

Article

# Periodic Transformer Encoder for Multi-Horizon Travel Time Prediction

Hui-Ting Christine Lin <sup>1</sup>  and Vincent S. Tseng <sup>1,2,\*</sup> 

<sup>1</sup> Department of Computer Science, National Yang Ming Chiao Tung University, Hsinchu 30010, Taiwan; htlin.c@nycu.edu.tw

<sup>2</sup> Department of Management Information Systems, National Chung Hsing University, Taichung 402, Taiwan

\* Correspondence: vtseng@cs.nycu.edu.tw

**Abstract:** In the domain of Intelligent Transportation Systems (ITS), ensuring reliable travel time predictions is crucial for enhancing the efficiency of transportation management systems and supporting long-term planning. Recent advancements in deep learning have demonstrated the ability to effectively leverage large datasets for accurate travel time predictions. These innovations are particularly vital as they address both short-term and long-term travel demands, which are essential for effective traffic management and scheduled routing planning. Despite advances in deep learning applications for traffic analysis, the dynamic nature of traffic patterns frequently challenges the forecasting capabilities of existing models, especially when forecasting both immediate and future traffic conditions across various time horizons. Additionally, the area of long-term travel time forecasting still remains not fully explored in current research due to these complexities. In response to these challenges, this study introduces the Periodic Transformer Encoder (PTE). PTE is a Transformer-based model designed to enhance traffic time predictions by effectively capturing temporal dependencies across various horizons. Utilizing attention mechanisms, PTE learns from long-range periodic traffic data for handling both short-term and long-term fluctuations. Furthermore, PTE employs a streamlined encoder-only architecture that eliminates the need for a traditional decoder, thus significantly simplifying the model's structure and reducing its computational demands. This architecture enhances both the training efficiency and the performance of direct travel time predictions. With these enhancements, PTE effectively tackles the challenges presented by dynamic traffic patterns, significantly improving prediction performance across multiple time horizons. Comprehensive evaluations on an extensive real-world traffic dataset demonstrate PTE's superior performance in predicting travel times over multiple horizons compared to existing methods. PTE is notably effective in adapting to high-variability road segments and peak traffic hours. These results prove PTE's effectiveness and robustness across diverse traffic environments, indicating its significant contribution to advancing traffic prediction capabilities within ITS.

**Keywords:** Intelligent Transportation Systems; travel time prediction; multi-horizon prediction; periodic data processing; transformer



**Citation:** Lin, H.-T.C.; Tseng, V.S. Periodic Transformer Encoder for Multi-Horizon Travel Time Prediction. *Electronics* **2024**, *13*, 2094. <https://doi.org/10.3390/electronics13112094>

Academic Editor: Ping-Feng Pai

Received: 9 May 2024

Revised: 24 May 2024

Accepted: 25 May 2024

Published: 28 May 2024



**Copyright:** © 2024 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

## 1. Introduction

Intelligent Transportation Systems (ITS) have experienced significant growth driven by increasing urbanization and the increasing demands for efficient transportation management. These systems leverage advanced data analytics and real-time control mechanisms to enhance the reliability, safety, and efficiency of transportation networks. As urban populations grow and transportation becomes more complex, ITS are required to handle not only the immediate fluctuations in traffic but also to anticipate long-term changes and challenges. Consequently, the evolution of these systems demands the development of advanced predictive technologies capable of operating across multiple time horizons. This capability is crucial for effectively managing day-to-day operations while also supporting

strategic decisions that ensure sustainability and resilience in transportation planning. By predicting traffic patterns minutes to hours or even days in advance, ITS can improve traffic flow, reduce congestion, and enhance urban mobility in both the short and long term.

Travel time prediction is a fundamental component of ITS that significantly contributes to the efficiency and reliability of transportation systems. Accurate predictions of travel times are essential for route planning, congestion management, and overall traffic enhancement. For example, they enable commuters to make informed decisions, reduce waiting times, and enhance their commuting experience. As transportation networks grow in complexity, the ability to predict travel times accurately over multiple horizons—from a few minutes ahead to several hours—becomes critical for operational efficiency and strategic planning.

Traffic prediction often employs traditional statistical methods such as Historical Averages (HA) and ARIMA (Autoregressive Integrated Moving Average) due to their straightforward computational frameworks and strong theoretical foundations [1–3]. These methods, particularly effective for simple, short-term forecasts, rely heavily on stable historical data patterns. HA averages historical data for predictions, which is suitable for short-term forecasting. However, these methods face limitations when dealing with multivariate time series or dynamic traffic conditions that require adaptation to rapid changes. These limitations highlight a need for more advanced forecasting techniques to handle the complexity and variability of long-term traffic predictions.

The advent of data-driven approaches, particularly those leveraging machine learning and deep learning, has transformed traffic prediction methodologies. These methods utilize vast amounts of data to learn complex patterns and dependencies, which are often invisible to traditional statistical methods. In traffic prediction, deep learning techniques such as recurrent neural networks (RNNs) and long short-term memory (LSTM) networks have initially shown promise in handling temporal dependencies and variable traffic conditions. Early studies highlighted the efficacy of LSTM networks in predicting traffic flow dynamics [4]. Further research by [5] expanded the use of LSTM models to encompass travel time prediction, validating their effectiveness in practical settings. Subsequent studies have innovated by integrating Convolutional Neural Networks (CNNs) with LSTMs, enabling the simultaneous processing of spatial and temporal traffic data to enhance prediction precision [6]. Moreover, recent advancements have included weather data integration into deep learning frameworks to enhance the accuracy of travel time predictions [7]. Despite these successes, many current models are primarily designed for short-term predictions and face challenges in long-term forecasting due to their limited ability to effectively handle extended-range dependencies.

In recent years, transformer-based methods have revolutionized deep learning, especially in the analysis of sequence data. By utilizing attention mechanisms, these models can interpret complex relationships across extended time periods, establishing their superiority over traditional models in scenarios that demand an understanding of long-term dependencies. Transformers have shown versatility and efficacy in traffic prediction within various aspects of ITS. They excel in applications ranging from traffic speed analysis [8] to travel time prediction [9], and traffic flow forecasting [10–13]. These methods show their strengths in handling diverse data types, demonstrating their capability to meet the complex demands of modern traffic systems. Models such as the Informer [14] and PDFormer [11] address significant challenges, including high computational demands and dynamic traffic conditions, respectively. The Informer efficiently processes long data sequences through a ProbSparse self-attention mechanism, improving both processing time and resource usage, which is beneficial for accurate long-term predictions. Meanwhile, the PDFormer introduces innovative approaches to accommodate propagation delays in dynamic scenarios, enhancing the practicality of forecasting models in ITS.

The adaptability of transformer models is further demonstrated by their capability to manage complex spatial-temporal relationships and integrate multimodal data [15]. The Spatial-Temporal Transformer Networks (STTNs) [16] particularly demonstrate the

ability of transformers to explore the nonlinear and dynamic dependencies typical in traffic data, making them suitable for complex forecasting tasks. Additionally, transformers adeptly incorporate external factors, such as weather conditions and special events, into their predictions. This integration significantly improves model accuracy across various traffic scenarios, as evidenced by models specifically designed to include such external influences [12,13]. Overall, these transformer models have proven their capability to handle complex, multifaceted data, significantly enhancing both performance and efficiency in long-term time series analysis. Their potential within the ITS domain is significant, demonstrating their ability to improve long-term traffic prediction by effectively managing the intricate and varied data typical of these systems.

