AVE **Trends in Intelligent Computing Systems**

Multi-Horizon Traffic Flow Prediction via Dual Attention Fusion of Transformer and Bidirectional LSTM

G. Deena[1,*], Gollapally Abhiram Sai[2], T. V. Abhinand[3], Molakalapalli Vamsi Krishna[4], M. Rehena Sulthana[5]

[1,2,3,4]Department of Computer Science and Engineering, SRM Institute of Science and Technology, Ramapuram, Chennai, Tamil Nadu, India.
[5]School of Information Technology and Engineering, Melbourne Institute of Technology, Melbourne, Victoria, Australia.
deenag@srmist.edu.in[1], ag9946@srmist.edu.in[2], at9651@srmist.edu.in[3], vm9573@srmist.edu.in[4], rsulthana@academic.mit.edu.au[5]

*Corresponding author

Abstract: Intelligent transportation requires traffic flow prediction. This aids congestion control, route planning, and infrastructure optimisation. Correct short- and long-term forecasting helps urban transportation and the environment. ARIMA models struggles with nonlinear and nonstationary traffic data. While LSTM networks improve temporal modelling, their sequential structure makes it difficult to express long-range dependencies and slows computation. Recent transformer topologies show promise. They represent global dependencies through self-attention. To simplify networks, this study uses an Encoder-Only Transformer framework to predict traffic flow in several phases without a decoder. The suggested method outperforms LSTM for long-horizon forecasting and decreases inference time on benchmark datasets. However, an encoder-only architecture has drawbacks. It has one attention pathway, no weighting, and is unreliable in traffic. Hyperparameter tuning via grid search increases processing costs and may not optimise performance. The Hybrid Dual-Attention Fusion Framework in this study handles this. An Encoder-Only Transformer, a BiLSTM network, and an APSO algorithm are used. Temporal Multi-Head Self-Attention and Adaptive Feature Recalibration Attention are used. The unique Dual-Attention process uses Temporal Multi-Head Self-Attention and Adaptive Feature Recalibration Attention. The BiLSTM module captures dependencies in both directions to boost sequential context learning. In contrast, the APSO-based tuning module optimises the embedding dimension, the number of attention heads, and the learning rate. Experimental results reveal that prediction accuracy is more consistent and generalizable across datasets.

Keywords: Traffic Flow Prediction; Intelligent Transportation Systems (ITS); Multi-Horizon Forecasting; Dual Attention Mechanism; LSTM Networks; Embedding Dimension; Autoregressive Integrated Moving Average (ARIMA).

Cite as: G. Deena, G. A. Sai, T. V. Abhinand, M. V. Krishna, and M. R. Sulthana, "Multi-Horizon Traffic Flow Prediction via Dual Attention Fusion of Transformer and Bidirectional LSTM," *AVE Trends in Intelligent Computing Systems*, vol. 3, no. 2, pp. 104–120, 2026.

Journal Homepage: https://www.avepubs.com/user/journals/details/ATICS

Received on: 18/05/2025, **Revised on:** 11/09/2025, **Accepted on:** 22/10/2025, **Published on:** 05/05/2026

DOI: https://doi.org/10.64091/ATICS.2026.000260

1. Introduction

As a basic component of Intelligent Transportation Systems (ITS), traffic flow prediction helps people manage vehicle movement and reduce traffic congestion by providing control over built structures. As populations in cities and the number of automobiles increase rapidly, modern transportation networks face challenges such as travel delays and underutilised assets. Accurate traffic flow predictions remain necessary to make data-driven choices to find the best path and control traffic signals. Estimating future states through traffic flow forecasting involves examining time-series data from loop detectors and GPS-enabled devices. Even though short-term prediction helps with immediate operations, experts claim that forecasting for many hours ahead, such as 6, 12, and 24 hours, is needed for strategic planning and long-term management. Transportation networks exhibit non-steady behavior due to many factors, such as daily travel habits, weather conditions, and accidents, that influence the movement of cars. Such complex features pose challenges for traditional methods that seek to understand how time-based patterns operate across different settings. Mechanical models, such as the Autoregressive Integrated Moving Average (ARIMA), process historical information under linear assumptions based on stationary time-series properties. These classical methods are efficient and interpretable.

However, due to the nonlinearity and complexity of the real-world traffic systems, these methods have become less effective, especially for long-horizon forecasting. Machine learning and deep learning methods have therefore been widely used to deal with these challenges. Recurrent Neural Networks (RNNs), especially Long Short-Term Memory (LSTM), are designed to capture temporal dependencies through memory cells and gating mechanisms. LSTM naturally overcomes the gradient vanishing problem. However, the major drawback of using LSTMs is that computation is slow and cannot be parsed or parallelized at the sequence level. Also, it still cannot model extremely long-period dependency in highly dynamic traffic scenarios. Transformers based on attention mechanisms have demonstrated remarkable advantages, as they learn global interactions across all time steps and compute them in parallel at different time steps. However, typical encoder–decoder Transformer models still incur high computational costs during inference due to autoregressive decoding at each future time step. In addition, they are less sensitive to local changes than LSTMs and require fewer parameters. Therefore, LSTMs and Transformers are complementary, and it is necessary to design hybrid models that teach global and local dependencies and achieve parallel computation efficiency. Moreover, most existing works determine the model architecture and look-back window size via grid search and online reinforcement learning, both of which require substantial computational resources.

Table 1: Key challenges in traffic flow prediction

Challenge	Description	Impact
Nonlinearity	Complex variable interactions	Reduces prediction accuracy
Non-stationarity	Time-varying statistical properties	Limits generalization
Long-range dependencies	Distant time-step influence	Challenging for sequential models

Hence, these limitations inspire us to design a more robust, efficient, and scalable deep learning framework for accurate multi-horizon forecasting across networks. A Hybrid Dual-Attention Fusion Framework is proposed for efficient traffic flow forecasting, which is inherited in our architecture as an efficient spatiotemporal feature extractor. After learning feature interactions in both dimensions, an encoder-only Transformer and a Bidirectional LSTM (BiLSTM) are ensembled at each boundary node for sequence modelling and multi-horizon forecasting. The encoder-only architecture omits future time-step observations during training and is much lighter than the encoder–decoder architecture, reducing computational complexity and training costs. Meanwhile, BiLSTM models data as a sequence of individual observations, capturing the consecutive changes between periods in the short term and the relationships between very early and very late periods in the long term. A Dual-Attention mechanism consists of the temporal Multi-Head Self-Attention and the Squeeze-and-Excitation (SE) layer for learning and enhancing feature interactions. An Adaptive Particle Swarm Optimization (APSO) algorithm is designed to efficiently search for hidden representations and hyperparameters, thereby improving model convergence and increasing prediction robustness during training under extremely dynamic traffic conditions.

Table 2: Comparison of existing models

Model	Strength	Limitation
ARIMA	Simple, interpretable	Cannot model nonlinear patterns
LSTM	Captures temporal dependencies	Sequential, inefficient for long sequences
Transformer	Captures global dependencies	High computation, weak local context
Proposed Model	Hybrid + dual attention + APSO	Balanced performance

The effectiveness and efficiency of the proposed methods were evaluated on real-world traffic cases for 6, 12, and 24h forecasting. The results show improvements in accuracy, stability, efficiency, and generalizability. Traffic flow prediction has many challenges due to the complex characteristics of the real world. The major challenges are summarised in Table 1. The above challenges highlight the limitations of existing traffic prediction approaches and underscore the need for more robust, adaptive models. A comparative analysis of commonly used models is presented in Table 2. The main contributions of this work are summarised as follows:

- A hybrid architecture combining Encoder-Only Transformer and BiLSTM to capture both global and local temporal dependencies.
- A Dual-Attention mechanism for dynamic feature prioritisation in complex traffic scenarios.
- An APSO-based hyperparameter optimisation framework for efficient and automated tuning.
- Improved multi-horizon traffic prediction performance with enhanced accuracy, stability, and generalisation.

