Traffic & Transportation

Trend-Season Dual-Branch Fusion Network with Harmonic Weighting for Traffic Prediction

Yangyi LIU[1], Siyuan ZHANG[2], Wenjun ZHOU[3], Yifan WANG[4], Quan ZHANG[5], Bo PENG[6]

Original Scientific Paper
Submitted: 4 Jul 2025
Accepted: 3 Nov 2025
Published: 29 June 2026

[1] liuyangyi_ioe@163.com, Intelligent Policing Key Laboratory of Sichuan Province, Sichuan Police College, Luzhou, China
[2] 202322000565@stu.swpu.edu.cn, School of Computer Science and Soft Engineering, Southwest Petroleum University, Chengdu, China
[3] Corresponding author, zhouwenjun@swpu.edu.cn, School of Computer Science and Soft Engineering, Southwest Petroleum University, Chengdu, China
[4] yifan.wang@swpu.edu.cn, School of Computer Science and Soft Engineering, Southwest Petroleum University, Chengdu, China
[5] zhangquan@swpu.edu.cn, School of Computer Science and Soft Engineering, Southwest Petroleum University, Chengdu, China
[6] bopeng@swpu.edu.cn, School of Computer Science and Soft Engineering, Southwest Petroleum University, Chengdu, China

Publisher:
Faculty of Transport and Traffic Sciences, University of Zagreb

ABSTRACT
Traffic flow prediction is an important component of intelligent transportation systems. Most existing studies adopt spatio-temporal methods for real-time short-term prediction. However, when dealing with complex and dynamically changing historical data in long-term prediction tasks, several challenges arise, such as high complexity, high computational cost and unstable spatial dependencies. In response to this deficiency, this paper proposes a trend-season dual-branch fusion network with harmonic weighting for long-sequence traffic flow prediction. This model decomposes the original sequence into trend and seasonal components. Specifically, the trend branch applies multi-scale sampling, and the trend components at different scales are fused in a top-down manner to enhance information representation. Finally, a linear layer is used to predict the trend results. The seasonal branch encodes the seasonal components through wavelet transform, and a harmonic weighting mechanism is designed to adaptively fuse the prediction results of both branches. Experiments conducted on multiple public transportation datasets demonstrate that the proposed model significantly outperforms existing mainstream methods in long-term predictions at time steps of 96, 192, 288 and 336, thereby verifying the effectiveness and robustness of time-series modelling approach in long-sequence prediction scenarios.

KEYWORDS
multi-scale fusion; long-term prediction; traffic flow; wavelet transform; harmonic weighting.

1. INTRODUCTION

Traffic flow prediction is one of the core tasks of intelligent transportation systems (ITS) [1]. Accurate predictions can provide key insights for transportation planners, assist in designing more resilient and sustainable transportation systems and promote urban liveability and economic growth. Based on different application scenarios, traffic prediction can be divided into two major categories: short-term prediction and long-term prediction.

Among them, short-term prediction (for the next few minutes to several hours) mainly adopts spatio-temporal graph neural networks (GNNs, such as STGCN [2] and DCRNN [3]) or spatio-temporal convolutional networks (STCNN [4]), which utilise the spatial adjacency information of the road network to

capture the traffic interaction between nodes and improve the short-term prediction accuracy. These approaches emphasise real-time and dynamic performance and are widely used in real-time signal control, public transportation dispatching and highway management. Compared to short-term traffic prediction, long-term traffic prediction focuses more on capturing the long-term dynamics and complex patterns of traffic flow to support strategic decision-making, mainly applied to transportation infrastructure planning, urban traffic policy formulation, land-use and urban development planning, and cross-day congestion management and mitigation. Spatio-temporal methods perform well in short-term prediction. However, they encounter two major limitations when applied to long-term forecasting: (1) Traditional GCN-based traffic prediction models require predefined adjacency matrices for graph convolution operations [5]. The fixed adjacency matrix may introduce redundant noise and degrade predictive performance. (2) The complex structure of graph neural networks leads to high computational costs and makes it difficult to handle long-sequence inputs efficiently. Especially on complex and dynamic real-world datasets, the contribution of spatial information is significantly weakened in long-term predictions.

Time-series models are widely applied in traffic prediction, especially in long-term traffic prediction（for the next hours to several days [6]）, for instance, linear regression models and autoregressive integrated moving average models [7–9] are effective for handling stable and linearly varying time series [10]. Machine learning methods like SVR [11–13], KNN [14] and NN [15–16] can effectively capture dynamic traffic flow trends due to their flexibility in handling nonlinear problems. However, they rely on manual feature engineering, limiting their ability to fully exploit the multi-scale periodicity and non-stationarity of long-term traffic flow. To address these challenges, deep learning time-series models have demonstrated superior modelling capabilities. RNNs and their variants (LSTM and GRU) [17–20] networks capture long-term dependencies through recurrent structures, suitable for modelling daily and weekly traffic periodicity, but their accuracy is low for long sequences. DLinear employs trend-seasonal decomposition and linear layers for simple, efficient modelling, effectively capturing periodicity and trends in sequences [21], but its linear predictions struggle with multi-scale, non-linear patterns. TimeMixer employs multi-scale information mixing to enhance the modelling capability of the model [22], but it's difficult to adaptively adjust the strength of periodicity. AdaWaveNet utilises adaptive wavelet transform to effectively decompose seasonal components [23], but it adopts a simple approach for trend components.

In response to the above issues, this paper proposes a dual-branch information fusion model called TSDFNet. Unlike previous long-sequence forecasting models focusing on a single aspect of temporal or frequency-domain analysis, TSDFNet processes the trend and seasonal components of long-term traffic data differently according to their distinct characteristics. Specifically, the trend branch applies multi-scale sampling and performs a top-down, coarse-to-fine fusion of trend information, enabling more accurate modelling of trend components, while the seasonal branch extracts intricate information from seasonal components through adaptive transformations and refined fusion techniques. Finally, predictions from both trend and seasonal branches are combined using adaptive weights. This design allows TSDFNet to more effectively capture the multi-scale, non-stationary and periodic properties inherent in long-term traffic data.

The main contributions of this paper are as follows:

1) We propose a novel trend-season dual-branch fusion network with harmonic weighting for long-term traffic prediction. TSDFNet introduces a unified architecture that explicitly models and integrates the dynamics of trend and seasonal components, enabling more interpretable and stable long-horizon predictions.
2) We design a hierarchical multi-scale fusion mechanism within the trend branch to progressively aggregate coarse-to-fine information, enhancing the modelling of long-term dependencies and slow-varying dynamics. In contrast, the seasonal branch employs adaptive wavelet transformation to effectively capture high-frequency periodic variations. Finally, a harmonic weighting mechanism dynamically balances the contributions of both branches in the frequency domain, providing adaptive integration rather than fixed combination.
3) Extensive experiments on the PEMS04, PEMS08, LargeST-SD and Glasgow datasets demonstrate the superior performance of TSDFNet across various prediction horizons (96, 192, 288 and 336), thereby confirming its practical effectiveness and applicability to real-world long-term traffic forecasting scenarios.

2. RELATED WORK

Traffic flow forecasting has been extensively studied, depending on different methods, it is commonly categorised into two categories: spatio-temporal models and time-series models. Spatio-temporal models typically target short-term forecasting tasks, while time-series models are more suitable for long-term prediction scenarios.