In advancing travel time prediction within ITS, this study introduces the Periodic Transformer Encoder (PTE). The PTE leverages the strengths of transformer models to overcome the limitations of existing traffic prediction methods, which are primarily designed for short-term forecasting and remain insufficient in handling the intricacies of long-term predictions. The PTE is specifically designed for multi-horizon prediction, adeptly handling both immediate and extended forecasting challenges by effectively utilizing periodic data, an aspect that existing methods have not fully explored. The contributions of this study are detailed as follows:

- The proposed PTE framework represents a significant advancement in travel time prediction. By learning from both short-term and long-term traffic patterns and effectively capturing temporal dependencies across various horizons, the PTE enhances prediction performance for all time scales, consistently outperforming existing methods.
- The PTE introduces a streamlined encoder-only architecture that eliminates the need for a traditional decoder, thereby reducing model complexity and resource requirements. This design simplifies the training and inference processes while reducing computational resources.
- A series of evaluations conducted on a comprehensive real-world traffic dataset demonstrates the superior performance of the PTE. These results show that the PTE outperforms existing methods in predicting travel times across multiple horizons. Notably, the PTE proves its robustness in handling diverse traffic conditions and varying degrees of complexity, affirming its capability to deliver more accurate and reliable travel time predictions.

The rest of this paper is structured as follows. Section 2 provides an overview of related work. Section 3 describes the proposed framework in detail. Section 4 presents the experimental setup, and Section 5 discusses the results and implications of the experiments. Section 6 provides a summary of the study.

## 2. Related Work

### 2.1. Short-Term Traffic Prediction

Early research in traffic prediction primarily utilized traditional machine learning techniques. For instance, support vector regression (SVR) was employed in early works to tackle traffic forecasting challenges [17]. The Autoregressive Integrated Moving Average (ARIMA) model was also used to predict travel times based on historical data, achieving effective one-step-ahead predictions [2]. Additionally, enhancements to conventional models were proposed by the introduction of traffic trend adjustments into the K-nearest neighbors (KNN) model, known as KNN-T [18]. Another approach combined random vector functional link networks with empirical mode decomposition, specifically addressing short-term traffic predictions [19].

In recent years, deep learning has significantly expanded across various fields, enhancing its utilization in traffic forecasting. Long Short-Term Memory (LSTM) networks have demonstrated effectiveness in traffic flow prediction due to their superior capacity for capturing long-term dependencies [4]. Similarly, the deployment of Long Short-Term Memory networks combined with deep neural network layers (LSTM-DNN) has proven effective for

analyzing travel time data on highways [20]. Convolutional neural networks (CNNs) have been adapted to handle spatio-temporal traffic data, with local receptive fields designed to improve predictive performance [21]. Furthermore, the integration of Gated Recurrent Units (GRU) and XGBoost has been employed to extract hidden patterns in traffic data, therefore enhancing the precision of predictions [22]. Significant advancements include the development of the bidirectional spatial-temporal adaptive transformer (Bi-STAT), which features an encoder-decoder architecture. This architecture includes spatial-adaptive and temporal-adaptive transformer components for accurate traffic forecasting [23]. Moreover, to capture the inherent periodicity and continuity in traffic data, models have integrated graph convolutional networks (GCNs) for spatial dependency mapping and transformers for temporal analysis [24]. These approaches highlight the diversity of techniques available for short-term forecasting. However, they typically face challenges when extended to longer durations or iterative multi-step forecasts, which can lead to error accumulation. The challenge of modeling long-term dependencies with deep learning models remains significant [12].

## 2.2. Long-Term Traffic Prediction

Long-term traffic prediction is a critical requirement for various real-world applications, leading to the development of models adept at managing complex datasets. To address both short-term and long-term traffic flow predictions, hybrid forecasting algorithms were introduced [25]. Following this, a DNN model was utilized to improve daily traffic flow predictions in Seattle by leveraging both contextual variables and raw traffic data [26]. Moreover, RNNs are employed to predict long-term traffic flows in urban areas, integrating meteorological and contextual data to refine predictions [27]. Subsequent advancements led to the creation of a hybrid model that merges wavelet decomposition with CNN and LSTM technologies [28]. This model preprocesses traffic data using wavelet technology to effectively extract temporal features. Furthermore, a gradient boosting model, augmented by Fourier filtering to mitigate noise and amalgamate diverse data sources, has been adopted to tackle the challenges of long-term traffic prediction [29]. Additionally, the deep ensemble stacked LSTM model (DE-SLSTM) incorporates weather data to adjust for biases in forecasts that span several hours into the future [7]. These works demonstrate the evolving scope of traffic prediction technologies, illustrating significant advancements in addressing long-term forecasting challenges. Each work has contributed to enhancing predictive accuracy and adapting to the complexities of traffic data in various applications.

Recent advancements have shown the effectiveness of attention-based methods in capturing long-term dependencies. A transformer model with multi-head attention has been introduced for long-term traffic flow forecasting [30]. The multisize patched spatial-temporal transformer network (MSP-STTN) utilizes self-attention for enriched context modeling and cross-attention for global memory learning, aiming at both short- and long-term grid-based crowd flow predictions [12]. Furthermore, the Temporal Fusion Transformer (TFT) merges short-term and long-term temporal patterns and uses various input types to effectively manage prediction horizons ranging from 5 to 150 min [31]. Although these models work to forecast traffic conditions over extended periods, the complexity of traffic dynamics and the need for long-duration predictions present significant challenges. Particularly, error accumulation in multi-step forecasting remains a critical issue. To tackle this, advanced models integrating attention mechanisms within an encoder-decoder architecture have been proposed [32]. This work uses periodic data to enhance forecasting accuracy, particularly for long-term scenarios. By focusing on relevant segments of data through attention mechanisms, these models aim to reduce the common problem of error accumulation observed in multi-step long-term forecasting. However, effectively bridging the gap between short-term and long-term prediction capabilities in traffic management remains an ongoing challenge.

### 3. Proposed Methodology

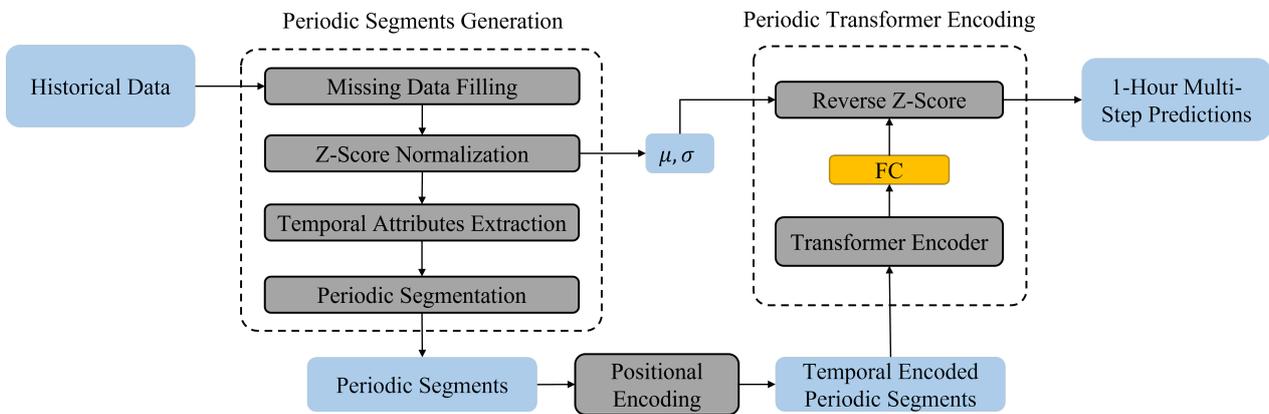
#### 3.1. Problem Formulation

This study proposes the Periodic Transformer Encoder (PTE) as a solution for estimating future travel times from past traffic data along specific road segments. The proposed architecture is illustrated in Figure 1. Fundamental to our approach is the construction of a transformation function  $f(\cdot)$  for each road segment  $r$ . This function systematically processes historical traffic data to generate forecasts of future travel times, represented as:

$$X^{r,\tau} \xrightarrow{f_r^\Delta} Y^{r,\tau,\Delta} \tag{1}$$

where the historical traffic data  $X^{r,\tau} = \{x_{\tau-T+1}^r, \dots, x_\tau^r\}$  comprise observations from the last  $T$  time steps until the current point  $\tau$ , and the predicted future data  $Y^{r,\tau,\Delta} = \{y_{\tau+\Delta+1}^r, \dots, y_{\tau+\Delta+T'}^r\}$  represent the travel times for the next  $T'$  steps beginning  $\Delta$  time steps after  $\tau$ .