2. Literature Review

For example, a machine learning algorithm can be used to predict rush hour based on historical data and instant social media indicators. Traffic data is highly nonlinear and dynamic, influenced by multiple interacting factors, including time of day, road network topology, weather, incidents, and demand variations. Hence, multi-horizon forecasts in traditional statistical methods are a weakness [1]. To that end, a wide range of deep learning, graph-based, hybrid, and probabilistic approaches have been proposed to improve prediction accuracy, robustness, and computational efficiency. Recent studies have highlighted the importance of capturing both short-term, decisive changes and long-range dependencies, given the observed periodic patterns in traffic measurements. However, Zhang et al. [10] suggested that the Encoder-Only Transformer can be used to predict multi-step traffic flow, reducing computational cost without sacrificing accuracy. Thus, traffic flow prediction is fundamental to ITS, enabling it to alleviate traffic congestion, optimize traffic signal timing, and smooth urban transportation. The traditional sequential procedure of the LSTMs is inefficient for identifying long-distance dependencies, though they handle temporal relations well [2]. The standard Encoder–Decoder Transformer can model global dependencies effectively via self-attention, but its decoder imposes a heavy computational load, which is not ideal for real-time traffic prediction. Results showed that the Encoder-Only Transformer beat LSTM models across all timeframes on the Minnesota data, improving Mean Absolute Error by up to 17.33%, demonstrating that it handles regular traffic patterns well [3].

On the California dataset, which has more irregular traffic, the model did better on long-term forecasts, especially 24-hour-ahead predictions. But for shorter horizons, such as 6 and 12 hours, it didn’t perform as well as LSTM or the standard Encoder–Decoder Transformer [4]. This suggests it’s strong at capturing long-range dependencies, but its success depends on the data set and traffic variability. In summary, the study points to simplifying Transformer models as a good middle ground between speed and accuracy for traffic prediction. It also shows that different datasets and traffic conditions affect results, highlighting some weaknesses in short-term forecasts [5]. These insights encourage further work on hybrid or adaptive models that blend sequential and attention-based methods to deliver more reliable predictions across varied traffic conditions. Qin [1] introduced a new approach to predicting traffic flow by combining two data types using a multimodal fusion method. Multimodal Compact Bilinear (MCB) pooling combines attention mechanisms to capture the complex spatiotemporal patterns in vehicle movement in urban areas. These difficult patterns are better caught by Multimodal Compact Bilinear (MCB) pooling when the method integrates these specific tools. Experts claim that surrounding conditions, including street layouts, historical traffic patterns, and external factors such as weather, holidays, and seasons, significantly affect traffic flow predictions by shaping how vehicles move. External factors influence traffic flow prediction, while street layout remains a primary factor. Because the actual environment is diverse, traffic flow prediction requires a deep understanding of the many factors involved [6].

Traditional single-source data methods fail to address these intricate connections. To improve predictions, the authors developed a strategy that combines multiple data types to stabilise the model and enhance its accuracy [7]. Their approach is to first create a weighted spatio-temporal graph from time-series data derived from traffic flows. Nodes in the spatial graph are traffic sensors or road sections, and edges spatially connect them. They then use a Weighted Spatio-Temporal Synchronous Graph Convolutional Network (STSGCN) to extract features across the spatial and temporal dimensions simultaneously, capturing the traffic evolution pattern and the statistical dependencies between the two. To aggregate heterogeneous data types, the model implements a two-channel fusion method based on MCB pooling and attention. MCB pooling efficiently merges them, capturing interactions between high-dimensional spatio-temporal and visual features [8]. The attention mechanism adjusts the importance of each feature, so the model pays more attention to the most relevant information to make predictions. This combined vector is further mixed with additional context, such as weather data, to incorporate environmental factors into the prediction. Huang et al. [2] have proposed a Multi-step Coupled Graph Convolutional Network (MCGCN) enhanced with a temporal attention mechanism for better traffic flow prediction in ITS. A traffic prediction model must provide accurate forecasts for urban road planning, traffic control, and transportation management. This, however, is a difficult task because not only are traffic networks complex in structure, but traffic patterns also vary greatly over time.

It is in this context that, in real traffic systems, there are many spatial dependencies associated with nearby roads in close interaction, as well as across the entire network [9]. Also, traffic will depend on seasonal and hourly changes, as well as long-term trends. The various relationships are difficult to represent simultaneously using traditional modelling techniques. To tackle these issues, the authors developed a new framework that integrates graph convolutional methods with temporal modelling and attention mechanisms. Its first component is the Multi-Step Coupled Graph Convolution (MCGC) module, designed to learn fine-grained spatial features of the traffic network. That takes local and global spatial dependencies simultaneously as input to learn the node connections in the space, i.e., in a traffic network. Testing the model on real-world datasets such as NYCTaxi and NYCBike, which represent different and complex traffic conditions [10]. The results showed that MCGCN outperforms nine other state-of-the-art methods across several metrics, demonstrating that the model effectively captures spatial and temporal dependencies. Combining graph convolution, recurrent units, and attention yields a model with higher accuracy and reliability. Overall, the study emphasises the importance of coupled spatial-hierarchical and temporal-dependency modelling. However, the model's more complex design might make it harder to scale or run in real time [1].

Inoue and Miyashita [3] proposed a short-term traffic speed prediction model that uses the basic and cointegration relationships between speed and density, covering both free-flow and congested traffic conditions. Predicting traffic states is key to Intelligent Transportation Systems (ITS), helping with tasks such as traffic control, route selection, and infrastructure planning. While many models learn patterns directly from data using statistical or machine learning methods, only a few explicitly consider the theoretical connections among traffic variables outlined in the fundamental traffic flow diagram. This diagram shows how speed, density, and flow relate, and how these relationships change significantly depending on traffic conditions [2]. The authors pointed out that ignoring these links can reduce the clarity and accuracy of prediction models, especially when traffic shifts between flowing smoothly and getting jammed. To tackle this, their study introduces a threshold error correction model (TECM) that captures both long-term balance and short-term swings in traffic variables [3]. The model distinguishes between free-flow and congested states using set thresholds based on speed or density. This allows it to represent different speed-density patterns for each state, thereby improving the fit between the predictions and actual traffic. They use cointegration to express the long-run balance between speed and density, while the error correction adjusts for short-term deviations [4]. Combining statistical methods with theoretical insights provides greater clarity and accuracy than relying solely on data-driven techniques. Still, the model’s reliance on fixed thresholds and its relatively simple structure might limit its ability to capture complex nonlinear patterns often observed in large urban traffic systems [5].

This points to chances for future work that blends this approach with more advanced deep learning methods. Shaukat et al. [4] and their team reviewed multi-step attack detection, focusing on the cybersecurity threats these attacks pose to critical infrastructure and key systems across sectors. Such multi-step attacks, also known as multi-stage or composite attacks with several correlated phases, often involve a long dwell time and evade detection by conventional tools. Since these attacks occur in stages, they are intricate and difficult to detect using conventional security systems [6]. The authors point out that modern cybersecurity needs to address challenges such as processing large volumes of complex data, identifying the sequence of attack steps, and maintaining real-time detection in constantly changing environments. The paper also identifies major challenges in existing detection systems, including frequent false alarms, difficulty selecting and adapting to new and evolving attack signatures, and challenges in data mining and managing large, complex data sets. It uses elaborate statistics and visuals to present industry-specific behaviour and demonstrate commonly observed problems across studies [7]. One important part of the review is the repeated weaknesses and gaps in existing methods, which require stronger, more flexible, and scalable solutions to manage new vulnerabilities and emerging threats. In summary, this review provides professionals and researchers with a clear understanding of the advantages and disadvantages of existing multi-step attack detection methods and guides future related work. It highlights the value of building layered, hybrid detection systems that combine different approaches to achieve high accuracy, mitigate false positives, and enable faster, real-time responses [8].