Spatio-temporal models leverage both spatial dependencies among road network nodes and temporal dynamics to improve forecasting accuracy. Representative models, for instance, Yu et al. [2] proposed STGCN, a spatio-temporal graph convolutional network that integrates spatial graph convolutions with temporal convolutions in a fully convolutional framework, enabling efficient parallel computation and alleviating gradient-vanishing issues. Li et al. [3] developed DCRNN, which employs diffusion convolution to capture spatial dependencies and combines recurrent neural networks to model temporal dynamics, effectively learning complex spatial-temporal correlations. More recently, Jiang et al. [24] proposed PDFormer, which explicitly incorporates propagation delay into the Transformer architecture to model long-range temporal dependencies more accurately. Luo et al. (2024) introduced LSTTN, a long-short term transformer-based spatio-temporal network that integrates long- and short-term temporal representations within a unified framework to enhance forecasting robustness (available as a preprint on arXiv:2403.16495). Similarly, Xiao and Long [25] proposed a multi-channel spatio-temporal Transformer, which integrates multiple feature channels and graph learning mechanisms to better capture complex urban mobility patterns. These spatio-temporal architectures achieved strong performance in short-term traffic prediction. However, their limited temporal modelling capacity and reliance on static graph structures reduce their suitability for long-term traffic prediction due to its increased fluctuations and uncertainty.

In contrast, time-series models focus solely on temporal dependencies derived from historical traffic data, without relying on spatial relationships. Kumar and Vanajakshi [7] applied ARIMA and SARIMA models for short-term traffic forecasting, demonstrating their effectiveness for stationary and linear time series. However, such statistical approaches often struggle with nonlinear and non-stationary traffic patterns. To address these limitations, machine learning methods have been explored. Wang et al. [13] optimised support vector regression (SVR) using Bayesian techniques to improve short-term traffic prediction accuracy, while Evans et al. [26] employed a random forest (RF) algorithm that considers contextual traffic features. Tang et al. [15] enhanced neural network-based models by incorporating fuzzy logic to capture the periodic characteristics of traffic speed. Despite their flexibility, these traditional machine learning models rely heavily on handcrafted features and lack the ability to model complex long-range dependencies.

With the acquisition of massive traffic data, deep learning methods, particularly time-series models, have been able to effectively mine useful information from vast historical datasets and uncover latent patterns. For instance, LSTM and GRU networks capture long-term dependencies through recurrent structures. Some variants, such as that proposed by Wang et al. [27], integrated encoder–decoder architectures and attention mechanisms to better capture long-range dependencies in traffic flow. Li et al. [28] developed a hybrid CNN–LSTM model that combines convolutional feature extraction with recurrent modelling to predict traffic flow over 24-hour horizons. As research deepens, scholars have developed more novel time-series models. Wu et al. [29] proposed Autoformer, a Transformer-based forecasting model that leverages series decomposition and auto-correlation mechanisms to improve long-term prediction accuracy. Zeng et al. [21] presented DLinear, which applies trend-seasonal decomposition with linear layers for efficient modelling while maintaining strong long-horizon performance. Wang et al. [22] designed TimeMixer, which employs multi-scale information mixing and past decomposable mixing (PDM) modules to enhance the modelling of complex temporal dynamics. Yu et al. [23] developed AdaWaveNet, which introduces adaptive wavelet transforms to better handle non-stationary time series and capture multi-scale seasonal variations. Ye et al. (2024) proposed ATFNet, an adaptive time-frequency ensembled network that integrates time- and frequency-domain representations to improve robustness in long-term forecasting (available as a preprint on arXiv:2404.05192).

Recently, several studies have extended these purely temporal forecasting frameworks to the traffic prediction domain. Chan et al. [30] proposed the cluster-augmented LSTM (CAL) model, which enhances long-horizon traffic speed forecasting through segmented time-frame clustering and data augmentation, thereby improving feature diversity without using spatial structures. Wang et al. [31] developed the hierarchical temporal aggregation recurrent residual network (HTARRN), which aggregates multi-resolution temporal features and employs residual recurrent modelling to capture both fine-grained and coarse-grained dependencies, achieving superior long-horizon prediction accuracy. These domain-specific temporal models

demonstrate that purely time-series architectures can achieve competitive performance in long-term traffic forecasting, complementing the progress made by general-purpose long-sequence forecasting frameworks.

Although these studies have significantly advanced long-term time-series forecasting, traffic flow remains highly nonlinear and exhibits multi-scale periodicity and dynamic uncertainty, posing persistent challenges for existing models. Therefore, we propose TSDFNet to effectively capture these complex patterns in long-term traffic flow forecasting.

3. METHODOLOGY

3.1 Overall architecture

We define the long-term traffic prediction problem as predicting future data from historical input data, where input $X \in \mathbb{R}^{T\times N\times C}$ and T represents the number of time steps, N represents the number of nodes, C represents the number of features. Since this work only uses traffic flow as feature, $C = 1, X \in \mathbb{R}^{N\times T}$.

TSDFNet is a dual-branch information fusion framework designed for capturing complex shifts and anomalies in long-term traffic data. The model consists of three main components: a multi-scale trend modelling branch, an adaptive seasonal modelling branch and a dynamic trend-seasonal fusion weight module. *Figure 1* shows the overview of TSDFNet.

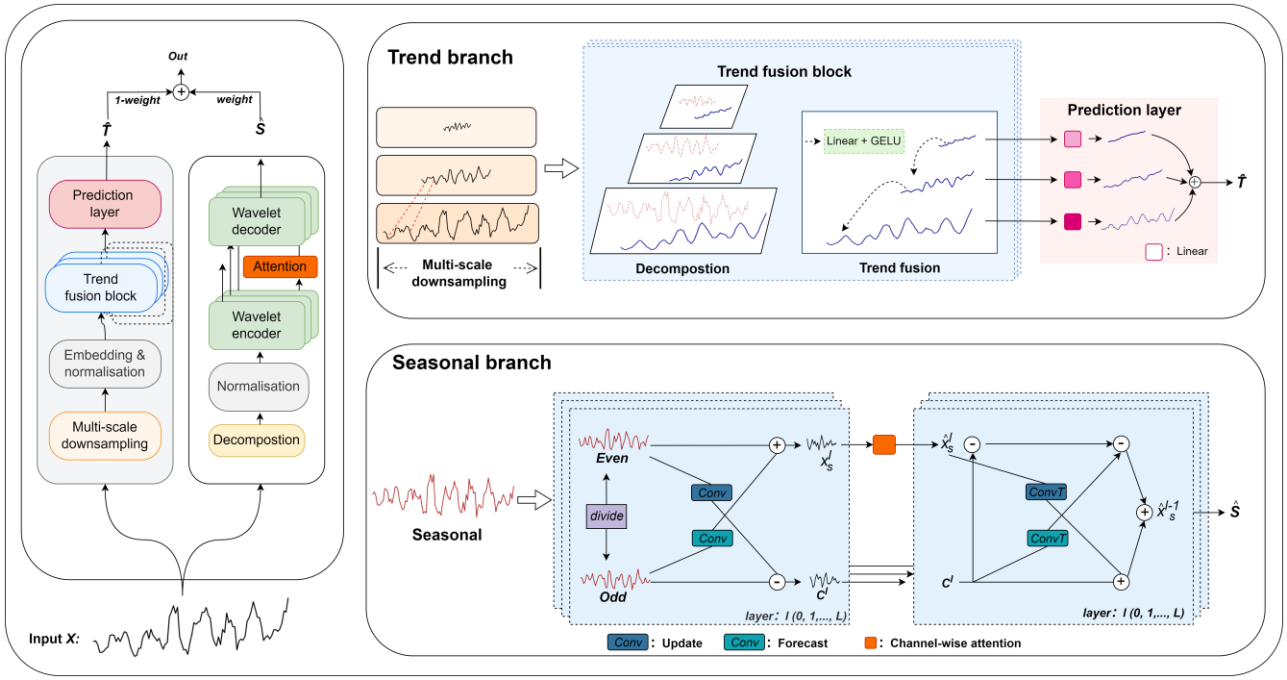

Figure 1 – Overview of TSDFNet

Given an input traffic time series X, the trend branch employs multi-scale downsampling to extract complete representations at different temporal scales. It then decomposes multi-scale sequences to obtain trends at different scales and fuses them from coarse to fine scale and make the final prediction. This design enables more accurate modelling of complex trend behaviours across multiple temporal scales. The adaptive seasonal modelling branch employs an adaptive wavelet-based decomposition to extract fine-grained seasonal patterns and effectively handle non-stationary periodic components in traffic data.