To enhance clarity and ease of understanding, the subsequent sections of this work will not include the time and road segment indicators ( $r, \tau$ , and  $\Delta$ ) in the notations for  $X$  and  $Y$ .



**Figure 1.** Overview of the proposed framework. FC: fully connected layers,  $\mu$ : mean,  $\sigma$ : standard deviation of the travel time.

#### 3.2. Data Preprocessing and Segmentation

The initial data preprocessing stages are visualized in Figure 1, highlighting the flow from raw historical data through various transformations necessary for effective model input preparation, following the established procedures of [32]. Initially, missing values in the traffic data, such as travel time and speed, are filled by linear interpolation, ensuring the temporal integrity of the data is maintained, which is crucial for the reliability of subsequent predictions. To standardize the data and mitigate the influence of outliers, each data point is normalized using the Z-score method, where each value is adjusted based on the mean ( $\mu$ ) and standard deviation ( $\sigma$ ) of the dataset. In addition, the traffic dataset is complemented by a variety of temporal attributes, including day types (weekends, weekdays, national holidays), day of the week, months, times of day, and peak periods. Each data instance is transformed into a 316-dimensional vector, where the first two dimensions capture critical traffic metrics—travel time and speed—and the remaining 314 dimensions are dedicated to binary encodings of these temporal attributes, as detailed in Table 1.

**Table 1.** Temporal attributes.

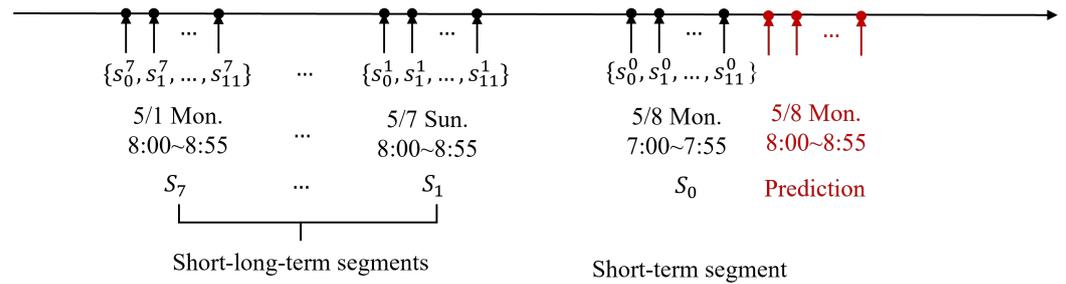
Attribute	Dimension	Description
Day type	3	Includes all calendar days categorized into Weekends, Week-days, and National holidays
Day of the week	7	Organizes days from Monday to Sunday
Month	12	Enumerates months from January to December
Peak hours	4	Identifies four distinct traffic peaks: Morning-peak, Noon-peak, Night-peak, and Off-peak
Time slot	288	Represents the number of 5 min intervals in a day, accounting for each day's full 24 h, totaling 288 time slots

Periodic segment  $S_d$  is extracted from the preprocessed historical data, where  $d$  denotes the day prior to the current time, ranging from  $d = 0$  to  $d = 7$ . The set  $S$  is thus composed of these segments,  $S = \{S_0, S_1, \dots, S_7\}$ , which include one short-term segment and seven corresponding week-long short long-term segments. For instance, if the current time is 8:00 A.M. on May 8, and the goal is to predict travel times for the next hour from 8:00 to 8:55 A.M.,  $S_0$  covers the one-hour period immediately preceding the current time. Specifically,  $S_0 = \{s_0^0, s_1^0, \dots, s_{11}^0\}$  includes traffic data from 7:00 to 7:55 A.M., sampled at 5 min intervals. Conversely, the short long-term segments  $S_1$  to  $S_7$  represent the same one-hour period, each containing twelve time points from 8:00 to 8:55 A.M. Each corresponds to the same hour on the days leading up to the prediction day, stretching back from May 7 to May 1. For a visual explanation of how these segments are structured, refer to Figure 2. These segments,  $\{S_0, S_1, \dots, S_7\}$ , are aggregated into a periodic tensor  $S$  with dimensions  $96 \times 316$ . The dimension 96 represents the total number of one-hour time points across all segments, with each  $S_d$  containing 12 time points capturing hourly traffic data. The second dimension, 316, is a 316-dimensional vector representing various traffic-related metrics and temporal attributes for each time point. The aggregated periodic tensor is embedded with temporal positional information, resulting in a component termed the temporal encoded periodic segment. This step delivers important information to the model about the sequence order, employing Positional Encoding techniques, as described in [33]. Positional encoding is applied to the input embeddings at each position using sine and cosine functions of different frequencies:

$$PE(pos, 2i) = \sin\left(\frac{pos}{10000^{\frac{2i}{d_{\text{model}}}}}\right),$$

$$PE(pos, 2i + 1) = \cos\left(\frac{pos}{10000^{\frac{2i}{d_{\text{model}}}}}\right).$$

Here,  $pos$  represents the position within the sequence, and  $i$  denotes the dimension index. Each position in the input sequence receives a unique positional encoding. The dimension of the model,  $d_{\text{model}}$ , is set at 316, which also corresponds to the dimension of each embedding. The output dimensions remain identical to those of the input. These segments, termed temporal encoded periodic segments, are subsequently fed into the Periodic Transformer Encoder.



**Figure 2.** Visual representation of periodic segments extracted from traffic data, showing both short-term and week-long short long-term segments based on the current time at 8:00 A.M. on May 8, with predictions aimed one hour ahead starting from 9:00 A.M.

### 3.3. Periodic Transformer Encoder

As illustrated in Figure 3, the proposed Periodic Transformer Encoder (PTE) employs a Transformer Encoder to process the aggregated periodic tensor to capture dependencies across various time horizons. This process enhances interactions between different periodic elements, leading to a unified and robust representation. The temporal encoded periodic segments,  $S$ , are input into the Transformer Encoder with dimensions of  $96 \times 316$ , which represents 8 periodic cycles, each comprising 12 segments that correspond to one-hour time points with 316 features each, covering both short-term and long-term dependencies. The multi-head self-attention mechanism of the encoder is particularly effective for capturing the long-range dependencies typical of week-long traffic data. It uses three sets of weights, Query ( $Q$ ), Key ( $K$ ), and Value ( $V$ ), to transform the input tensor:

$$Q = SW_Q, \quad K = SW_K, \quad V = SW_V$$

These transformations project the input tensor into spaces that facilitate the computation of attention scores, calculated as follows:

$$\text{Attention}(Q, K, V) = \text{softmax}\left(\frac{QK^T}{\sqrt{d_k}}\right)V$$

where  $d_k$  is set as the input dimension. This mechanism enables the encoder to weigh and prioritize different segments of the input based on their relevance to the output prediction. Following the attention mechanism, the process continues as described in [33], involving the addition and normalization steps and passing through feed-forward networks (FFNs) within each encoder layer.