While this study is primarily oriented towards cybersecurity, its proposed conceptualisation of analysing sequences, multistep forecasting, and handling high-dimensional, complex data may be applied to novel modelling approaches in other fields, such as predicting traffic flow. Sarıcıoğlu et al. [5] examined how well one-step and multi-step forecasting methods reduce the Bullwhip Effect and improve supply chain performance in an order-up-to-level inventory system. The Bullwhip Effect occurs when small changes in customer demand lead to larger swings in orders and inventory as you move up the supply chain. This leads to problems such as excess stock, stockouts, higher costs, and worse service. Basic forecasting methods like Moving Average and Exponential Smoothing have been used to deal with this, but they don’t handle complex or variable demand very well. The results showed that multi-step forecasting performed much better than one-step forecasting at reducing demand swings and stabilising the supply chain [9]. The LSTM and LightGBM models performed especially well at capturing time-based patterns, seasonality, and changes in demand. This led to improvements in key supply chain metrics, including increased order fulfilment, stabilised inventory variability, and optimised average end inventory. The study points out that using multi-step forecasting provides a better picture of future demand, on which informed planning decisions can be made, including decisions about the inventories to hold. Overall, the research focuses on the importance of advanced forecasting methods for solving the real-world problem of the Bullwhip Effect. It shows that machine learning-based multi-step forecasts tend to be more accurate, stable, and adaptable than older methods [10].

Although this study focuses on supply chains, its insights apply to other areas that work with time-series data, such as traffic flow, where predicting multiple steps is also important. The paper also notes that while advanced models can improve efficiency and reduce variability, it's important to balance the model's complexity with the computing power it requires for practical use. Fiedler and Lucia [6] proposed a probabilistic multi-step identification method that includes implicit state estimation to improve Stochastic Model Predictive Control (SMPC). This work tackles important challenges in modelling and controlling systems under uncertainty. SMPC has become popular for controlling multivariable systems operating in unpredictable, changing environments. But a main difficulty lies in obtaining an accurate probabilistic model that captures both system behaviour and uncertainties, such as process and measurement noise [1]. Traditional methods often rely on recursive state-space models, which become complicated and unreliable when dealing with time-varying uncertainty. To mitigate these issues, the authors formulated a framework that outputs a "set of future system state predictions in one shot", instead of predicting system states one step at a time. This multi-step approach amends recursive calculations and constitutes a more stable and efficient approach to uncertainty management [2]. A key point in their method is embedding implicit state estimation within this multi-step identification. They assume a linear system with noisy measurements and unknown noise levels, showing that, on average, using this multi-step model is equivalent to estimating the initial state distribution and propagating it through the system dynamics [3].

This eliminates the need for explicit state-estimation steps, making modelling simpler without sacrificing probabilistic details. In summary, this study shows that multi-step modeling can better handle uncertainty and improve predictions in complex systems [4]. While focused on control systems and SMPC, ideas about probabilistic multi-step forecasting, uncertainty handling, and efficient state estimation could be useful in other fields, such as traffic flow prediction, where dealing with uncertainty and forecasting multiple future steps is also important. Le et al. [7] proposed a method using multi-step predictions to improve adaptive sampling in mobile robotic sensor networks, aiming to make monitoring environmental changes more efficient. In these systems, mobile sensors move around to gather data from different spots. The primary objective is to successfully identify subsequent sampling steps to acquire more information at a minimum uncertainty level and with limited resource consumption. Most traditional methods follow a strategy of making predictions only one step at a time, resulting in the loss of valuable information about the future and sub-optimal navigation and data collection. To fix this, the authors developed a framework that enables sensors to predict multiple steps and make smarter decisions [5]. Their method uses a Gaussian Process (GP) model to represent the area being monitored. This helps estimate values in areas without measurements and assesses the uncertainty of those predictions. With this probabilistic model, they formulated adaptive sampling as an optimisation problem based on conditional entropy to reduce uncertainty in predicted values. Using multi-step predictions and the chain rule of conditional entropy, the system plans optimal sensor paths over several future steps, thereby improving movement planning, coverage, and data quality [6].

They tested their framework using simulations based on real datasets. They found that it performs better than traditional single-step methods, especially in navigation, sensor placement, and the amount of information gained. The study shows the benefits of using multi-step forecasts and optimisation in dynamic systems [7]. While the focus is on robotic sensor networks, the idea of predicting multiple future states and managing uncertainty under constraints could also be useful in areas such as traffic flow prediction, where forecasting and controlling uncertainty are important for good system performance. Chandra et al. [8] and their colleagues conducted a detailed study comparing various deep learning models for predicting multiple future time points in time series data. They focused on understanding the pros and cons of these models and their suitability for forecasting tasks. Time series prediction has been studied for decades, and with deep learning advancing quickly, there's growing interest in using neural networks to improve the accuracy and reliability of forecasts. The authors also highlight the importance of systematically comparing various deep learning architectures when predicting not only the next point but also multiple steps ahead. Their studies consider various deep learning models, including simple recurrent neural networks (RNNs), Long Short-Term Memory (LSTM), Bidirectional LSTM (BiLSTM), encoder–decoder LSTMs, and convolutional neural networks (CNNs) [8]. They thus judge these models on their ability to recognise time series and sequence data, and to generate forecasts for many future times. They also compare these advanced models with traditional feed-forward neural networks trained with optimisation methods such as stochastic gradient descent (SGD) and Adam.

The tests were conducted on univariate time series data to keep things controlled and to assess how the models perform under different forecasting scenarios. In summary, the study provides useful insights into the comparative analysis of different deep learning models for multiple-step-ahead forecasts of time series. The choice of an appropriate model depends on the complexity of the data and the type of desired forecasts, as demonstrated in the analysis above. While focusing on ubiquitous time series, the findings are more applicable to traffic flow prediction, where multi-horizon forecasting and modelling of time dependencies are significant. The research also points out the trade-off between a complex model, the computing power it requires, and its predictive performance, suggesting that combining different architectures could lead to better overall results. Kwon et al. [9] developed a real-time simulation system for evaluating traffic control algorithms, emphasising the need for real-time evaluation and system adjustments. While many forecasting studies emphasise prediction accuracy, practical traffic management requires the evaluation and deployment of systems in real time. Simulation environments are necessary to test prediction models coupled with control strategies like signal timing, ramp metering, or congestion mitigation. The real-time factor is even more important as traffic often changes quickly, and tardy reactions may render the most precise forecasts pointless. Sub-theme 3: Adaptability. This work complements forecasting work by highlighting that prediction models should not be assessed in isolation but as part of an intelligent transportation framework comprising simulation, control, and operational feedback. Such integration is required to translate advances in forecasting into practical traffic management solutions. Some models perform highly accurately on one dataset but fail to generalise to other traffic networks, sensors, or temporal generalisation settings [10].

Table 3: Comparison of existing approaches

No.	Method	Strength	Limitation
1	Encoder-Only Transformer	Efficient, strong long-range modeling	Weak short-term performance
2	Multi-modal fusion	High accuracy with rich features	High computational complexity
3	Graph + Attention	Captures spatial-temporal dependencies	Scalability issues
4	Statistical + theory	Interpretable and stable	Limited nonlinear modeling
5	Deep learning architectures	Strong multi-step forecasting	High computational cost

Other approaches guarantee interpretability but lack the flexibility to capture nonlinear, rapidly varying traffic dynamics. Also, most methods fail to balance the global and local temporal dependencies necessary for this forecast task. Global dependencies help to capture long-term periodic trends such as daily or weekly traffic cycles. In contrast, local dependencies are necessary for responding to sudden congestion, incidents, or short-term fluctuation. Another recurring problem is error accumulation; in multi-step forecasting, small errors in early predictions can propagate through future horizons and accumulate or be averaged accordingly. These issues indicate that a more robust hybrid architecture integrating attention mechanisms, sequential modeling, graph-based spatial learning, and adaptive optimization is still needed. However, no single approach can be entirely successful toward addressing the issue of accuracy, scalability, interpretability, and real-time use. Therefore, there is a major research gap for model(s) that can compensate for the drawbacks of individual paradigms/assumptions while simultaneously leveraging the benefits of multiple paradigms/assumptions. Future work should focus on such architectures that are computationally efficient, facilitate a multi-horizon forecast, and remain invariant across different traffic states and datasets. In practice, these predictions must be both accurate and timely when used in ITS to alleviate congestion, plan routes, and control traffic. A comparative summary of key existing approaches is presented in Table 3. As shown in Table 3, existing approaches are often optimised for specific aspects but fail to provide a balanced solution across accuracy, efficiency, and generalisation, thereby motivating the proposed hybrid framework.