Finally, the dynamic trend-seasonal fusion weight module adaptively combines the outputs of the two branches. It computes weights based on the dominant frequency characteristics of the input sequence, allowing the model to dynamically adjust the relative contributions of trend and seasonal predictions according to the underlying temporal patterns. This architecture enables TSDFNet to flexibly model the complex, nonlinear and multi-scale characteristics inherent in long-term traffic flow data.

3.2 Multi-scale trend modelling branch

In traffic time series, the trend component reflects the overall directional movement of traffic flow (increase or decrease). However, the trend component at the original scale still contains high-frequency noise or local anomalies. Due to limitations of simple decomposition methods, these fluctuations can interfere with the identification of the true underlying trend. The trend components obtained through downsampling at different scales reflect varying levels of abstraction, with coarser scales better representing the overall trends, as shown in *Figure 2*.

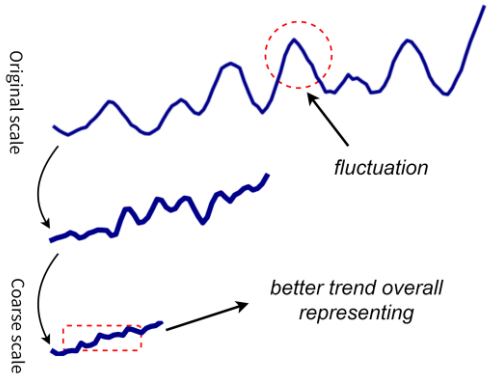

Figure 2 – Multi-scale trend component examples

Therefore, the top-down coarse-to-fine modelling strategy helps more accurately capture the underlying trend direction.

To be specific, the input sequence $X \in \mathbb{R}^{N \times T}$ is first downsampled into M scales using a moving average method, denoted as $X_M = \{x_1, x_2, \ldots, x_M\}$, where $x_m \in \mathbb{R}^{\lfloor \frac{T}{2^m} \rfloor \times N}$, $m \in \{1,2,\ldots,M\}$. Then embedding and normalisation operations are applied to X_M. Both processes are shown in *Equations 1* and *2*. The embedding operation transforms the raw features into higher-dimensional representations. Normalisation operation accelerates training convergence and improves model stability, both X_{embed} and $X_{normalised}$ consist of M scales.

$$X_{embed} = Embed(X_M) \tag{1}$$

$$X_{normalised} = normalise(X_{embed}) \tag{2}$$

Then $X_{normalised}$ is fed into the stacked trend fusion blocks (with L layers). A single trend fusion block is first applied to the sequence to perform seasonal-trend decomposition, obtaining different scale trend components $T_M^l = \{t_0^l, t_1^l, \ldots, t_M^l\}$, where t_0^l is the original scale and t_M^l is the coarsest scale, $l \in \{1,2,\ldots,L\}$. This process can be formalised as *Equation 3*. The seasonal components S_M^l are used to reconstruct the complete sequence at a specific scale within the l-th layer.

$$T_M^l,\ S_M^l = SeriesDecomp(x_m), x_m \in X_{normalised} \tag{3}$$

The coarsest trend scale focuses the least on detailed variations and emphasises overall information, while the original trend scale retains local fluctuations. Based on this observation, a top-down approach is adopted to progressively guide the modelling of finer-scale trend components, as shown in *Equation 4.* Where $Top - Down - Fusion(\cdot)$ is two linear layers equipped with $GELU$ activation functions.

$$t_m^l \leftarrow t_m^l + Top - Down - Fusion\left(t_{m+1}^l\right), for\ m = M-1, M-2, \ldots, 0 \tag{4}$$

The input dimension is $\lfloor \frac{T}{2^{m+1}} \rfloor$, the output dimension is $\lfloor \frac{T}{2^m} \rfloor$ as shown in *Figure 3*. After information fusion, the trend components are added to the corresponding seasonal components at each scale to reconstruct complete sequences, which are fed into the next fusion block. The final fusion block outputs the trend components T_M^L at all M scales.

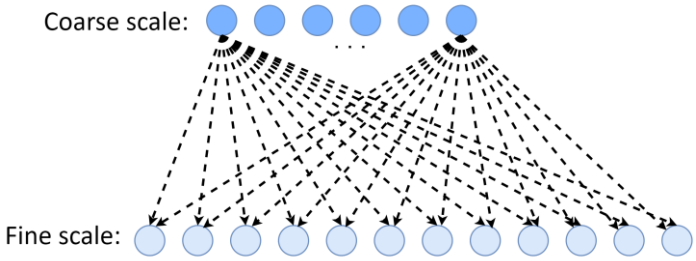

Figure 3 – Trend top-down-fusion

Then, for each scale m, a linear layer with a corresponding length is applied to forecast the trend at that scale. The final output of the trend branch is obtained by aggregating the predictions from all M scales, and the process can be formalised as *Equation 5*.

$$\hat{T}_m = \text{Linear}_m(T_m^L), m \in \{1,2,\dots,M\}, \hat{T} = \sum_{m=1}^{M} \hat{T}_m \tag{5}$$

In summary, this hierarchical coarse-to-fine design enables the trend branch to effectively suppress local fluctuations, capture stable long-term dynamics at multiple temporal resolutions and enhance both robustness and interpretability of trend prediction in long-horizon traffic forecasting scenarios.

3.3 Adaptive seasonal modelling branch

The seasonal component typically exhibits periodic or quasi-periodic fluctuations driven by daily, weekly or monthly commuting patterns. Although basic decomposition methods can isolate the seasonal component, the residual seasonal signal often remains noisy, non-stationary or structurally unstable. Therefore, a more refined modelling process is required to extract informative and temporally adaptive seasonal features.

The wavelet transform is a classical time-domain technique that supports multi-resolution analysis. Unlike global transforms such as the Fourier transform, the wavelet transform can simultaneously capture both temporal and scale information. In particular, the lifting scheme [32] reformulates traditional wavelet decomposition into a modular "split–predict–update" process, enabling more efficient and flexible computation within neural architectures.

Following this principle, the seasonal branch incorporates a learnable lifting-based structure, in which the seasonal component is recursively divided and reconstructed across multiple layers. Specifically, the seasonal component $x_s^{l-1}(l \in 0,\dots,L, x_s^0 = x_s)$ is divided into even- and odd-indexed parts according to their positions. The calculation is shown in *Equation 6*, here, $[::2]$ denotes selecting every second element starting from index 0 (even indices), while $[1::2]$ denotes selecting every second element starting from index 1 (odd indices).

$$e^l = x_s^{l-1}[::2], o^l = x_s^{l-1}[1::2], \forall i \in N \tag{6}$$

During the prediction step, the even-indexed components are used to estimate the odd-indexed components. This relies on the observation that many real-world signals exhibit smooth local transitions, allowing the odd-indexed elements to be inferred from their even-indexed neighbours. By predicting the odd-indexed components from a smoothed representation (even indices), the model captures local variations, and the difference between the actual and predicted values forms the detail coefficients c^l, as shown in *Equation 7*. Here, W_p^l are the convolutional filter weights, and b_p^l denotes the bias at level l.

$$c^l \leftarrow o^l - Conv_{W_p^l, b_p^l}(e^l) \tag{7}$$

Next, the update step leverages the detail coefficients to enhance the even-indexed components, as defined in *Equation 8*. Similarly, W_u^l are the convolutional filter weights, and b_u^l is the biases at level l. These steps progressively refine the signal representation by capturing both global trends and fine-grained details.

$$e^{l'} \leftarrow e^{l} + \mathrm{Conv}_{\mathrm{W_u^l},\mathrm{b_u^l}}(c^l) \tag{8}$$

The approximation of x_s at level l is obtained by combining the even components with the corresponding detail coefficients, as shown in *Equation 9*. The resulting x_s^l is fed to the next level.

$$x_s^l \leftarrow e^{l'} + c^l \tag{9}$$

To further enhance the seasonal component x_s^l, the seasonal branch applies channel-wise attention to the output x_s^N of the final wavelet decomposition layer, where N denotes the final level of decomposition. This mechanism refines the approximation across channels while keeping the detail coefficients c^N unchanged. As a result, it yields the final approximation $\widehat{x_s}^N$ of the seasonal component, thereby capturing key patterns and improving representation without increasing the sequence length.