Finally, the output from the Transformer Encoder, maintaining the dimensions  $96 \times 316$ ,  $S_{enc}$ , is further processed through fully connected layers to predict travel times:

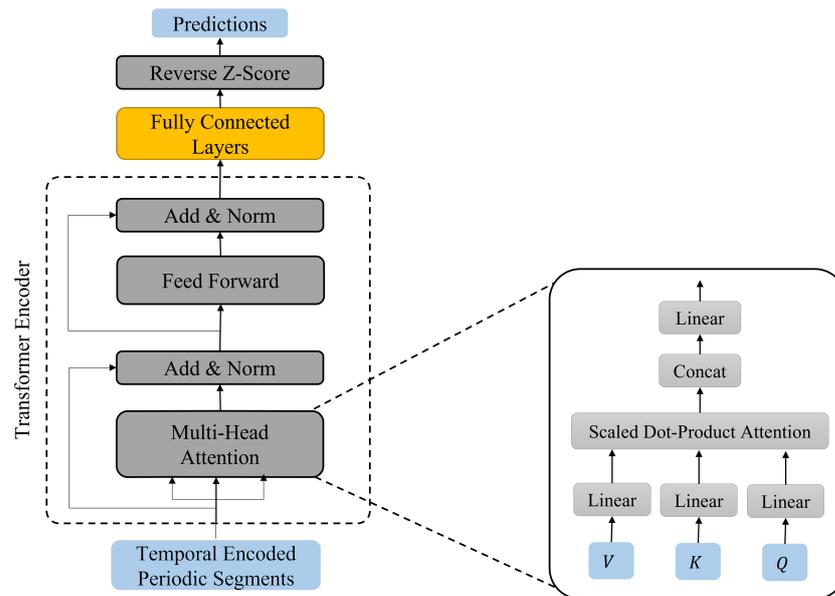
$$Y = \text{ReLU}(S_{enc}W_1 + b_1)W_2 + b_2$$

The dimensions transform from  $96 \times 316$  to  $12 \times 1$ , where each output corresponds to a predicted travel time at 5 min intervals over one hour.

Before generating the final predictions, a reverse Z-score transformation is applied. This step is for converting normalized prediction values back to their original scale. The reverse Z-score transformation is performed using the formula:

$$X = Z \times \sigma + \mu$$

where  $Z$  is the normalized prediction,  $\sigma$  is the standard deviation used during the initial normalization, and  $\mu$  is the mean of the original traffic data.



**Figure 3.** Architecture of the proposed Periodic Transformer Encoder (PTE).

### 4. Experimental Setup

#### 4.1. Datasets

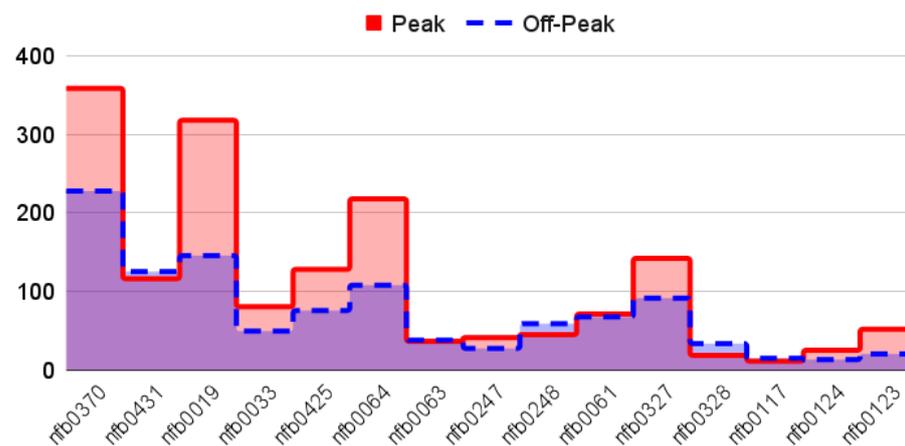
To evaluate the effectiveness of the proposed PTE, comprehensive experiments were conducted using the Taiwan Expressway dataset from the Freeway Bureau of Taiwan, China [34]. This dataset encompasses traffic data for 322 road segments. Following the procedures outlined in [32], we selected the same 15 road segments representing diverse regions of Taiwan—Northern, Central, and Southern. These segments were chosen for their varying travel time statistics, such as mean, standard deviation, and coefficient of variation. This selection aims to capture a broad range of traffic conditions throughout Taiwan, thereby enhancing the representativeness of our dataset. Such diversity is expected to strengthen the validity of our findings across different traffic environments. Details regarding these segments and their corresponding statistical data on travel times are described in Table 2. The dataset covers the period from 1 October 2019 to 31 January 2021. To ensure an equitable experimental design, the data from the first year were used for training, the subsequent month for validation, and the final three months for testing.

**Table 2.** Statistical summary of travel time across selected road segments.

Area	Segment ID	Mean	Standard Deviation	Coefficient of Variation
North	NFB0370	886.61	337.96	0.381
	NFB0431	519.68	128.97	0.248
	NFB0019	357.34	198.74	0.556
	NFB0033	313.64	62.90	0.201
	NFB0425	239.40	71.69	0.299
Central	NFB0064	562.01	92.96	0.165
	NFB0063	554.10	29.72	0.054
	NFB0247	428.68	34.12	0.079
	NFB0248	418.68	37.22	0.089
	NFB0061	402.96	45.75	0.114
South	NFB0327	498.39	61.51	0.123
	NFB0328	489.99	54.32	0.111
	NFB0117	398.29	16.64	0.042
	NFB0124	394.05	15.32	0.039
	NFB0123	392.02	26.46	0.067

#### 4.2. Segmentation of Traffic Conditions

To further illustrate performance under varying traffic densities, experimental results are segmented into two time categories: peak and off-peak. Peak hours, defined as intervals from 7:00 A.M. to 9:00 A.M. and 3:30 P.M. to 7:30 P.M., capture the typical rush hours and are contrasted against less congested off-peak periods. This segmentation helps address the expected greater variability during peak periods due to regular traffic patterns and random events such as accidents, visually supported by Figure 4.



**Figure 4.** The standard deviation of travel times during peak (solid red area) and off-peak (dashed blue line) hours for each road segment, visually depicted using a step line chart.

#### 4.3. Road Segment Complexity Analysis

Empirical evaluation results are categorized based on the complexity levels of road segments, assessing PTE on road segments with High, Moderate, and Low Variability, determined by the standard deviation of travel times. Road segments are classified as High Variability (standard deviation > 100), Moderate Variability (standard deviation between 50 and 100), and Low Variability (standard deviation < 50), indicating varying levels of forecasting challenges. The coefficient of variation (CV) is calculated as the ratio of the standard deviation to the mean travel time and is defined as  $CV = \frac{\sigma}{\mu}$ . CV is used to quantify the variability relative to the mean, enhancing our understanding of each segment's predictability.

#### 4.4. Competitive Methods

To effectively assess the capabilities of the proposed PTE, it is compared against a range of established and contemporary models in the field. The following methods are included in the comparative analysis:

- HA [1] employs historical averages for traffic forecasting by applying straightforward time series regression techniques based on past data.
- LSTM [35] represents a type of recurrent neural network known for its capability to manage sequence-based prediction challenges, including travel time forecasts.
- DNN [26] is a multi-layer deep neural network structured to address various predictive tasks related to traffic, such as estimating travel durations and analyzing flow dynamics.
- DE-SLSTM [7] augments conventional LSTM models by incorporating both short-term and extensive historical traffic data to enhance travel time prediction precision.
- MTSMTT [36] is an approach designed for multivariate time series forecasting, which combines BiLSTM units with attention mechanisms to extract and analyze underlying data intricacies.

- DHM [22] combines Gated Recurrent Units (GRU) with the XGBoost algorithm to forecast freeway travel times and integrate these predictions using linear regression techniques.
- TFT [31] utilizes Temporal Fusion Transformers to effectively merge different types of inputs, demonstrating flexibility in forecasting speeds across varying freeway conditions.
- PASS2S [32] integrates an attention mechanism into a sequence-to-sequence LSTM model, focusing specifically on addressing the complexities of long-term travel time predictions.

To ensure a fair comparison across all competitive methods, each model utilizes both short-term and historical data from one week prior, aligning with the experimental settings described in [32]. Furthermore, consistency in evaluation is maintained by utilizing the test results from [32], thereby aligning our experimental data with the outcomes of their research.