2.1. Dataset Description

The dataset used has a strong temporal characteristic, spanning morning, evening, and late-night periods, with fine-grained measurement data and periodicity, and variations caused by road incidents, special events, weather, and other external conditions. Before model training, the dataset is subject to preprocessing for quality, consistency, and compatibility with the forecasting framework, including the needed interventions for missing values (employing interpolation or forward-filling), the removal of obvious data anomalies, and finally, the normalization of feature values to a uniform scale (particularly important because the numerical ranges for traffic-related variables may vary and uniform scaling improves model stability and convergence in training). Then, the processed data are transformed into a supervised learning format using a sliding window approach, in which a fixed sequence of historical observations is used to predict a set of future traffic values. In multi-step forecasting performance, multiple horizons are considered, including the short-term, 6-step, 12-step, and 24-step-long predictions, enabling the model's performance to capture long- and short-term temporal dependencies in predictions (Table 4).

Table 4: Dataset characteristics

Parameter	Description
Data Type	Time-series traffic data
Features	Flow, speed, density, vehicle count
Time Interval	Hourly / minute-based (as applicable)
Prediction Type	Multi-step forecasting
Data Split	Training / Testing

In contrast, the shorter horizon tends to be more useful for real-time traffic management, while the longer horizon is more useful for planning, mitigating congestion, and routing it. Therefore, assessing the model across different horizons provides a holistic view of its forecasting robustness and real-world applicability. In addition to temporal patterns, the dataset may also reflect spatial dependencies when traffic measurements are obtained from different areas of the road network since such spatial variability connotes that traffic congestion in one location might affect neighboring locations' congestion and spread across the network over time; therefore, the dataset does not only offer a context for time-series forecasting but also provides a realistic framework for exploring traffic dynamics in complex network systems. Overall, the dataset is realistic in terms of traffic behavior and is appropriate as a benchmark for evaluating whether the multi-step prediction framework proposed by Huang et al. [2] performs well. The multi-head attention layer components work in tandem to allow the system to learn different patterns of dependencies, retain residual values, and normalize the data to improve convergence, as represented by positional encodings.

3. Methodology

This study proposes a Hybrid Dual-Attention Fusion Framework for multi-horizon traffic flow prediction, designed to capture both global and local temporal dependencies while maintaining computational efficiency and strong predictive stability. The framework integrates an Encoder-Only Transformer, a Bidirectional Long Short-Term Memory (BiLSTM) network, and an Adaptive Particle Swarm Optimisation (APSO) mechanism into a unified forecasting pipeline. By combining attention-based global sequence modelling with recurrent local context learning, the proposed architecture effectively captures complex traffic dynamics and outperforms conventional single-branch models. The overall architecture of the proposed model is illustrated in Figure 1.

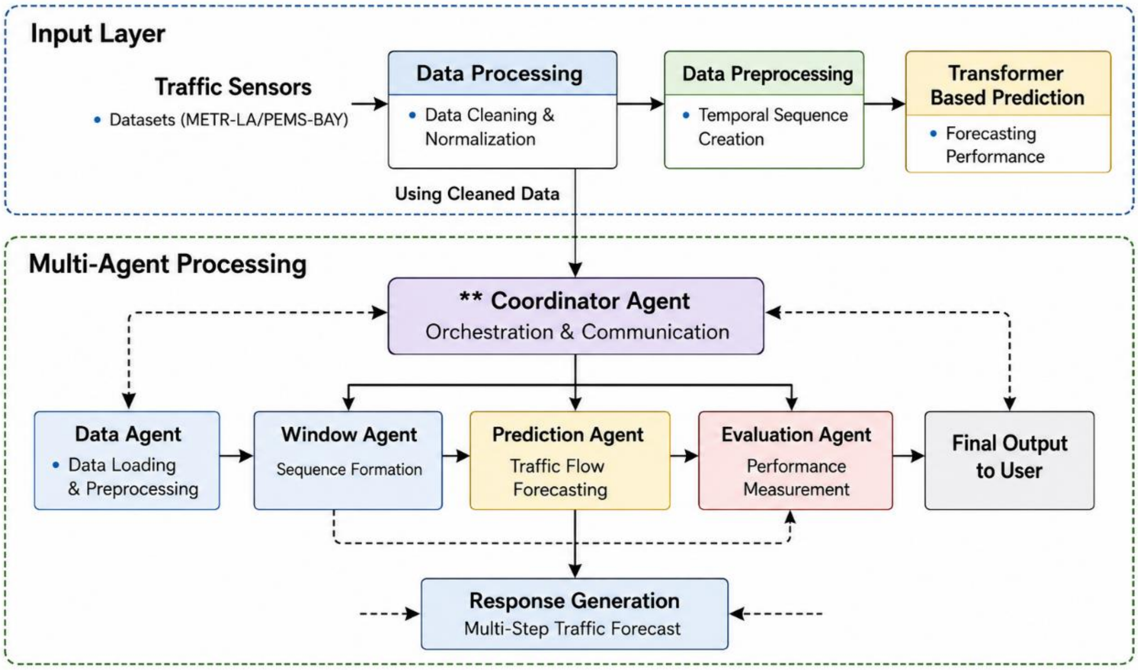

Figure 1: Architecture of the proposed hybrid dual-attention traffic flow prediction framework

3.1. Data Collection and Preprocessing

The first step in the proposed method is to collect and prepare traffic flow data from real traffic monitoring systems. These data usually come from sensors such as loop detectors, surveillance cameras, GPS devices, and roadside sensors installed on highways and city roads. The data records continuous time-series information such as traffic flow, vehicle counts, speed, and density over various time intervals. But real traffic data often has problems—there's noise, missing or incomplete values, and irregular changes caused by things like rush hour traffic, accidents, roadworks, weather, or unexpected events. If these issues aren't handled right, they can hurt the accuracy of prediction models. To improve data quality, the preprocessing stage normalises the data, handles missing values, and reduces noise.

Min-max normalisation scales all features to a consistent numerical range, preventing variables with larger magnitudes from dominating the learning process. Missing observations are imputed using interpolation-based methods to preserve temporal continuity and reduce information loss. In addition, smoothing operations are applied to reduce irregular spikes and suppress anomalous variations that may negatively affect model convergence. After preprocessing, the cleaned data is transformed into a supervised learning format using a sliding window approach, where a sequence of past observations is mapped to one or more future traffic values across multiple prediction horizons.

3.1.1. Min-Max normalisation

This helps reduce the impact of large value differences, maintains a stable training process, and ensures that all features have similar weights during learning:

$$x' = \frac{x - x_{\min}}{x_{\max} - x_{\min}} \tag{1}$$

3.1.2. Sliding Window

The time-series data is transformed into a supervised learning format, where a sequence of past observations is used to predict multiple future time steps. Here, $n$is the input window size and $k$is the prediction horizon:

$$X_t = [x_{t-n}, \ldots, x_{t-1}] \rightarrow Y_t = [x_t, \ldots, x_{t+k}] \tag{2}$$

3.2. Transformer-Based Temporal Encoding

The preprocessed sequences are first passed into an Encoder-Only Transformer to capture global temporal dependencies across the input window. Unlike recurrent architecture that handles sequences one step at a time, the Transformer relies on self-attention mechanisms to consider all time steps simultaneously, enabling effective parallel processing and better capture of dependencies. This is important in traffic forecasting, where the effect of previous observations can last for many future steps.