To fully restore the seasonal signal at the original resolution, a layer-by-layer inverse reconstruction is performed. Specifically, to recover the original seasonal component $\widehat{x_s}$ from its refined form $\widehat{x_s}^l$, which is obtained after generating the final approximation $\widehat{x_s}^N$, the seasonal branch adopts an inverse reconstruction process based on convolutional transpose networks. This process leverages transposed convolutional layers to iteratively upsample the seasonal representation and integrate the corresponding detail coefficients at each level.

Finally, the inverse of the combining step is implemented by subtracting the detail coefficients from the seasonal component at the current level l in an element-wise manner, as shown in *Equation 10*.

$$e^{l'} \leftarrow \widehat{x_s}^l - c^l \tag{10}$$

To reconstruct the original even components and refine the decomposed representations, transposed convolutional layers are applied in the reverse direction of the forward process, as defined in *Equation 11*. The reconstruction of the original odd components is then achieved by adding the predicted detail coefficients to their corresponding positions, as shown in *Equation 12*.

$$e^{l} \leftarrow e^{l'} - C\mathrm{onvT}_{\mathrm{W_u^{l,T}},\mathrm{b_u^{l,T}}}(c^l) \tag{11}$$

$$o^{l} \leftarrow c^{l} + C\mathrm{onvT}_{\mathrm{W_p^{l,T}},\mathrm{b_p^{l,T}}}(e^l) \tag{12}$$

The matrices $\mathrm{W_u^{l,T}}$ and $\mathrm{W_p^{l,T}}$, denote the transposed convolutional weights for the inverse update and prediction steps, respectively, while $\mathrm{b_u^{l,T}}$ and $\mathrm{b_p^{l,T}}$ represent the corresponding bias terms. At the final stage, interleaving the odd- and even-indexed components yields the reconstructed seasonal sequence at level $l - 1$. The seasonal output is denoted as $\hat{S}$, and the reconstruction process can be formalised as *Equation 13*.

$$\widehat{x_s}^{l-1}[2i] = e^l[i], \quad \widehat{x_s}^{l-1}[2i+1] = o^l[i], \; \hat{S} = \widehat{x_s}^0 \tag{13}$$

In summary, the adaptive seasonal modelling branch leverages a learnable lifting-based decomposition and hierarchical reconstruction strategy to effectively capture dynamic periodic structures and localized variations, thereby improving the model's capability to represent complex seasonal dynamics in long-term traffic forecasting.

3.4 Dynamic trend-seasonal fusion

In long-term traffic forecasting, traffic sequences often exhibit varying degrees of periodicity. For instance, workday traffic exhibits strong daily and weekly cycles, whereas off-peak or irregular patterns tend to be less periodic and more trend-driven. Rigidly combining the predictions from trend and seasonal branches without considering the actual periodic strength may lead to suboptimal results.

To adaptively combine the outputs of the trend and seasonal branches, a frequency-aware weighting mechanism based on the periodicity of the input traffic sequence is introduced to this work. To implement this, a discrete Fourier transform (DFT) is first applied to estimate the spectral energy distribution, as illustrated in

Equation 14, inspired by the dominant harmonic energy principle, where $F[k]$ is the complex spectral value of the $k-th$ frequency component, $e^{-2\pi i \cdot kn/L}$ is the basis function of the DFT, representing a sinusoid with frequency k/L in complex exponential form.

$$F[k] = \sum_{n=0}^{L-1} x[n] \cdot e^{-2\pi i \cdot kn/L}, \quad k = 0,1,\dots,L-1 \tag{14}$$

Subsequently, the total spectral energy is computed (typically using only half of the frequency components due to the conjugate symmetry property of the DFT). The calculation is shown in *Equation 15*. $|F[k]|^2$ is the energy of the frequency component.

$$E_f = \sum_{k=0}^{\lfloor L/2 \rfloor} |F[k]|^2 \tag{15}$$

Then, the fundamental frequency is identified, and the energy of the dominant harmonic group (i.e. the first N harmonics) is calculated, as shown in *Equation 16*. Here, k_0 denotes the index of the fundamental frequency, $f_0 = \frac{k_0}{L}$ represents the corresponding frequency, E_h indicates the total energy of the dominant harmonic group (i.e. the harmonic energy), and N is the predefined number of harmonics.

$$k_0 = \arg\max_k |F[k]|, \quad k \in [1, \lfloor L/2 \rfloor], E_h = \sum_{n=1}^{N} |F[n \cdot k_0]|^2 \tag{16}$$

Finally, the frequency-domain module weight is computed using *Equation 17*, which represents the contribution of the seasonal component, while the trend component weight is defined as $1 - w_f$.

$$w_f = \frac{E_h}{E_f} \tag{17}$$

The final output is calculated as *Equation 18*.

$$out = \hat{T} \times (1 - w_f) + \hat{S} \times w_f \tag{18}$$

4. EXPERIMENT

4.1 Dataset

To comprehensively evaluate the effectiveness and generalisability of the proposed model in long-term traffic forecasting, four widely used benchmark datasets are employed: PEMS04 [33], PEMS08 [33], LargeST-SD [34] and the Glasgow urban traffic flow dataset [35]. These datasets differ in scale, region and data characteristics, covering both freeway and urban road networks with varying degrees of periodicity and fluctuation.

PEMS04 and PEMS08 are real-world freeway traffic datasets collected by the caltrans performance measurement system (PeMS). They provide fine-grained vehicle flow measurements at high temporal resolution, making them suitable for analysing temporal patterns such as daily rush hours and weekly commuting cycles. Since both datasets have already been pre-processed, no additional data-cleaning procedures were required in this study.

LargeST-SD is a subset of the LargeST benchmark dataset, a large-scale traffic forecasting dataset featuring a high number of sensors, long temporal coverage and comprehensive metadata. The full LargeST dataset includes traffic flow data from 8,600 sensors across California, collected over a five-year period at 5-minute intervals. Each sensor is annotated with attributes such as geographic coordinates, district and highway type, enabling rich spatio-temporal and contextual analysis. The SD subset specifically covers 716 mainline sensors in San Diego County. Data from 2019 was resampled to 30-minute intervals by aggregating every six 5-minute readings. Sensors with more than 30% missing values or exhibiting abnormal temporal stationarity were

removed, leaving 681 valid sensors. Missing values were imputed using linear interpolation, producing a consistent dataset for long-term forecasting tasks.

Glasgow urban traffic flow dataset contains hourly traffic flow records collected through the SCOOT-based urban traffic control system operated by the Glasgow City Council. It includes 470 sensors distributed across major and local roads within the city. Data from 1 October 2019 to 30 September 2020 (8,760 hourly time steps) were used. Sensors outside the city boundary or with excessive missing or zero-flow rates were excluded, and missing values were linearly interpolated. This dataset captures fine-grained urban traffic dynamics and enables evaluation of model adaptability to complex intra-city patterns distinct from freeway data.

The information of the four datasets is summarised in *Table 1*. All datasets are partitioned into training, validation and test sets using a 6 : 2 : 2 ratio.