#### 4.5. Parameter Settings and Evaluation Metrics

The configuration of our model's hyperparameters is outlined in Table 3. Additionally, three standard metrics commonly used in travel time forecasting are employed, defined mathematically as follows:

$$\text{MAE} = \frac{1}{N \times l} \sum_{i=1}^N \sum_{j=1}^l |y_{i,j} - \hat{y}_{i,j}|, \quad (2)$$

$$\text{RMSE} = \sqrt{\frac{1}{N \times l} \sum_{i=1}^N \sum_{j=1}^l (y_{i,j} - \hat{y}_{i,j})^2}, \quad (3)$$

$$\text{SMAPE} = \frac{100\%}{N \times l} \sum_{i=1}^N \sum_{j=1}^l \frac{|y_{i,j} - \hat{y}_{i,j}|}{(|y_{i,j}| + |\hat{y}_{i,j}|)/2}. \quad (4)$$

where:

- $y_{i,j}$  denotes the actual travel time recorded for the  $i$ -th sample at time point  $j$ .
- $\hat{y}_{i,j}$  represents the predicted travel time for the  $i$ -th sample at time point  $j$ .
- $N$  is the total count of samples included in the dataset.
- $l$  indicates the total number of time points within the forecast period.

These metrics provide a comprehensive view of the model's performance, accounting for different aspects of forecasting performance.

**Table 3.** Experimental settings.

Setting	Value
Learning rate	0.0001
Batch size	128
Attention heads	4 per encoder block
Number of layers (Encoder, FC)	1, 2
Output dimension	12
Training epochs	15
Loss function	Mean Square Error (MSE)

## 5. Results and Discussions

This section conducts a comparative analysis of the proposed PTE against established competitive methods, focusing on two primary prediction types: short-term (within the next

hour) and long-term (from 1 day to a week ahead). For detailed experimental conditions regarding peak and off-peak hours, please refer to Section 4.2. The impact of road segment variability on the prediction performance is also evaluated to provide insights into PTE's robustness across fluctuating traffic conditions; for more information, see Section 4.3. In the results tables, the top-performing results are highlighted in bold, with the second-best outcomes underscored.

### 5.1. Short-Term Travel Time Prediction

For short-term forecasts throughout the entire day, Table 4 demonstrates PTE's exceptional capability in short-term forecasting, achieving the lowest error rates across all evaluation metrics for predictions up to one hour. PTE significantly outperforms its nearest competitors, with improvements of 10.16% in MAE, 9.91% in RMSE, and 9.16% in SMAPE, reflecting its superior forecasting abilities for imminent travel times. Compared to traditional statistical models like HA, which typically underperform, these results indicate the importance of advanced machine learning techniques in enhancing predictive performance. While LSTM and DNN models are also competent at handling sequence predictions, PTE excels particularly in managing dependencies in historical traffic patterns. Among the evaluated methods, DHM ranks highly, benefiting from the combined effects of GRU and XGBoost algorithms, and excelling in short-term forecasts by effectively capturing traffic dynamics. However, PTE's periodic encoding and direct prediction strategies lead to the most notable performance enhancements.

**Table 4.** Comparison with competitive methods in short-term prediction.

	All			Peak			Off-Peak		
	MAE	RMSE	SMAPE(%)	MAE	RMSE	SMAPE(%)	MAE	RMSE	SMAPE(%)
HA [1]	35.036	75.516	6.408	43.326	87.989	7.076	32.654	70.332	6.161
LSTM [35]	23.626	62.972	4.210	36.152	81.328	5.541	21.001	57.223	3.931
DNN [26]	23.695	57.216	4.389	32.605	68.975	5.342	21.827	53.581	4.189
DE-SLSTM [7]	20.994	51.870	3.934	29.003	62.941	4.861	19.315	48.498	3.740
MTSMFF [36]	26.048	59.649	5.092	32.49	70.536	5.837	23.742	54.207	4.825
DHM [22]	<u>19.591</u>	<u>52.872</u>	<u>3.712</u>	<u>26.035</u>	<u>64.071</u>	<u>4.449</u>	<u>17.284</u>	<u>47.132</u>	<u>3.448</u>
TFT [31]	31.964	70.441	6.118	44.852	89.494	8.029	28.362	62.695	5.620
PASS2S [32]	20.381	<u>50.250</u>	3.860	28.222	<u>61.011</u>	4.778	18.738	<u>46.972</u>	3.668
PTE	<b>17.600</b>	<b>45.264</b>	<b>3.372</b>	<b>22.148</b>	<b>53.835</b>	<b>3.743</b>	<b>16.647</b>	<b>42.530</b>	<b>3.294</b>
Improvement ratio (%)	10.165	9.922	9.160	14.928	11.761	15.868	3.687	9.456	4.464

#### 5.1.1. Performance during Peak and Off-Peak Hours

As shown in Table 4, all evaluated methods face greater challenges during peak hours, as reflected by increased error rates across all evaluation metrics. PTE consistently outperforms all competing models, offering substantial improvements during peak hours, with performance gains of 14.92% in MAE, 11.76% in RMSE, and 15.86% in SMAPE compared to other leading methods. These results not only demonstrate PTE's robustness, maintaining consistent performance despite traffic variability. In off-peak hours, where traffic patterns are generally more stable, PTE maintains the lowest error rates across all metrics, demonstrating its ability to adapt to different traffic densities.

#### 5.1.2. Impact of Road Segment Complexity

In Table 5, PTE demonstrates its robustness by consistently achieving the lowest MAE, RMSE, and SMAPE values across all types of road segments, including High, Moderate, and Low Variability in short-term prediction scenarios. These results highlight PTE's ability to handle immediate traffic conditions effectively, regardless of the road segment's underlying complexity. While error rates increase with segment variability for all models,

PTE excels particularly in higher-variability segments, demonstrating superior performance in managing dynamic and unpredictable traffic patterns.

**Table 5.** Comparison of short-term prediction results on different types of road segments against competitive methods.

Variability	Metric	DE-SLSTM	MTSMFF	DHM	TFT	PASS2S	PTE	Improvement Ratio (%)
High	MAE	38.638	44.152	<u>34.706</u>	57.547	37.125	<b>31.326</b>	9.738
	RMSE	87.794	97.656	88.262	124.223	<u>82.990</u>	<b>73.597</b>	11.318
	SMAPE (%)	5.849	6.939	<u>5.254</u>	8.868	5.697	<b>4.834</b>	7.990
Moderate	MAE	18.308	22.695	<u>17.805</u>	25.550	17.847	<b>15.547</b>	12.524
	RMSE	57.005	62.921	57.428	68.055	<u>55.799</u>	<b>51.370</b>	7.937
	SMAPE (%)	4.776	6.051	<u>4.608</u>	6.912	4.648	<b>4.066</b>	11.753
Low	MAE	<u>8.081</u>	13.198	8.185	14.919	8.118	<b>7.529</b>	6.829
	RMSE	18.510	25.795	20.343	27.214	<u>19.268</u>	<b>17.583</b>	5.006
	SMAPE (%)	1.777	2.912	1.830	3.297	<u>1.806</u>	<b>1.690</b>	4.872

### 5.2. Long-Term Travel Time Prediction

Table 6 displays PTE’s performance across various time segments for long-term predictions. Across the board, PTE shows superior performance, achieving the lowest MAE, RMSE, and SMAPE values, particularly in the comprehensive all-day category. It notably outperforms specialized models like PASS2S and DE-SLSTM, which are designed for long-term dependencies but fall short of PTE’s advanced data handling capabilities.

