi:10.1109/TITS.2019.2935152.
26. Wang Y, Chen P. Network traffic prediction based on transformer and temporal convolutional network. *PLoS One*. 2025;20(4):e0320368. doi:10.1371/journal.pone.0320368.
27. Zhang Y, Shang K, Cui Z, Zhang Z, Zhang F. Research on traffic flow prediction at intersections based on DT-TCN-attention. *Sensors*. 2023;23(15):6683. doi:10.3390/s23156683.
28. Vaswani A, Shazeer N, Parmar N, Uszkoreit J, Jones L, Gomez AN, et al. Attention is all you need. In: *Proceedings of the 31st International Conference on Neural Information Processing Systems*; 2017 Dec 4–9; Long Beach, CA, USA.
29. Guo S, Lin Y, Feng N, Song C, Wan H. Attention based spatial-temporal graph convolutional networks for traffic flow forecasting. *Proc AAAI Conf Artif Intell*. 2019;33(1):922–9. doi:10.1609/aaai.v33i01.3301922.

30. Shi X, Qi H, Shen Y, Wu G, Yin B. A spatial–temporal attention approach for traffic prediction. *IEEE Trans Intell Transp Syst.* 2021;22(8):4909–18. doi:10.1109/TITS.2020.2983651.
31. Lim B, Arık SÖ, Loeff N, Pfister T. Temporal fusion transformers for interpretable multi-horizon time series forecasting. *Int J Forecast.* 2021;37(4):1748–64. doi:10.1016/j.ijforecast.2021.03.012.
32. Wu H, Xu J, Wang J, Long M. Autoformer: decomposition transformers with auto-correlation for long-term series forecasting. In: *Proceedings of the 35th International Conference on Neural Information Processing Systems*; 2021 Dec 6–14; Virtual.
33. Nie Y, Nguyen NH, Sinthong P, Kalagnanam J. A time series is worth 64 words: long-term forecasting with transformers. In: *Proceedings of the International Conference on Learning Representations 2023*; 2023 May 1–5; Kigali, Rwanda.
34. Zeng A, Chen M, Zhang L, Xu Q. Are transformers effective for time series forecasting? *Proc AAAI Conf Artif Intell.* 2023;37(9):11121–8. doi:10.1609/aaai.v37i9.26317.
35. Ashouri M, Kin Hing Phoa F, Angela Cremona M. Analyzing Taiwanese traffic patterns on consecutive holidays through forecast reconciliation and prediction-based anomaly detection techniques. *IEEE Access.* 2025;13(2):108500–18. doi:10.1109/ACCESS.2025.3581839.
36. Ma D, Song XB, Zhu J, Ma W. Input data selection for daily traffic flow forecasting through contextual mining and intra-day pattern recognition. *Expert Syst Appl.* 2021;176(6):114902. doi:10.1016/j.eswa.2021.114902.