Table 1 – Summary of datasets used in experiments

Datasets	Nodes	Time steps	Time Range
PEMS04	307	16992	01/2018 – 02/2018
PEMS08	170	17856	07/2016 – 08/2016
LargeST-SD	681	17520	01/2019 – 12/2019
Glasgow	470	8760	10/2019 – 09/2020

4.2 Evaluation metrics

To comprehensively assess predictive performance, three widely used evaluation metrics are adopted: mean absolute error (MAE), root mean squared error (RMSE) and mean absolute percentage error (MAPE). Specifically, MAE measures the average magnitude of prediction errors, reflecting the overall prediction accuracy. RMSE assigns higher penalties on larger errors, thus indicating model stability and sensitivity to outliers. MAPE quantifies prediction accuracy as a percentage, facilitating intuitive interpretation. Therefore, lower values of these metrics indicate superior predictive performance and higher model reliability.

$$\mathrm{MAE} = \frac{1}{N}\sum_{i=1}^{N}|y_i - \hat{y}_i| \tag{19}$$

$$\mathrm{RMSE} = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(y_i - \hat{y}_i)^2} \tag{20}$$

$$\mathrm{MAPE} = \frac{1}{N}\sum_{i=1}^{N}\left|\frac{y_i - \hat{y}_i}{y_i}\right| \times 100\% \tag{21}$$

where N denotes the total number of observations, y_i and $\hat{y}_i$ represent the ground truth and predicted value, respectively.

4.3 Experimental settings and baseline models

Implementation details

All experiments were performed on a machine running Ubuntu 22.04.4 LTS, using Python 3.9.7 and PyTorch 2.1.1. The hardware environment includes a single NVIDIA GeForce RTX 3070 GPU with 8 GB memory. The main hyperparameters of TSDFNet are summarised in *Table 2*. These include model dimensions, trend and seasonal branch configurations, and the parameters of the frequency-aware fusion module.

Table 2 – Hyperparameters of TSDFNet

Parameter	Value	Description
d_model	256	Hidden embedding size
d_ff	512	Intermediate expansion size
Downsampling layers	3	Number of multi-scale decomposition levels in the trend branch
Downsampling window	2	Reduction factor for temporal downsampling
Lifting levels	4	Hierarchical decomposition depth in the seasonal branch
Kernel size	7	Convolutional kernel size for lifting operations
Number of harmonics	3	Dominant harmonic group count in frequency-aware module
Optimiser	Adam	Optimiser used for model training
Learning rate	0.003	Initial learning rate for TSDFNet
Batch size	16 / 32	Selected according to dataset scale
Epochs	20	Early stopping strategy applied

Baseline models were trained with their official default configurations as provided in the corresponding open-source implementations, remaining consistent with their original designs. To maintain a unified training protocol across all experiments, each model was trained for 20 epochs with an early stopping strategy (patience = 10), and the input and prediction lengths were set equal, i.e. input = output $\in \{96, 192, 288, 336\}$.

Baseline models

To comprehensively evaluate the effectiveness of the proposed model, a diverse set of baseline methods is selected, including both traditional and state-of-the-art deep learning models:

— SVR [13] is a classical regression method that uses kernel functions to capture nonlinear relationships and has been widely applied in early traffic forecasting studies.
— LSTM leverages recurrent units with memory cells to model long-range temporal dependencies in time series data.
— DLinear [21] is a decomposition-based model that directly models trend and seasonal components with simple linear projections, offering high efficiency and competitive performance.
— Autoformer [29] introduces an auto-correlation mechanism to capture long-term dependencies and a decomposition architecture to model series-level trends and periodic patterns.
— Pyraformer [36] introduces a pyramidal attention structure for efficient long-term sequence modelling, reducing computational cost while preserving global context.
— iTransformer [37] employs instance-level decomposition and dynamic receptive field learning to improve adaptation to diverse temporal patterns.
— CycleNet [38] leverages a cycle-aware hierarchical representation to explicitly capture multi-scale periodic patterns.

These baseline models collectively cover a wide spectrum of modelling strategies – from traditional statistical learning to modern Transformer-based architectures, ensuring a comprehensive comparison for long-term traffic forecasting tasks.

4.4 Results

To assess the effectiveness and robustness of the proposed method, this section reports quantitative results on four benchmark datasets under various long-term forecasting horizons. All methods are evaluated under consistent experimental settings using MAE, RMSE and MAPE as the performance metrics. The prediction lengths are set to 96, 192, 288 and 336, aligned with common long-term forecasting benchmarks.

Results on freeway traffic datasets

Table 3 – Prediction performance of different methods on PEMS04 dataset

Time	Metric	Value							
		SVR	LSTM	Autoformer	Pyraformer	DLinear	iTransformer	CycleNet	TSDFNet
96	MAE	43.64	31.60	109.22	28.27	55.70	32.59	34.18	**27.72**
	RMSE	66.64	47.26	146.48	42.53	78.62	49.89	53.52	**41.56**
	MAPE(%)	38.75	26.62	125.46	24.61	69.98	**23.40**	26.27	25.36
192	MAE	39.14	34.34	93.86	30.23	42.82	30.51	34.17	**29.04**
	RMSE	61.69	50.51	123.19	45.00	64.02	49.98	53.30	**43.91**
	MAPE(%)	32.98	28.70	138.63	26.39	46.36	**22.18**	26.04	27.17
288	MAE	37.44	35.99	101.20	29.18	37.22	30.53	33.67	**27.98**
	RMSE	59.29	53.55	127.69	45.33	57.59	49.34	52.49	**42.40**
	MAPE(%)	30.66	29.65	152.51	25.11	35.00	**22.14**	25.69	25.93
336	MAE	38.82	38.18	103.72	28.74	38.32	31.90	33.63	**28.45**
	RMSE	61.10	56.62	132.31	43.99	59.19	52.12	52.33	**43.19**
	MAPE(%)	32.31	30.30	153.02	24.94	35.79	**22.34**	25.66	26.99

Table 4 – Prediction performance of different methods on PEMS08 dataset

Time	Metric	Value							
		SVR	LSTM	Autoformer	Pyraformer	DLinear	iTransformer	CycleNet	TSDFNet
96	MAE	43.64	31.08	44.71	26.89	48.81	40.87	31.19	**25.84**
	RMSE	66.64	47.59	62.15	39.75	68.44	59.78	47.59	**38.13**
	MAPE(%)	38.75	23.14	34.38	20.63	43.03	27.99	21.41	**19.68**
192	MAE	36.99	32.22	82.99	27.66	38.08	48.19	32.13	**26.75**
	RMSE	55.96	48.27	109.29	42.46	57.53	67.88	48.58	**40.20**
	MAPE(%)	25.35	22.25	76.24	21.76	30.65	34.55	22.36	**20.02**
288	MAE	31.35	31.93	87.78	28.62	33.37	39.60	31.75	**25.51**
	RMSE	50.11	47.80	114.27	42.27	52.41	58.29	48.18	**38.34**
	MAPE(%)	20.19	21.87	88.18	21.24	25.26	27.45	22.21	**19.38**
336	MAE	32.15	31.89	91.04	28.26	34.17	40.32	32.26	**25.24**
	RMSE	50.87	47.82	117.03	41.66	53.52	58.20	48.65	**38.21**
	MAPE(%)	26.09	24.68	87.46	21.24	26.07	28.94	22.46	**19.17**

Table 3 summarises the experimental results on the PEMS04 dataset. At the short forecasting horizon of 96 steps, TSDFNet achieves MAE = 27.72 and RMSE = 41.56, obtaining the lowest absolute prediction errors among all compared methods. Although iTransformer attains the lowest MAPE at this horizon, TSDFNet still maintains a competitive MAPE of 25.36% while outperforming other baselines in MAE and RMSE. At 192

steps, TSDFNet continues to deliver the best MAE and RMSE values, with MAE = 29.04 and RMSE = 43.91, indicating its strong capability in reducing overall forecasting deviation. For longer horizons of 288 and 336 steps, TSDFNet further achieves MAE = 27.98 / 28.45 and RMSE = 42.40 / 43.19, consistently ranking first in both metrics. Compared with baseline models such as LSTM, DLinear and CycleNet, TSDFNet exhibits more stable long-term forecasting performance, demonstrating the effectiveness of the proposed trend-season dual-branch fusion architecture in capturing long-horizon traffic flow patterns.