**Table 6.** Comparison with competitive methods in long-term prediction.

	All			Peak			Off-Peak		
	MAE	RMSE	SMAPE(%)	MAE	RMSE	SMAPE(%)	MAE	RMSE	SMAPE(%)
HA [1]	34.516	74	6.304	43.335	87.007	7.081	32.659	70.134	6.141
LSTM [35]	31.520	74.062	5.743	43.440	89.895	6.958	29.011	69.080	5.488
DNN [26]	31.279	70.295	5.886	41.065	84	6.861	29.219	66.087	5.681
DE-SLSTM [7]	31.948	72.549	5.842	41.68	84.723	6.841	29.899	68.765	5.632
MTSMFF [36]	31.639	69.964	5.91	41.23	83.999	7.026	28.193	63.141	5.509
DHM [22]	33.281	71.896	6.245	43.506	86.560	7.459	29.609	64.803	5.808
TFT [31]	32.316	70.084	6.089	46.439	91.406	7.965	29.023	<u>62.462</u>	5.705
PASS2S [32]	<u>28.929</u>	<u>66.695</u>	<u>5.373</u>	<u>37.596</u>	<u>78.879</u>	<u>6.211</u>	<u>27.105</u>	62.947	<u>5.197</u>
PTE	<b>27.380</b>	<b>65.764</b>	<b>5.063</b>	<b>35.689</b>	<b>78.137</b>	<b>5.783</b>	<b>25.632</b>	<b>62.016</b>	<b>4.912</b>
Improvement ratio (%)	5.353	1.396	5.764	5.071	0.941	6.886	5.435	0.713	5.482

#### 5.2.1. Performance during Peak and Off-Peak Hours

During peak hours, PTE effectively handles increased prediction complexity due to irregular congestion events, as evidenced by its outstanding metrics compared to other models, detailed in Table 6. This ability is vital for dynamic routing and congestion management applications. In less challenging off-peak hours, PTE maintains superior performance, essential for reliable planning and operational decisions, demonstrating its consistent predictive strength across different traffic conditions.

#### 5.2.2. Impact of Road Segment Complexity

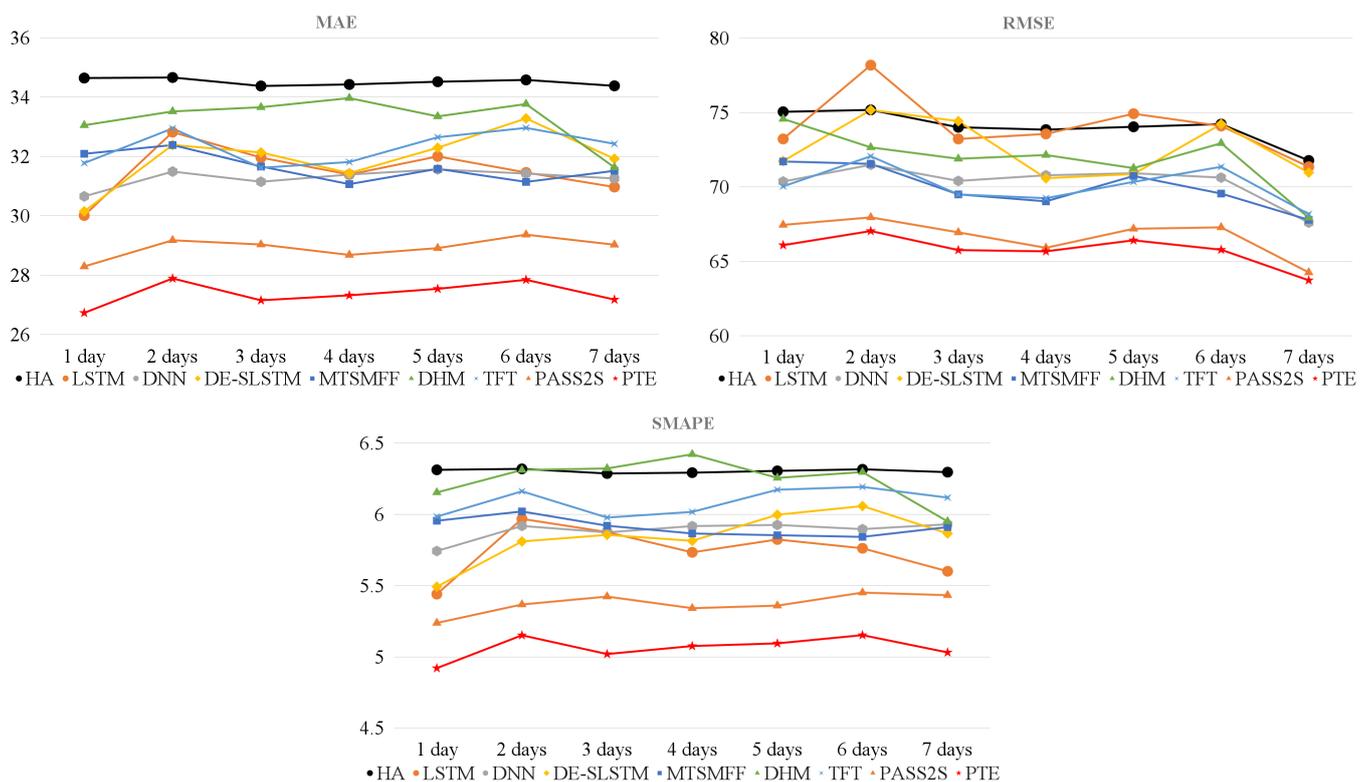
Table 7 demonstrates PTE’s robust performance across road segments of varying complexity. Consistently recording the lowest error rates in MAE, RMSE, and SMAPE, PTE proves highly effective in complex traffic networks, adeptly managing the inherent uncertainties of diverse traffic patterns. This demonstrates its capacity to deliver reliable long-term predictions across a spectrum of traffic scenarios.

**Table 7.** Comparison of long-term prediction results on different types of road segments against competitive methods.

Variability	Metric	DE-SLSTM	MTSMFF	DHM	TFT	PASS2S	PTE	Improvement Ratio (%)
High	MAE	61.200	58.257	61.833	59.128	<u>54.747</u>	<b>51.990</b>	5.035
	RMSE	135.985	126.716	130.079	126.357	<u>121.140</u>	<b>120.380</b>	0.628
	SMAPE (%)	8.970	8.675	9.316	8.880	<u>8.217</u>	<b>7.744</b>	5.757
Moderate	MAE	23.805	24.828	25.607	25.252	<u>22.404</u>	<b>21.569</b>	3.726
	RMSE	63.383	64.113	64.545	63.877	<u>61.597</u>	<b>59.899</b>	2.756
	SMAPE (%)	6.395	6.706	6.985	6.858	<u>5.996</u>	<b>5.754</b>	4.050
Low	MAE	12.999	13.997	14.604	14.683	<u>11.764</u>	<b>10.746</b>	8.650
	RMSE	25.796	26.571	28.311	27.328	<u>24.723</u>	<b>24.161</b>	2.276
	SMAPE (%)	2.867	3.074	3.191	3.250	<u>2.587</u>	<b>2.369</b>	8.429

### 5.2.3. Day-by-Day Performance

Figure 5 details PTE’s ability to deliver accurate long-term forecasts over a seven-day period, showcasing how well it manages daily traffic fluctuations. This aspect is crucial for long-term planning within Intelligent Transportation Systems (ITS), where understanding multi-horizon prediction capabilities is vital.



**Figure 5.** Comparison with competitive methods in long-term prediction over a 7-day horizon.

For the 1-day-ahead prediction, PTE leads with robust performance, effectively utilizing prior data to achieve the lowest MAE and RMSE scores, demonstrating its precise grasp of immediate traffic conditions. From 2-day to 7-day forecasts, PTE maintains lower error rates, highlighting its consistency in adapting to daily changes, including weekday and weekend variations. While models such as DE-SLSTM and PASS2S are designed to handle long-term dependencies, PTE consistently outperforms them and other advanced models such as MTSMFF, DHM, and TFT, illustrating its superior capability in long-term forecasting. PTE’s performance remains notably reliable across the entire week, demonstrating its strong predictive performance and adaptability in varying traffic conditions.