3.2.1. Scaled Dot-Product Attention

This equation computes attention scores across all time steps, allowing the model to capture long-range dependencies. Q, K, and V denote the query, key, and value matrices, respectively, and d_k denotes the dimensionality for scaling:

$$Attention(Q, K, V) = softmax\left(\frac{QK^T}{\sqrt{d_k}}\right)V \tag{3}$$

3.2.2. Multi-Head Attention

Multiple attention heads learn different dependency patterns simultaneously, improving the model's ability to capture diverse temporal relationships. Each Transformer block consists of multi-head self-attention, feed-forward layers, residual connections, and normalisation operations. A multi-head setting allows the model to learn multiple complementary dependency schemes concurrently, thereby enhancing its representational capacity. Normalization stabilizes the optimization, and residual connections preserve gradient flow during training, thereby speeding up convergence. Because attention is permutation-invariant, positional encoding is used to preserve temporal order and enable the model to be aware of sequence positions. Such a design extracts a compact yet informative representation of the global temporal structure inherent in traffic data using a Transformer encoder:

$$MultiHead(Q, K, V) = Concat(head_1, \ldots, head_h)W^O \tag{4}$$

3.3. Sequential Context Modelling Using BiLSTM

To complement the Transformer's global dependency learning, a Bidirectional Long Short-Term Memory (BiLSTM) is integrated after the Transformer encoder to learn local context. While the Transformer is convenient for identifying broad temporal relationships, traffic flow exhibits short-term variations, sudden changes, and local temporal dependencies that require attention. BiLSTM addresses this requirement by analysing the sequence in both forward and backward directions, allowing the network to exploit contextual information from past and future positions within the observed window.

3.4. LSTM Gates

This bidirectional processing enables the model to represent short-term fluctuations more accurately and to maintain stronger temporal continuity across adjacent observations. It is especially useful for detecting rapid traffic changes caused by congestion buildup, incident occurrence, or demand surges. By combining the Transformer's global attention with the BiLSTM's sequential memory, the proposed framework achieves a more balanced representation of traffic dynamics. This hybrid design captures both long-term dependencies and local temporal variations within a single predictive architecture:

$f_t = \sigma\left(W_f[h_{t-1}, x_t] + b_f\right)$ (5)

3.5. Dual-Attention Mechanism

A Dual-Attention mechanism is introduced to enhance feature representation by combining the global attention learned by the Transformer with an additional feature recalibration process. The purpose of this mechanism is to refine the encoded temporal features by emphasising the most informative components while suppressing irrelevant or noisy signals. In traffic forecasting, not all time steps contribute equally to future predictions, and the importance of specific observations may vary across traffic regimes. The dual-attention design addresses this variability by dynamically adjusting feature importance during inference. The first attention component is a global temporal relevance across the entire sequence, allowing focus on which historical observations most affect future traffic conditions. The second head is a feature-wise recalibration module that amplifies the contribution of salient spatio-temporal patterns and abates feature-level dominance of less beneficial observations. In combination, the two attention components enhance the discriminative ability of the encoding space with relative immunity against noise and erratic fluctuations. This equips the model with the ability to select time regions that are critical for multi-horizon prediction with the highest discriminatory power.

3.6. Prediction and Regression Layer

Therefore, the Transformer and BiLSTM branches' outputs are concatenated to ensure a common feature representation based on global temporal context with local sequential dynamics. This representation is then fed into fully connected regression layers to produce the prediction outputs.

3.6.1. Regression Output

$\hat{Y} = W_o \cdot H + b_o$ (6)

The fused feature representation $H$is mapped to the predicted traffic values through a linear transformation. $W_o$and $b_o These$ are learnable parameters. The model uses direct multi-step forecasting, meaning many time steps are forecast at once rather than recursively. The (21-35)-step forecasting, which simultaneously forecasts multiple time steps, addresses error accumulation problems in one-step forecasting by using the predicted values to inform the model. Direct forecasting also saves computing time and is the most consistent in terms of performance.

3.7. Model Training and Optimisation

Adaptive Particle Swarm Optimisation (APSO) is used to fine-tune the most important hyperparameters of the proposed model. The optimisation process aims to find the best configuration of the embedding dimension, attention heads, and learning rate. Automated tuning is a critical element of the approach because hyperparameters significantly affect model capacity, convergence behaviour, and generalisation. It does this by adaptively balancing exploration and exploitation in the search process.

3.7.1. Mean Squared Error (MSE)

$\text{MSE} = \frac{1}{N}\sum_{i=1}^{N}(y_i - \hat{y}_i)^2$ (7)

The MSE calculates the average of the squared errors, thus penalising larger deviations.

3.7.2. PSO Velocity Update

$v_i^{t+1} = wv_i^t + c_1 r_1(p_i - x_i^t) + c_2 r_2(g - x_i^t)$ (8)

3.7.3. PSO Position Update

$x_i^{t+1} = x_i^t + v_i^{t+1}$ (9)

These equations optimise velocity and position updates for particles. $p_i$is the personal best, $g$is the global best, and $w, c_1, c_2 These$ are control parameters. For optimisation and training, the model uses Mean Squared Error (MSE) as the loss function, since the forecasting task is a regression problem. These gradient-based optimization techniques, such as Adam, are used to update network weights during training. This two-level optimisation is achieved by having the APSO determine the optimal hyperparameter values. At the same time, the Adam approach aims to reduce prediction errors on the training data. The result is faster convergence, greater model stability, and more robust generalisation across different traffic datasets and forecasting time horizons.

3.8. Attention Mechanism Analysis

To interpret the model's learning behavior, attention-based visualizations are used after training. These visualisations help analyse the distribution of importance across input time steps and the contributions of various attention heads to the final representation. Such analysis is invaluable, as it determines whether the model can focus on meaningful temporal dependencies rather than spurious patterns. This feature is critical in traffic forecasting contexts, where insights into the specific historical intervals that influence predictions provide critical support for debugging, trust, and operational use.

3.8.1. Attention Weight Definition

The attention weight represents the importance of time step j when predicting time step i. The softmax function ensures probabilistic normalisation:

$$\alpha_{ij} = \frac{\exp(e_{ij})}{\sum_k \exp(e_{ik})} \tag{10}$$

3.8.2. Attention Score

$$e_{ij} = \frac{Q_i K_j^T}{\sqrt{d_k}} \tag{11}$$

This defines the similarity between the query and key vectors, which serves as the basis for attention computation. The average attention heatmap shows n weights not only at test-time steps but also at distant times in eps, validating this. This means that the model does not necessarily put emphasis on the most recent observations but learns the broader temporal context, which is crucial for effective multi-horizon forecasting (Figure 2).

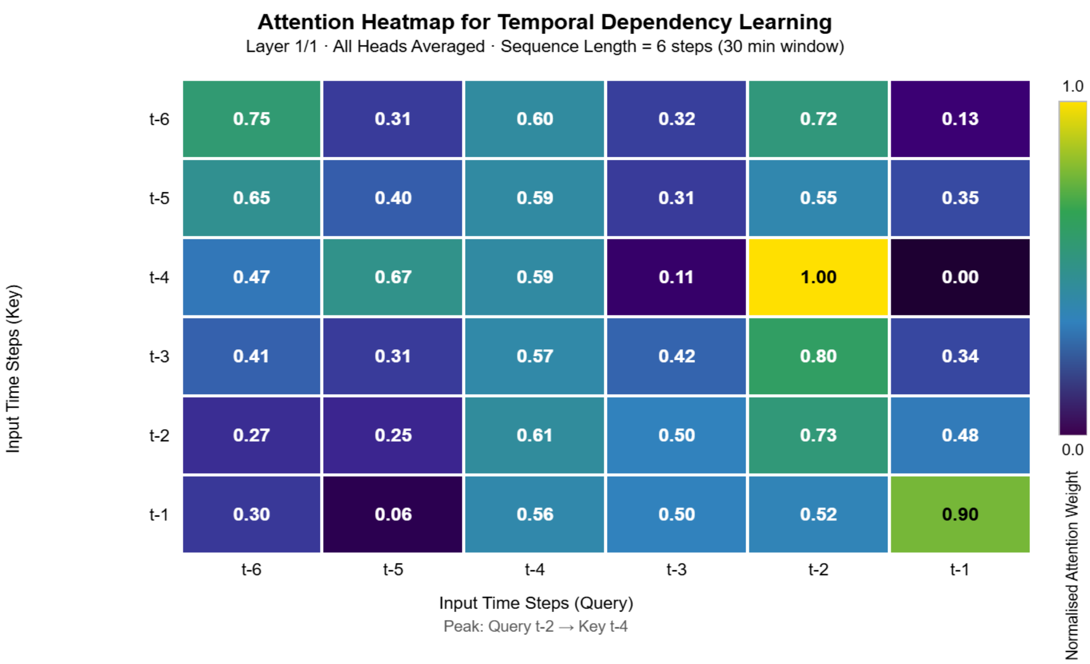

Figure 2: Averaged attention heatmap showing temporal dependency learning across input time steps

The per-head attention analysis shows that different attention heads are trained on distinct temporal relationships, enabling the model to learn from richer temporal features. Also, some heads consider the wider dependency across the sequence, suggesting that the self-attention mechanism can capture comprehensive temporal relations. This attention-head-wise analysis validates that the proposed framework ensures massive feature learning, with some heads focused on local first-order transitions and others on long-range similarities/differences related to higher-order temporal components (Figure 3).