Table 4 summarises the experimental results on the PEMS08 dataset. At the short forecasting horizon of 96 steps, TSDFNet achieves MAE = 25.84, RMSE = 38.13 and MAPE = 19.68%, outperforming all baseline models, including Pyraformer (MAE = 26.89, MAPE = 20.63%) and CycleNet (MAE = 31.19, MAPE = 21.41%). These results demonstrate that TSDFNet provides more accurate short-term forecasts while maintaining low variability across metrics. As the prediction horizon increases, the errors of most models grow considerably, whereas TSDFNet maintains steady performance. At 192 steps, TSDFNet records MAE = 26.75, RMSE = 40.20 and MAPE = 20.02%, remaining the most accurate model overall. For 288 and 336 steps, TSDFNet continues to lead with MAE = 25.51 / 25.24, RMSE = 38.34 / 38.21 and MAPE = 19.38% / 19.17%, showing minimal fluctuation across forecasting horizons. In contrast, baseline models such as Autoformer (MAPE = 88.18%) and iTransformer (MAPE ≈ 28%) exhibit substantial performance degradation, highlighting the robustness and long-term predictive stability of TSDFNet.

Table 5 – Prediction performance of different methods on LargeST-SD dataset

Time	Metric	Value							
		SVR	LSTM	Autoformer	Pyraformer	DLinear	iTransformer	CycleNet	TSDFNet
96	MAE	44.85	42.82	54.06	56.29	45.21	**29.32**	44.56	31.46
	RMSE	72.29	72.77	80.75	79.69	69.99	50.30	67.25	**48.82**
	MAPE(%)	29.86	31.19	48.41	54.18	36.25	**21.31**	37.30	23.88
192	MAE	43.59	40.49	55.69	40.17	38.04	**29.40**	38.98	29.92
	RMSE	71.02	68.51	83.82	62.66	63.52	52.11	64.24	**50.84**
	MAPE(%)	28.44	30.32	51.79	31.79	27.93	**21.48**	30.36	24.46
288	MAE	45.62	40.60	59.25	38.35	28.41	**28.33**	29.18	28.91
	RMSE	71.70	67.67	84.42	60.51	54.05	50.72	54.66	**49.86**
	MAPE(%)	26.46	30.79	52.41	29.15	**18.84**	20.49	20.99	22.16
336	MAE	45.86	42.35	51.47	38.16	28.28	28.40	27.98	**27.80**
	RMSE	72.05	69.04	81.68	60.50	53.22	50.91	53.90	**41.56**
	MAPE(%)	26.65	31.81	45.30	28.74	**17.84**	20.59	19.69	21.06

Table 5 reports the experimental results on the LargeST-SD dataset. At the short forecasting horizon of 96 steps, TSDFNet achieves MAE = 31.46, RMSE = 48.82 and MAPE = 23.88%, clearly outperforming baseline models such as Autoformer (MAPE = 48.41%) and CycleNet (MAPE = 37.30%). As the prediction horizon increases to 192 steps, the overall error of TSDFNet remains moderate (MAE = 29.92, RMSE = 50.84, MAPE = 24.46%), while most baselines experience pronounced degradation. At longer horizons, the performance of TSDFNet remains remarkably stable – achieving MAE = 28.91, RMSE = 49.86, MAPE = 22.16% at 288 steps, and MAE = 27.80, RMSE = 41.56, MAPE = 21.06% at 336 steps, demonstrating minimal error growth as the forecasting distance increases. Although CycleNet attains a slightly lower MAPE (19.69%) at the longest horizon, TSDFNet consistently achieves the lowest RMSE across all horizons and remains competitive in MAE and MAPE, demonstrating robust long-horizon forecasting performance on LargeST-SD.

Visual comparisons in *Figure 4* also confirm that TSDFNet's predictions closely align with ground truth, effectively capturing both global traffic trends and fine-grained variations.

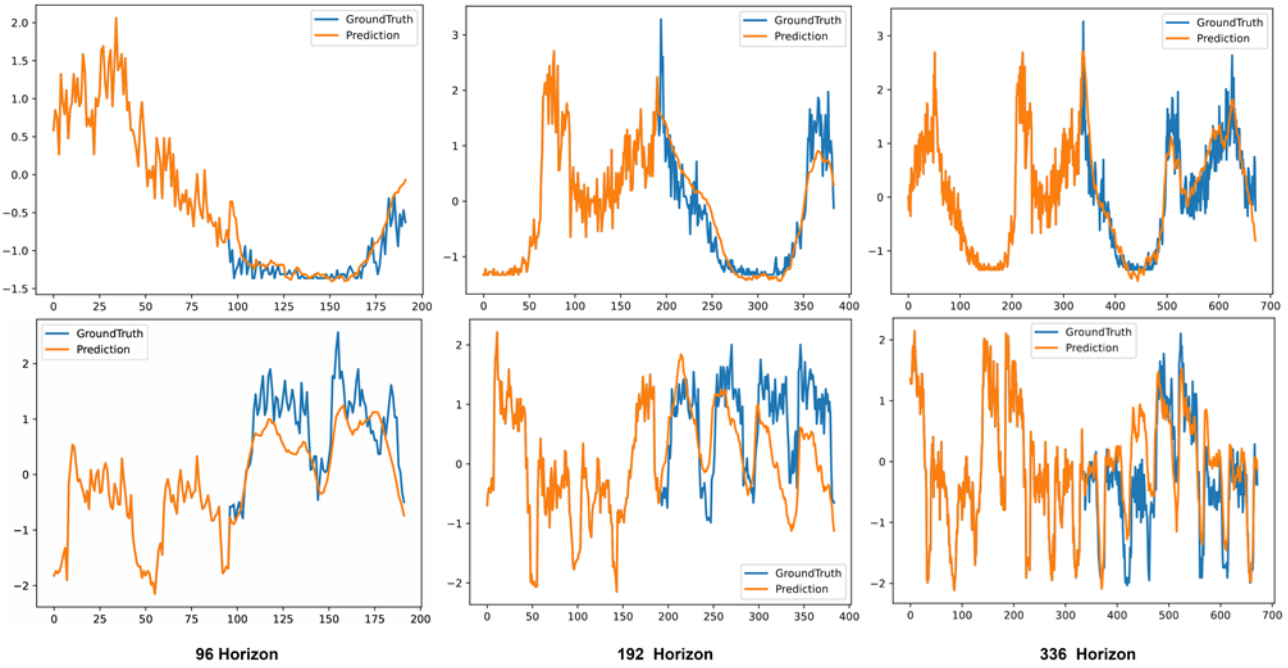

Figure 4 – Visualisation of long-term forecasting results on PEMS08 and LargeST-SD datasets

Results on urban traffic dataset (Glasgow)

Table 6 – Prediction performance of different methods on Glasgow dataset

Time	Metric	Value							
		SVR	LSTM	Autoformer	Pyraformer	DLinear	iTransformer	CycleNet	TSDFNet
96	MAE	30.59	25.22	21.30	14.01	18.97	15.01	18.05	**13.98**
	RMSE	39.92	31.62	27.78	23.47	24.12	23.42	26.04	**22.49**
	MAPE(%)	85.12	72.48	79.18	**60.36**	71.92	68.29	62.69	67.27
192	MAE	32.07	28.63	20.26	14.93	15.46	**12.69**	17.17	13.18
	RMSE	39.68	30.21	32.76	23.78	22.59	**18.91**	24.15	21.37
	MAPE(%)	83.45	74.03	84.79	64.13	67.88	**59.59**	66.53	65.75
288	MAE	36.99	29.27	22.83	18.61	16.02	14.79	16.17	**14.29**
	RMSE	45.56	38.07	33.35	25.65	25.20	23.75	24.38	**22.47**
	MAPE(%)	89.96	88.19	87.22	71.37	69.22	**58.76**	64.74	61.21
336	MAE	36.30	29.11	25.14	15.12	19.94	16.35	20.05	**13.62**
	RMSE	44.18	40.59	33.02	23.56	27.14	23.69	27.32	**20.99**
	MAPE(%)	81.61	83.90	80.11	70.01	**65.96**	69.85	72.81	71.07

Table 6 presents the results on the Glasgow dataset, which reflects complex and highly dynamic urban traffic conditions. At the short forecasting horizon of 96 steps, TSDFNet achieves MAE = 13.98, RMSE = 22.49 and MAPE = 67.27%. It delivers the lowest MAE and RMSE among all methods, outperforming strong baselines such as Pyraformer (MAE = 14.01) and iTransformer (MAE = 15.01), although Pyraformer attains a slightly

lower MAPE (60.36%). As the prediction horizon increases, TSDFNet sustains a stable level of accuracy. At 192 steps, it achieves MAE = 13.18, RMSE = 21.37 and MAPE = 65.75%. When the horizon extends further to 288 steps, TSDFNet remains competitive with MAE = 14.29, RMSE = 22.47 and MAPE = 61.21%, outperforming most models such as Autoformer (MAPE = 87.22%) and DLinear (69.22%). Even at the longest horizon of 336 steps, TSDFNet maintains MAE = 13.62, RMSE = 20.99 and MAPE = 71.07%, indicating robust long-term forecasting capability under irregular and highly variable traffic patterns.