#### 5.2.4. Individual Segment Analysis

The analysis presented in Table 8 details PTE’s MAE performance across selected road segments with varying traffic complexities. This day-by-day examination provides a deeper insight into PTE’s performance trends over a week.

**Table 8.** Performance comparison on MAE for particular road segments.

Segment ID	Method	1 Day	2 Days	3 Days	4 Days	5 Days	6 Days	7 Days
NFB0370	DE-SLSTM [7]	112.711	124.196	116.160	113.254	113.503	126.214	111.231
	MTSMFF [36]	112.253	112.012	104.217	100.242	107.015	103.390	104.751
	DHM [22]	105.368	112.994	118.828	115.879	115.242	118.125	103.395
	TFT [31]	110.994	110.249	108.464	111.004	110.218	110.996	114.244
	PASS2S [32]	<u>100.001</u>	<u>102.632</u>	<u>99.829</u>	<u>101.235</u>	<u>99.102</u>	<u>101.020</u>	<u>99.723</u>
	PTE	<b>94.864</b>	<b>98.443</b>	<b>96.953</b>	<b>94.781</b>	<b>97.978</b>	<b>99.904</b>	<b>98.822</b>
NFB0425	DE-SLSTM [7]	30.622	31.292	32.338	31.655	31.893	33.589	34.009
	MTSMFF [36]	30.998	31.596	32.223	31.638	31.705	31.800	<u>30.499</u>
	DHM [22]	32.830	33.745	34.410	35.336	33.328	33.466	31.666
	TFT [31]	29.216	31.592	30.776	<u>30.884</u>	31.463	31.534	31.898
	PASS2S [32]	<u>28.946</u>	<u>29.879</u>	<u>30.632</u>	30.910	<u>29.713</u>	<u>30.301</u>	30.690
	PTE	<b>27.959</b>	<b>28.784</b>	<b>28.210</b>	<b>29.053</b>	<b>28.475</b>	<b>28.616</b>	<b>28.474</b>
NFB0247	DE-SLSTM [7]	15.791	15.885	16.359	18.616	16.650	18.372	16.187
	MTSMFF [36]	18.159	17.782	17.777	17.807	17.323	17.596	17.639
	DHM [22]	29.039	20.302	17.884	19.488	18.607	18.829	16.430
	TFT [31]	19.083	18.063	20.052	19.229	20.460	18.786	18.373
	PASS2S [32]	<u>13.838</u>	<u>14.505</u>	<u>15.057</u>	<u>15.831</u>	<u>15.213</u>	<u>15.119</u>	<u>15.167</u>
	PTE	<b>13.421</b>	<b>14.341</b>	<b>14.642</b>	<b>14.746</b>	<b>14.750</b>	<b>15.046</b>	<b>15.001</b>

PTE consistently excels on the high-variability segment NFB0370, demonstrating the lowest MAE each day. This demonstrates its adept handling of the unpredictable and complex traffic patterns typical of such roads. Its robustness in these conditions likely stems from the effective use of attention mechanisms that selectively focus on crucial temporal features for the forecast period. Notably, its performance remains stable across the week, highlighting PTE’s ability to manage daily traffic fluctuations effectively.

On the moderately variable road segment NFB0425, PTE also maintains superior performance throughout the week. Although the traffic patterns here are less complex than those of high-variability roads, PTE’s capacity to discern and analyze key temporal dynamics ensures consistent and reliable predictions, which is vital for long-term planning in environments with moderate traffic variability.

For the low-variability segment NFB0247, PTE continues to lead, showing the best performance throughout the week. Traffic on such roads is generally more predictable and stable. Despite slight day-to-day variations in MAE, PTE consistently outperforms other models, proving its efficacy in managing both dynamic and stable traffic conditions.

## 6. Conclusions

This study introduces the Periodic Transformer Encoder (PTE), designed to enhance travel time predictions across multiple horizons within Intelligent Transportation Systems (ITS). By leveraging the strengths of transformer models and focusing on periodic data, the PTE successfully addresses the limitations of existing methods that primarily target short-term forecasting and often struggle with complex long-term prediction scenarios. The PTE’s encoder-only architecture eliminates the need for a traditional decoder, thereby simplifying the model and significantly reducing computational demands. This design not only streamlines the training and inference processes but also enhances the model’s operational efficiency. Empirical evaluations on a comprehensive real-world traffic dataset

have demonstrated that the PTE significantly outperforms existing methods across both short-term and long-term prediction intervals. Its ability to handle complex traffic patterns across diverse road conditions also proves the model's versatility and robustness. The significant improvements demonstrated by the PTE across multiple prediction horizons indicate its potential to substantially advance the field of traffic time prediction by offering more reliable and precise solutions.

Despite these significant advancements, opportunities for further exploration remain. The model's integration of multi-horizon data has yet to fully reveal the distinct impacts of different temporal scales on prediction performance, pointing to a potential area for more focused studies. Future work will delve into these topics more thoroughly by conducting analyses to understand how various types of temporal data influence prediction performance across different horizons. Investigating the interpretability of the attention mechanism is a promising direction for this research. These efforts are expected to lead to further improvements in how we manage and interpret these data interactions.

**Author Contributions:** Conceptualization, H.-T.C.L. and V.S.T.; methodology, H.-T.C.L. and V.S.T.; software, H.-T.C.L.; validation, H.-T.C.L.; formal analysis, H.-T.C.L. and V.S.T.; investigation, H.-T.C.L.; resources, V.S.T.; data curation, H.-T.C.L.; writing—original draft preparation, H.-T.C.L. and V.S.T. writing—review and editing, H.-T.C.L. and V.S.T.; visualization, H.-T.C.L.; supervision, V.S.T.; project administration, V.S.T.; funding acquisition, V.S.T. All authors have read and agreed to the published version of the manuscript.

**Funding:** This research was supported in part by National Science and Technology Council Taiwan under grant nos. 111-2221-E-A49-124-MY3 and 112-2634-F-A49-005.

**Data Availability Statement:** The original contributions presented in the study are included in the article; further inquiries can be directed to the corresponding author.