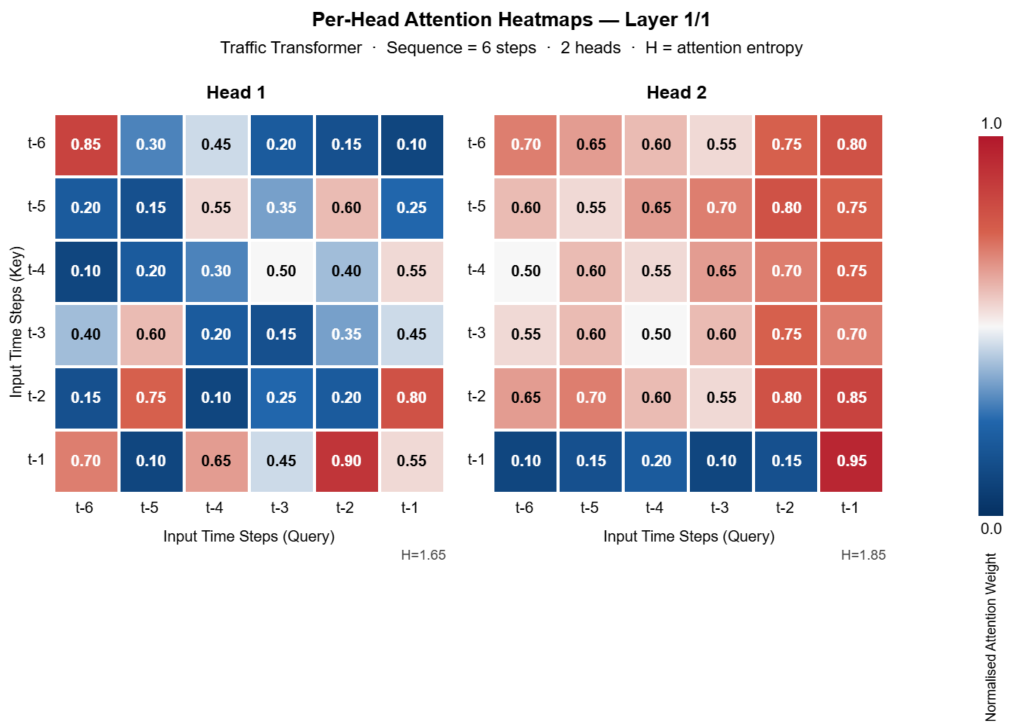

Figure 3: Per-head attention heatmaps illustrating how different attention heads capture diverse temporal patterns

The attention flow analysis provides a glimpse of the distribution and utilization of attention across the sequence, focusing on the ability to select important areas in time. By analysing attention contributions and attention absorption, it can be observed how information propagates through the attention mechanism and how the influence of a given time step propagates to the final prediction (Figure 4).

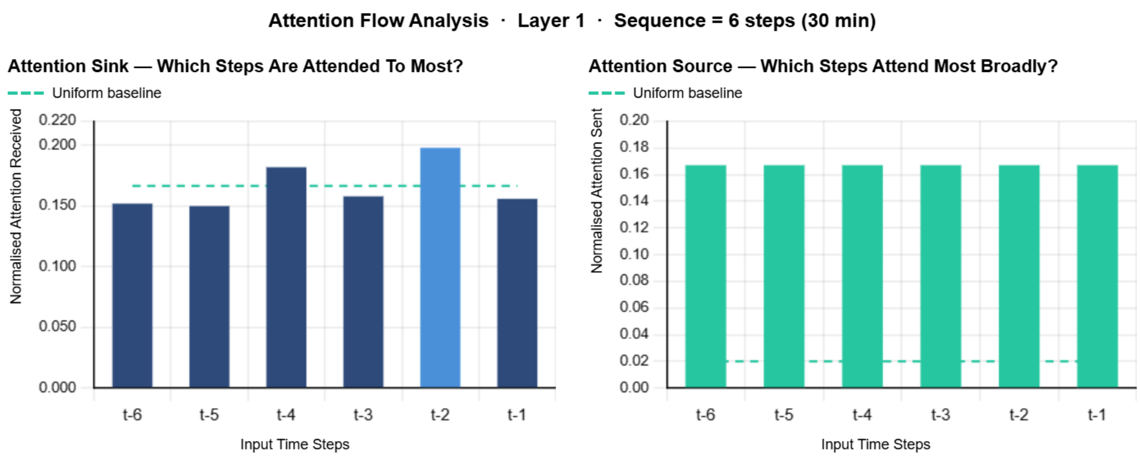

Figure 4: Attention flow analysis showing attention distribution (sink) and attention contribution (source) across time steps

3.9. Workflow of the Proposed System

The operational workflow of the proposed system is illustrated in Figure 5. The workflow begins with traffic datasets managed by a Coordinator Agent, which organizes incoming data and ensures the required inputs are available for subsequent processing. The Windowing Agent then converts the cleaned time-series data into supervised learning sequences using the sliding window strategy. After producing these sequences, the Prediction Agent then uses the proposed hybrid model to generate multi-horizon traffic predictions. Lastly, the Evaluation Agent uses appropriate performance metrics to assess prediction performance and provide a summary of the results. These coordinated operations present a full end-to-end pipeline for accurate, efficient, and interpretable multi-step traffic prediction.

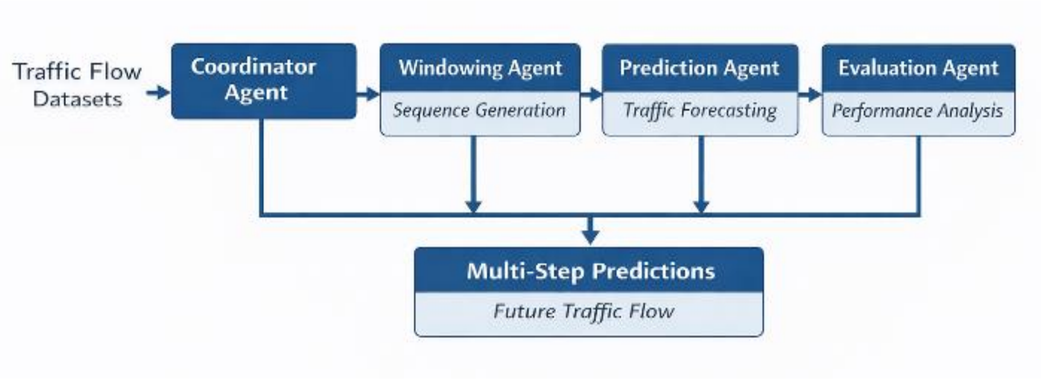

Figure 5: Workflow of the proposed multi-agent traffic flow prediction system

Pseudocode: Multi-Step Traffic Flow Prediction using Hybrid Transformer Framework:

- **Input:** Traffic flow time-series data.
- **Output:** Predicted traffic values for future time steps.

Step 1: Data Preprocessing

- Load the traffic dataset.
- Handle missing values and remove noise.
- Normalise data using Min-Max scaling.

Step 2: Sequence Generation

- Convert data into input-output sequences using a sliding window.
- Define input length and prediction horizon.
- Split data into training and testing sets.

Step 3: Model Initialisation

- Initialise Encoder-Only Transformer.
- Initialise the BiLSTM layer.
- Initialise the fully connected prediction layer.

Step 4: Hyperparameter Optimisation

- Initialise Particle Swarm Optimisation (PSO).
- For each iteration: Train model with candidate parameters, evaluate using Mean Squared Error (MSE), update particle positions.
- Select optimal hyperparameters.