Collectively, the results across all four datasets – PEMS04, PEMS08, LargeST-SD and Glasgow – demonstrate that TSDFNet consistently achieves accurate and stable long-term traffic forecasts across diverse roadway types, ranging from structured freeway networks to complex urban systems. This confirms its strong generalisation ability and practical reliability for real-world intelligent transportation applications.

4.5 Ablation experiment

To investigate the effectiveness of each module within the proposed architecture, an ablation study is conducted on the PEMS04 dataset with a prediction length of 96. The results, shown in *Table 7*, provide insight into the functional importance of the trend module, seasonal module and frequency-aware fusion mechanism.

Table 7 – Ablation results on the PEMS04 dataset

Method Variant	MAE	RMSE
Full model	27.72	41.56
w/o TrendModule	45.97	66.73
w/o WaveletSeason	31.12	46.40
w/o Weight	29.44	43.78
OnlyTrend	30.56	44.29
OnlySeason	39.31	57.01

As shown in *Table 7*, the complete TSDFNet attains the best performance with an MAE of 27.72 and an RMSE of 41.56. When the trend module is removed, the prediction accuracy drops sharply (MAE 45.97, RMSE 66.73), whereas the OnlyTrend variant still achieves competitive results (MAE 30.56, RMSE 44.29). Similarly, removing the seasonal module (w/o WaveletSeason) yields MAE 31.12 and RMSE 46.40, while the OnlySeason model performs considerably worse (MAE 39.31, RMSE 57.01).

Through comparison, it can be observed that both the absence and isolation of the trend branch lead to smaller performance degradation than those of the seasonal branch, and the results of trend-related variants remain closer to the complete model. This indicates that the trend branch contributes more significantly to the overall improvement and serves as the primary source of forecasting accuracy. The hierarchical multi-scale fusion mechanism in this branch effectively aggregates long-range temporal dependencies and captures the slow-evolving patterns of traffic flow, thereby mitigating cumulative prediction errors and enhancing stability over long horizons. In contrast, although the seasonal branch contributes less prominently, it still plays an important complementary role. By decomposing high-frequency components through wavelet transformation, this branch enhances the model's sensitivity to periodic fluctuations, such as daily and weekly traffic cycles, and refines short-term variations that the trend module alone cannot capture. Consequently, the combination of both branches enables TSDFNet to maintain global consistency while adapting to fine-grained temporal dynamics.

The frequency-aware weighting mechanism also contributes modestly to the overall accuracy. Removing it (w/o Weight) results in a slight degradation (MAE 29.44, RMSE 43.78), indicating that adaptive fusion of the two branches refines prediction balance rather than fundamentally altering the modelling capability.

Overall, the ablation results reveal that the trend branch provides the primary source of improvement by capturing stable, low-frequency structures, while the seasonal branch and weighting mechanism play supporting roles that enhance temporal adaptability and consistency.

5. CONCLUSIONS

This study proposed TSDFNet, a trend–season dual-branch fusion network for long-term traffic flow forecasting. The model decomposes traffic sequences into distinct trend and seasonal components and employs a harmonic weighting mechanism to adaptively integrate them in the frequency domain.

Experimental evaluations on four benchmark datasets (PEMS04, PEMS08, LargeST-SD and Glasgow) demonstrate that explicitly decoupling long-term traffic dynamics into low-frequency trends and high-frequency seasonal fluctuations significantly enhances forecasting stability and accuracy across extended horizons. TSDFNet generally achieves competitive or lower prediction errors than state-of-the-art baselines in MAE and RMSE across most datasets and horizons, confirming its robustness and generalisation capability under diverse temporal and spatial conditions.

The results further reveal that hierarchical multi-scale trend fusion serves as the primary contributor to performance improvement, effectively capturing slow-evolving structural variations. Meanwhile, the wavelet-based seasonal branch refines short-term periodic patterns, and the harmonic weighting mechanism adaptively balances the two, strengthening model resilience under non-stationary and noisy environments. These findings highlight that frequency-aware temporal decomposition offers a powerful and interpretable paradigm for modelling complex, multi-scale traffic dynamics.

Despite its effectiveness, TSDFNet still relies on fixed prediction horizons and univariate inputs. Future research will focus on extending this framework to multivariate and multimodal forecasting, incorporating external factors such as weather, events and spatial dependencies. Moreover, exploring dynamic frequency fusion and interpretable adaptive weighting mechanisms will further enhance the generalisation and practical applicability of TSDFNet, ultimately contributing to more adaptive and reliable decision support in intelligent transportation systems.

ACKNOWLEDGEMENTS

Supported by Intelligent Policing Key Laboratory of Sichuan Province, No. ZNJW2024KFQN013, ZNJW2024KFQN011 and ZNJW2024KFMS003.

REFERENCES

[1] Hossain MN, Ahmed N, Wazid Ullah SM. Traffic flow forecasting in intelligent transportation systems prediction using machine learning. *Proceedings of the 2022 International Conference on Futuristic Technologies (INCOFT)*, 25-27 Nov. 2022, Belgaum, India. 2022. p. 1–5. DOI: 10.1109/INCOFT55651.2022.10094319.

[2] Yu B, Yin H, Zhu Z. Spatio-temporal graph convolutional networks: a deep learning framework for traffic forecasting. *Proceedings of the 27th International Joint Conference on Artificial Intelligence (IJCAI'18)*. Stockholm, Sweden, 13–19 July 2018. AAAI Press; 2018. p. 3634–3640. DOI: 10.24963/IJCAI.2018/505.

[3] Li Y, Yu R, Shahabi C, Liu Y. Diffusion convolutional recurrent neural network: data-driven traffic forecasting. *Proceedings of the 6th International Conference on Learning Representations (ICLR 2018)*, 30 Apr. - 3 May 2018, Vancouver, Canada. 2018. arXiv:1707.01926.

[4] He Z, Chow CY, Zhang JD. STCNN: A spatio-temporal convolutional neural network for long-term traffic prediction. *Proceedings of the 20th IEEE International Conference on Mobile Data Management (MDM 2019)*. Hong Kong, China, 10–13 June 2019. 2019. p. 226–233. DOI: 10.1109/MDM.2019.00053.

[5] Chen YT, et al. Traffic flow prediction based on spatial-temporal multi factor fusion graph convolutional networks. *Scientific Reports*. 2025;15(1):12612. DOI: 10.1038/s41598-025-96801-1.

[6] Hou Z, Li X. Repeatability and similarity of freeway traffic flow and long-term prediction under big data. *IEEE Transactions on Intelligent Transportation Systems*. 2016;17(6):1786–1796. DOI: 10.1109/TITS.2015.2511156.

[7] Kumar SV, Vanajakshi L. Short-term traffic flow prediction using seasonal ARIMA model with limited input data. *European Transport Research Review*. 2015;7(3):21. DOI: 10.1007/s12544-015-0170-8.

[8] Wang H, et al. Empirical mode decomposition–autoregressive integrated moving average hybrid short-term traffic speed prediction model. *Transportation Research Record.* 2014;2460(1):66–76. DOI: 10.3141/2460-08.