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

## References

1. Smith, B.L.; Demetsky, M.J. Traffic Flow Forecasting: Comparison of Modeling Approaches. *J. Transp. Eng.* **1997**, *123*, 261–266. [[CrossRef](#)]
2. Billings, D.; Yang, J.S. Application of the ARIMA Models to Urban Roadway Travel Time Prediction—A Case Study. In Proceedings of the 2006 IEEE International Conference on Systems, Man and Cybernetics, Taipei, Taiwan, 8–11 October 2006.
3. Shaygan, M.; Meese, C.; Li, W.; Zhao, X.G.; Nejad, M. Traffic prediction using artificial intelligence: Review of recent advances and emerging opportunities. *Transp. Res. Part C Emerg. Technol.* **2022**, *145*, 103921. [[CrossRef](#)]
4. Tian, Y.; Pan, L. Predicting Short-Term Traffic Flow by Long Short-Term Memory Recurrent Neural Network. In Proceedings of the 2015 IEEE International Conference on Smart City/SocialCom/SustainCom (SmartCity), Chengdu, China, 19–21 October 2015.
5. Duan, Y.; L.V., Y.; Wang, F.Y. Travel time prediction with LSTM neural network. In Proceedings of the 2016 IEEE 19th International Conference on Intelligent Transportation Systems (ITSC), Rio de Janeiro, Brazil, 1–4 November 2016.
6. Wang, D.; Zhang, J.; Cao, W.; Li, J.; Zheng, Y. When Will You Arrive? Estimating Travel Time Based on Deep Neural Networks. In Proceedings of the AAAI Conference on Artificial Intelligence, New Orleans, LA, USA, 2–7 February 2018.
7. Chou, C.H.; Huang, Y.; Huang, C.Y.; Tseng, V.S. Long-term traffic time prediction using deep learning with integration of weather effect. In Proceedings of the 23rd Pacific-Asia Conference on Knowledge Discovery and Data Mining, Macau, China, 14–17 April 2019.
8. Wu, L.; Wang, Y.; Liu, J.; Shan, D. Developing a time-series speed prediction model using Transformer networks for freeway interchange areas. *Comput. Electr. Eng.* **2023**, *110*, 108860. [[CrossRef](#)]
9. Mashurov, V.; Chopurian, V.; Porvatov, V.; Ivanov, A.; Semenova, N. GCT-TTE: Graph Convolutional Transformer for Travel Time Estimation. *arXiv* **2024**, arXiv:2301.07945.
10. Zong, X.; Chen, Z.; Yu, F.; Wei, S. Local-Global Spatial-Temporal Graph Convolutional Network for Traffic Flow Forecasting. *Electronics* **2024**, *13*, 636. [[CrossRef](#)]
11. Jiang, J.; Han, C.; Zhao, W.X.; Wang, J. PDFFormer: Propagation Delay-Aware Dynamic Long-Range Transformer for Traffic Flow Prediction. In Proceedings of the AAAI Conference on Artificial Intelligence, Vancouver, BC, Canada, 20–27 February 2023.
12. Xie, Y.; Niu, J.; Zhang, Y.; Ren, F. Multisize Patched Spatial-Temporal Transformer Network for Short- and Long-Term Crowd Flow Prediction. *IEEE Trans. Intell. Transp. Syst.* **2022**, *23*, 21548–21568. [[CrossRef](#)]
13. Qi, X.; Mei, G.; Tu, J.; Xi, N.; Piccialli, F. A Deep Learning Approach for Long-Term Traffic Flow Prediction With Multifactor Fusion Using Spatiotemporal Graph Convolutional Network. *IEEE Trans. Intell. Transp. Syst.* **2023**, *24*, 8687–8700. [[CrossRef](#)]
14. Zhou, H.; Zhang, S.; Peng, J.; Zhang, S.; Li, J.; Xiong, H.; Zhang, W. Informer: Beyond Efficient Transformer for Long Sequence Time-Series Forecasting. *Proc. AAAI Conf. Artif. Intell.* **2021**, *35*, 11106–11115. [[CrossRef](#)]

15. Zhang, Y.; Liu, S.; Zhang, P.; Li, B. GRU- and Transformer-Based Periodicity Fusion Network for Traffic Forecasting. *Electronics* **2023**, *12*, 4988. [CrossRef]
16. Mingxing, X.; Dai, W.; Liu, C.; Gao, X.; Lin, W.; Qi, G.J.; Xiong, H. Spatial-Temporal Transformer Networks for Traffic Flow Forecasting. *arXiv* **2020**, arXiv:2001.02908.
17. Wu, C.H.; Ho, J.M.; Lee, D. Travel-time prediction with support vector regression. *IEEE Trans. Intell. Transp. Syst.* **2004**, *5*, 276–281. [CrossRef]
18. Qiao, W.; Haghani, A.; Hamed, M. Short-Term Travel Time Prediction Considering the Effects of Weather. *Transp. Res. Rec.* **2012**, *2308*, 61–72. [CrossRef]
19. Li, L.; Qu, X.; Zhang, J.; Li, H.; Ran, B. Travel time prediction for highway network based on the ensemble empirical mode decomposition and random vector functional link network. *Appl. Soft. Comput.* **2018**, *73*, 921–932. [CrossRef]
20. Liu, Y.; Wang, Y.; Yang, X.; Zhang, L. Short-term travel time prediction by deep learning: A comparison of different LSTM-DNN models. In Proceedings of the 2017 IEEE 20th International Conference on Intelligent Transportation Systems (ITSC), Yokohama, Japan, 16–19 October 2017.
21. Ran, X.; Shan, Z.; Shi, Y.; Lin, C. Short-Term Travel Time Prediction: A SpatioTemporal Deep Learning Approach. *Int. J. Inf. Technol. Decis. Mak.* **2019**, *18*, 1087–1111. [CrossRef]
22. Ting, P.Y.; Wada, T.; Chiu, Y.L.; Sun, M.T.; Sakai, K.; Ku, W.S.; Jeng, A.A.K.; Hwu, J.S. Freeway Travel Time Prediction Using Deep Hybrid Model—Taking Sun Yat-Sen Freeway as an Example. *IEEE Trans. Veh. Technol.* **2020**, *69*, 8257–8266. [CrossRef]
23. Chen, C.; Liu, Y.; Chen, L.; Zhang, C. Bidirectional Spatial-Temporal Adaptive Transformer for Urban Traffic Flow Forecasting. *IEEE Trans. Neural Netw. Learn. Syst.* **2022**, *34*, 6913–6925. [CrossRef]
24. Cai, L.; Janowicz, K.; Mai, G.; Yan, B.; Zhu, R. Traffic transformer: Capturing the continuity and periodicity of time series for traffic forecasting. *Trans. GIS* **2020**, *24*, 736–755. [CrossRef]
25. Hou, Z.; Li, X. Repeatability and Similarity of Freeway Traffic Flow and Long-Term Prediction Under Big Data. *IEEE Trans. Intell. Transp. Syst.* **2016**, *17*, 1786–1796. [CrossRef]
26. Qu, L.; Li, W.; Li, W.; Ma, D.; Wang, Y. Daily long-term traffic flow forecasting based on a deep neural network. *Expert Syst. Appl.* **2019**, *121*, 304–312. [CrossRef]
27. Belhadi, A.; Djenouri, Y.; Djenouri, D.; Lin, J. A recurrent neural network for urban long-term traffic flow forecasting. *Appl. Intell.* **2020**, *50*, 3252–3265. [CrossRef]
28. Li, Y.; Chai, S.; Ma, Z.; Wang, G. A Hybrid Deep Learning Framework for Long-Term Traffic Flow Prediction. *IEEE Access* **2021**, *9*, 11264–11271. [CrossRef]
29. Che-Ming Chen, C.C.L.; Chu, C.P. Long-term travel time prediction using gradient boosting. *J. Intell. Transp. Syst.* **2020**, *24*, 109–124. [CrossRef]
30. Reza, S.; Ferreira, M.; Machado, J.; Tavares, J. A Multi-head Attention-based Transformer Model for Traffic Flow Forecasting with a Comparative Analysis to Recurrent Neural Networks. *Expert Syst. Appl.* **2022**, *202*, 117275. [CrossRef]
31. Zhang, H.; Zou, Y.; Yang, X.; Yang, H. A temporal fusion transformer for short-term freeway traffic speed multistep prediction. *Neurocomputing* **2022**, *500*, 329–340. [CrossRef]
32. Huang, Y.; Dai, H.; Tseng, V.S. Periodic Attention-based Stacked Sequence to Sequence framework for long-term travel time prediction. *Knowl. Based Syst.* **2022**, *258*, 109976. [CrossRef]
33. Vaswani, A.; Shazeer, N.; Parmar, N.; Uszkoreit, J.; Jones, L.; Gomez, A.N.; Kaiser, L.; Polosukhin, I. Attention Is All You Need. In Proceedings of the Advances in Neural Information Processing Systems, Long Beach, CA, USA, 4–9 December 2017.
34. Freeway Bureau Taiwan R.O.C. Taiwan Expressway Dataset. Available online: <https://tisvcloud.freeway.gov.tw> (accessed on 10 March 2021).
35. Hochreiter, S.; Schmidhuber, J. Long short-term memory. *Neural Comput.* **1997**, *9*, 1735–1780. [CrossRef]
36. Du, S.; Li, T.; Yang, Y.; Horng, S.J. Multivariate time series forecasting via attention-based encoder–decoder framework. *Neurocomputing* **2020**, *388*, 269–279. [CrossRef]

**Disclaimer/Publisher’s Note:** The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.