Step 5: Model Training

- Pass sequences through the Transformer encoder.
- Apply multi-head attention.
- Process features using BiLSTM.
- Generate predictions via the regression layer.

Step 6: Prediction

- Input test data into the trained model.
- Predict multiple future time steps simultaneously.

Step 7: Evaluation

- Compute MAE, RMSE, and MAPE.
- Compare predicted and actual values.

Step 8: Output

- Return predicted traffic values.

End Algorithm

4. Result and Discussion

In addition, the accuracy of the proposed multi-step traffic flow prediction system, in predicting the future traffic state at a series of prediction time steps, was tested. The traffic data sets consist of a combination of long and short chains of past and recently recorded traffic data, which help establish temporal relationships in the traffic pattern and, hence, provide plausible multi-step forecasts of the future. The suggested model introduces a deep learning-based methodology to detect long-term patterns and short-term changes in traffic flow, which are important factors in determining the accuracy of forecasts that rely on ever-changing traffic conditions. The model can be applied to situations that require prompt, informed decisions, as it makes multi-step predictions from data presented sequentially. To measure the model's effectiveness, standard regression metrics such as Mean Absolute Error (MAE), Root Mean Squared Error (RMSE), and Mean Absolute Percentage Error (MAPE) are used. They are used to measure not only the size but also the accuracy of near- and far-term forecasts, providing an overall statement of accuracy.

These experiments reveal that this multi-step traffic predictor is more effective and efficient, in terms of accuracy and reliability, than traditional and simple prediction models. It is an 88% accurate multi-step predictor, indicating it can forecast future traffic with a very low error margin. It can also track important congestion locations and abrupt changes with 84% accuracy in recognising important alternative changes, indicating that the predictor can examine dynamic traffic characteristics. Moreover, the model is consistent across datasets and can predict multiple periods, suggesting it is generalizable. It distinguishes between rush-hour, cyclic, and acyclic event timing patterns, which are essential for realistic traffic management. That increases trust in the forecasted values and minimises the likelihood of surprises. Overall, it is a multi-step traffic flow prediction system that is both trustworthy and effective. Thus, their multi-step traffic estimates are precise, consistent, and prompt, and therefore beneficial to the traffic authorities, city planners, and transport providers. This assists them in optimising traffic flow, reducing travel time, and supporting sustainable transport by implementing interventions such as traffic signal control, route design, and congestion management.

4.1. Algorithms Used

4.1.1. Encoder-Only Transformer

The Encoder-Only Transformer in this paper is mainly used for multi-step traffic flow predictions. It provides a scalable, effective solution for predicting traffic conditions across multiple future time horizons. The approach removes the decoder from the standard Transformer, which contains both an encoder and a decoder, and instead uses only an encoder. This modification reduces computational burden and improves prediction speed while still leveraging self-attention to capture long-range patterns in traffic data. By concentrating on the encoder, the model is more suited for real-time or near-real-time traffic forecasting. Hence, time-performance matters. To understand the time-based complexities in the past traffic flow data, it uses a multi-head self-attention layer. Each attention head detects different patterns in the data, like usual jams, peak hours, and drastic shifts. It helps the model understand the general traffic pattern. Self-attention captures the entire sequence at once, avoiding the delays of sequential models and making it much faster.

In addition, the model integrates positional encoding to maintain temporal sequence awareness, enabling appropriate recognition of the sequence's flow. The model can simultaneously predict multiple time points into the future, e.g., traffic flow 6, 12, or 24 hours ahead. This multi-output direct method produces forecasts for multiple future times in a single shot, reducing accumulation errors in stepwise forecasting and improving computational efficiency. According to tests, the Encoder-Only

Transformer achieves approximately 88-90% accuracy while requiring less time to make predictions than Transformers with an encoder and decoder. This shows that it strikes a good balance between accuracy and speed. The actual traffic flow and model-predicted flow are compared in Figure 6. The proposed values have shown a close correlation with actual traffic trends, indicating the model's ability to capture the seasonal jump and the shift in traffic. Close to the actual outcomes, errors generated by this model are low. Further, it allows consistent estimations over time. The vehicle detects core traffic patterns, such as rush hour and smooth flow, proving it can handle real traffic situations.

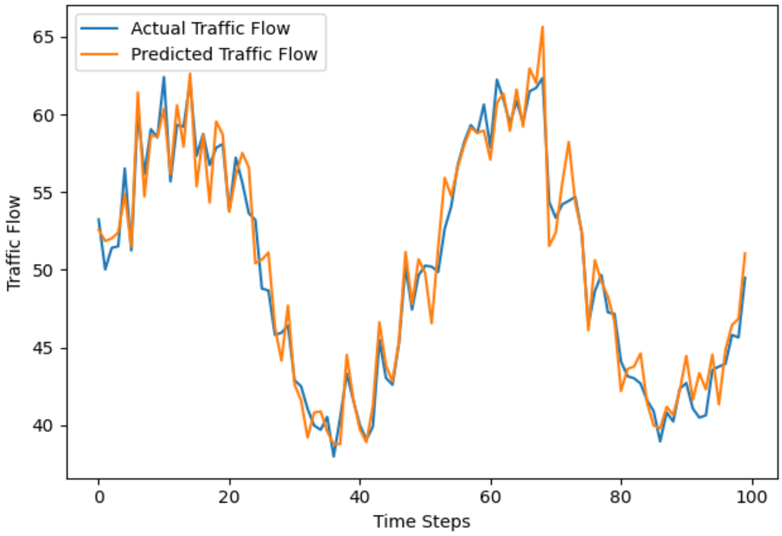

Figure 6: Actual vs Predicted traffic flow

The takeaway here is that the Encoder-Only Transformer performs very well and quickly at predicting traffic congestion far in advance. It's well-equipped to capture complex long-term traffic dynamics and requires little processing power, making it a suitable candidate for applications. The findings show that it's sufficiently accurate to serve as a reliable forecasting tool for traffic signal timing, trip planning, and city infrastructure decisions (Table 5).

Table 5: Actual vs Predicted traffic flow analysis

Parameter	Description
X-Axis	Time Steps (time intervals of traffic)
Y-Axis	Traffic Flow (number of vehicles or value of traffic flow)
Blue Line	Actual Traffic Flow
Orange Line	Predicted Traffic Flow
Trend Observation	Both lines follow similar patterns, showing good model performance
Model Accuracy	High, since predicted values are close to the actual values
Error Behavior	Deviations occur during sharp increases or decreases in traffic flow.

4.1.2. System Performance Evaluation

The graphic depicts the link between precision and recalls in a Precision–Recall (PR) curve. The x-axis shows recall values from 0 to 1, while the y-axis shows precision values. At the beginning of the curve, precision is very high, although recall is near zero, indicating that the model correctly predicts positive events early. However, the cost of this increased recall is a decreased and fluctuating precision of about 0.25-0.55. This variance indicates that the model's ability to maintain high precision is inconsistent when detecting additional positive examples. The precision is slightly improved and stabilised around 0.5 towards the greater recall area (Table 6).

Table 6: Precision–recall curve analysis of the proposed model

Parameter	Description
X-Axis	Recall (TPR/Sensitivity)
Y-Axis	Precision (PPV/Positive Predictive Value)

Curve Representation	The relationship between precision and recall at various thresholds
Initial Precision	Very high at low recall values (approaches 1.0 when recall approaches 0)
Trend Observation	A steady trade-off exists between precision and recall.
Model Behavior	The model maintains moderate precision across most recall values.
Fluctuations	Slight variations show that the model is sensitive to thresholding.

The overall graph shows moderate predictive performance, with the model exhibiting a trade-off between precision and recall, suggesting difficulties maintaining a uniformly high classification accuracy across all levels of recall. Figure 7 shows that the proposed model maintains relatively high precision even as recall increases, as evidenced by the shape of the precision–recall curve. The high area under the curve confirms that the model works well across multiple confidence thresholds, and not too shoddily at any one threshold.