[9] Chi Z, Shi L. Short-term traffic flow forecasting using ARIMA–SVM algorithm and R. *Proceedings of the 5th International Conference on Information Science and Control Engineering (ICISCE 2018)*. Chengdu, China, 20–22 July 2018. 2018. p. 517–522. DOI: 10.1109/ICISCE.2018.00114.

[10] Ye J, Xue S, Jiang A. Attention-based spatio-temporal graph convolutional network considering external factors for multi-step traffic flow prediction. *Digital Communications and Networks.* 2022;8(3):343–350. DOI: 10.1016/j.dcan.2021.09.007.

[11] Xu Y, et al. Urban traffic flow prediction: a spatio-temporal variable selection-based approach. *Journal of Advanced Transportation*. 2016;50(4):489–506. DOI: 10.1002/atr.1356.

[12] Chen Y, et al. Acting as a decision maker: traffic-condition-aware ensemble learning for traffic flow prediction. *IEEE Transactions on Intelligent Transportation Systems*. 2022;23(4):3190–3200. DOI: 10.1109/TITS.2020.3032758.

[13] Wang D, et al. Bayesian optimization of support vector machine for regression prediction of short-term traffic flow. *Intelligent Data Analysis*. 2019;23(2):481–497. DOI: 10.3233/IDA-183832.

[14] Cheng S, Lu F, Peng P, Wu S. Short-term traffic forecasting: an adaptive ST-KNN model that considers spatial heterogeneity. *Computers, Environment and Urban Systems*. 2018;71:186–198. DOI: 10.1016/j.compenvurbsys.2018.05.009.

[15] Tang J, et al. An improved fuzzy neural network for traffic speed prediction considering periodic characteristic. *IEEE Transactions on Intelligent Transportation Systems*. 2017;18(9):2340–2350. DOI: 10.1109/TITS.2016.2643005.

[16] Liu Z, et al. A hybrid short-term traffic flow forecasting method based on neural networks combined with K-nearest neighbor. *Promet – Traffic & Transportation*. 2018;30(4):445–456. DOI: 10.7307/ptt.v30i4.2651.

[17] Akin M, Sagiroglu S. Short-term traffic speed prediction with RNN method for roads characterized by density-based clustering method. *Journal of the Faculty of Engineering and Architecture of Gazi University*. 2022;37(2):581–593. DOI: 10.17341/gazimmfd.921035.

[18] Xia D, et al. RSAB-ConvGRU: A hybrid deep-learning method for traffic flow prediction. *Multimedia Tools and Applications*. 2023;83(7):20559–20585. DOI: 10.1007/s11042-023-15877-x.

[19] Lu H, et al. A temporal-aware LSTM enhanced by loss-switch mechanism for traffic flow forecasting. *Neurocomputing.* 2021;427:169–178. DOI: 10.1016/j.neucom.2020.11.026.

[20] Walch M, Neubauer M, Schildorfer W. Floating car data-based short-term travel time forecasting with deep recurrent neural networks incorporating weather data. *Journal of Transportation Engineering, Part A: Systems.* 2023;149(6):04023059. DOI: 10.1061/JTEPBS.TEENG-7647.

[21] Zeng A, Chen M, Zhang L, Xu Q. Are transformers effective for time series forecasting? *Proceedings of the Thirty-Seventh AAAI Conference on Artificial Intelligence (AAAI 2023).* 7–14 Feb. 2023, Washington, DC, USA. 2023. p. 11121–11128. DOI: 10.1609/aaai.v37i9.26317.

[22] Wang S, et al. TimeMixer: decomposable multiscale mixing for time series forecasting. *Proceedings of the International Conference on Learning Representations (ICLR 2024).* 7–11 May 2024, Vienna, Austria. 2024. DOI: 10.48550/arXiv.2305.08891.

[23] Yu H, Guo P, Sano A. AdaWaveNet: Adaptive wavelet network for time series analysis. *Transactions on Machine Learning Research.* 2024 Nov. Epub ahead of print. DOI: 10.48550/arXiv.2306.10931.

[24] Jiang J, Han C, Zhao WX, Wang J. Propagation delay-aware dynamic long-range Transformer for traffic flow prediction. *Proceedings of the AAAI Conference on Artificial Intelligence (AAAI 2023)*. 7–14 Feb. 2023, Washington, DC, USA. 2023. p. 4493–4501. DOI: 10.1609/aaai.v37i4.25556.

[25] Xiao J, Long B. A multi-channel spatial-temporal Transformer model for traffic flow forecasting. *Information Sciences.* 2024;671:120648. DOI: 10.1016/j.ins.2024.120648.

[26] Evans J, Waterson B, Hamilton A. Forecasting road traffic conditions using a context-based random forest algorithm. *Transport Planning and Technology.* 2020;43(2):123–138. DOI: 10.1080/03081060.2020.1728470.

[27] Wang Z, Su X, Ding Z. Long-term traffic prediction based on LSTM encoder-decoder architecture. *IEEE Transactions on Intelligent Transportation Systems.* 2021;22(10):6561–6571. DOI: 10.1109/TITS.2020.2995546.

[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. DOI: 10.1109/ACCESS.2021.3050836.

[29] Wu H, Xu J, Wang J, Long M. Autoformer: decomposition transformers with auto-correlation for long-term series forecasting. *Proceedings of the 35th Conference on Neural Information Processing Systems (NeurIPS 2021)*. 6–14 Dec. 2021, Virtual Conference. 2021. p. 22419–22430. DOI: 10.48550/arXiv.2106.13008.

[30] Chan RKC, Lim JMY, Parthiban R. Long-term traffic speed prediction utilizing data augmentation via segmented time frame clustering. *Knowledge-Based Systems.* 2025;308:112785. DOI: 10.1016/j.knosys.2025.112785.

[31] Wang X, Chen Y, Huang T. Long-horizon traffic flow forecasting using hierarchical temporal aggregation and recurrent residual modeling. *IEEE Transactions on Intelligent Transportation Systems.* 2024. Early Access. DOI: 10.1109/TITS.2024.3365129.

[32] Sweldens W. The lifting scheme: a construction of second generation wavelets. *SIAM Journal on Mathematical Analysis.* 1998;29(2):511–546. DOI: 10.1137/S0036141095289051.

[33] Chen C, et al. Freeway performance measurement system: mining loop detector data. *Transportation Research Record: Journal of the Transportation Research Board.* 2001;1748(1):96–102. DOI: 10.3141/1748-12.

[34] Liu X, et al. LargeST: A benchmark dataset for large-scale traffic forecasting. *Proceedings of the 37th Conference on Neural Information Processing Systems Datasets & Benchmarks*. 10–16 Dec. 2023, Virtual Conference. 2023. p. 1–12. DOI: 10.48550/arXiv.2306.08259.

[35] Li Y, Zhao Q, Wang M. High-resolution traffic flow data from the urban traffic control system in Glasgow. *Scientific Data.* 2025;12(1):253. DOI: 10.1038/s41597-025-04494-y.

[36] Liu S, et al. Pyraformer: Low-complexity pyramidal attention for long-range time series modeling and forecasting. *Proceedings of the International Conference on Learning Representations (ICLR 2022).* 25–29 Apr. 2022, Virtual Conference. 2022. DOI: 10.48550/arXiv.2201.12746.

[37] Liu Y, et al. iTransformer: Inverted transformers are effective for time series forecasting. *Proceedings of the International Conference on Learning Representations (ICLR 2024).* 7–11 May 2024, Vienna, Austria. 2024. DOI: 10.48550/arXiv.2310.06625.

[38] Lin S, et al. Cyclenet: Enhancing time series forecasting through modeling periodic patterns. *Proceedings of the 38th Conference on Neural Information Processing Systems (NeurIPS 2024)*. 9–15 Dec. 2024, New Orleans, LA, USA. 2024. p. 106315-106345. DOI: 10.48550/arXiv.2405.11892.