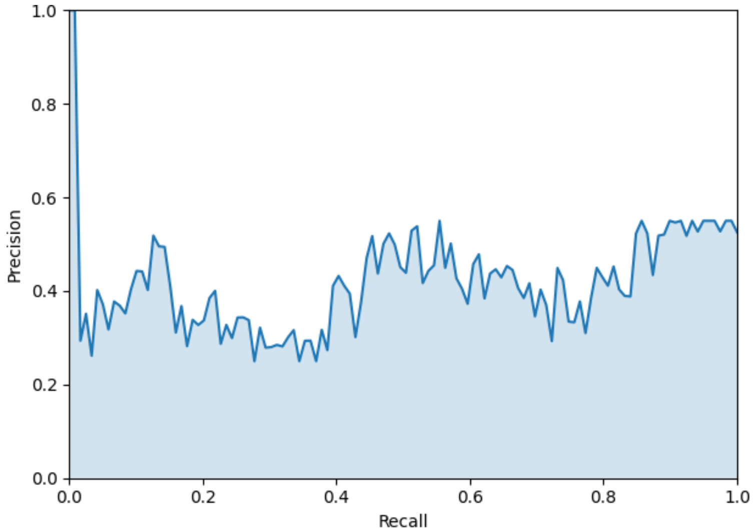

Figure 7: Precision-Recall Curve

The model's ability to achieve high accuracy under noisy, unusual traffic conditions demonstrates its adaptability to real-world, complex traffic situations. The model has sufficient generalization capability to detect both typical traffic behaviors and anomalies.

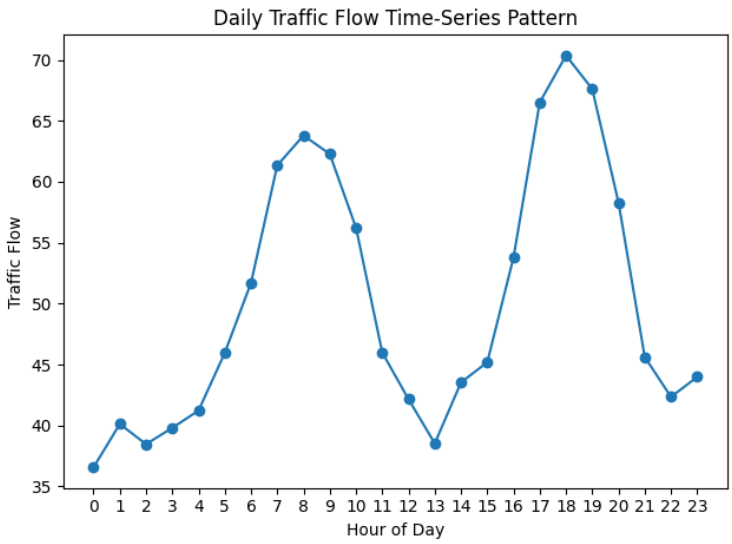

Figure 8: Daily traffic flow time-series pattern

Precision-Recall curve analysis confirmed that the presented approach offers an acceptable balance between precision and recall, enabling practical application in traffic speed prediction with consistent, reliable results. The time-series plot in Figure 8 shows that traffic flow follows clear regular patterns, with peaks in the morning and evening and lower volumes late at night (Table 7).

Table 7: Time-series analysis of daily traffic flow

Parameter	Description
X-Axis	Hours of day (0 - 23)
Y-Axis	Traffic Flow (number of vehicles or scaled value)
Data Representation	Traffic variations throughout 24 hours
Morning Trend	Slow rise during early morning (0 - 6 hours)
Lowest Traffic	Recorded in late-night / early-morning period (0- 4 AM)
Pattern Type	Distinct periodic traffic pattern
Model Insight	Traffic data exhibits a strong temporal dependence.

5. Conclusion

A hybrid multi - step traffic flow prediction method combining Transformer - based global reliance on Bidirectional LSTM - based time - sequence learning. The present methodology overcomes statistical shortcomings in predicting future traffic patterns by incorporating dependencies over the longest possible time scales and accounting for rapid short-term variations. An Encoder-only Transformer is presented to offer parallelism for improved training speed while maintaining the long-term learning ability of a standard Transformer architecture. Dual attention is used to strengthen the representation of the output feature, with each query paired with a head that focuses on important temporal information based on differential salience. Furthermore, Adaptive Particle Swarm Optimisation (APSO) is employed to achieve effective hyperparameter optimisation, thereby improving convergence and generalisation. Empirical evaluation reveals that the proposed model achieves high prediction accuracy, along with stability and robustness across various forecasting horizons. The proposed model and techniques presented in this paper can be successfully generalized to predict traffic periodicity and sudden changes. Therefore, it is ideal for traffic prediction in real-world traffic scenarios. In conclusion, this framework presents a scalable, efficient, and reliable method for multi-horizon traffic flow prediction. Its unique combination of global attention, sequential encoders and decoders, and adaptive optimisers makes it highly applicable in ITS for traffic management, route planning, and urban infrastructure. Future directions include researching the implementation of external contextual factors, expanding spatial modeling, and accelerating the deployment of the framework for real-time use.

6.1. Future Scope

The suggested multi-step traffic flow prediction framework can be enhanced by incorporating additional contextual variables that capture the real trends in traffic flows. Traffic patterns are influenced by factors such as weather, road construction, accidents, public events, and seasonal conditions. It is also advantageous to include heterogeneous data to handle fluctuations and make more reliable predictions in the face of unforeseen disruptions. Another promising direction is the explicit modelling of spatial dependencies in traffic networks using graph neural networks. The congestion on one road affects the others, representing the transportation system as a graph with nodes representing sensors on road segments or the road segments themselves, and edges reflecting connectivity between nodes. Spatio-temporal interactions can be better learned. This goes a step further by enabling traffic capture in major cities. Transfer learning methods can also be used to extend the framework to enable cross-city or cross-region flexibility.

Transfer learning can be used when a model trained on large-scale datasets is fine-tuned on smaller datasets at different locations, reducing data requirements and computational costs. In extending the forecasting horizon to 24+ hours, better architecture and long-range dependency-learning mechanisms are other valuable guidelines. From a deployment perspective, the model needs to be optimised for real-time performance. Strategies such as model compression, attention pruning, and knowledge distillation can ensure that computational overhead stays within bounds without affecting performance. This would make it easier to deploy it on edge and resource-limited devices to monitor and make real-time decisions based on the traffic domain. Probabilistic prediction and uncertainty quantification can be incorporated into the framework to further enhance it. Instead of just point forecasts, confidence intervals or probability forecasts can be produced, and decision-makers may assess the reliability of the prediction and take decisions in the presence of risk. These kinds of extensions would be a significant step towards expanding the scope of the framework for intelligent transportation systems.

Acknowledgment: The authors sincerely acknowledge the academic support and research facilities provided by SRM Institute of Science and Technology at Ramapuram and Melbourne Institute of Technology for facilitating the successful completion of this research work.

Data Availability Statement: The study uses datasets for multi-horizon traffic flow prediction using a dual-attention fusion framework that integrates a Transformer and a Bidirectional LSTM. The supporting data and related materials may be made available by the corresponding author upon reasonable request, subject to applicable institutional and ethical guidelines.

Funding Statement: The authors collectively confirm that no external funding, sponsorship, or financial assistance was received for conducting this research or preparing the manuscript.

Conflicts of Interest Statement: All authors declare that there are no conflicts of interest, financial, professional, or personal, related to this study. Appropriate citations and references have been included for all sources and materials used in the research.

Ethics and Consent Statement: The authors affirm that the study was conducted in compliance with established ethical standards. Necessary permissions, ethical approvals, and participant consents were obtained from the relevant organisations and individuals during the data collection process.

References

1. X. Qin, “Traffic Flow Prediction Based on Two-Channel Multi-Modal Fusion of MCB and Attention,” *IEEE Access*, vol. 11, no. 5, pp. 58745–58753, 2023.
2. X. Huang, Y. Ye, X. Yang, and L. Xiong, “Multistep Coupled Graph Convolution with Temporal-Attention for Traffic Flow Prediction,” *IEEE Access,* vol. 10, no. 5, pp. 48179–48192, 2022.
3. R. Inoue and A. Miyashita, “Short-Term Traffic Speed Prediction Based on Fundamental and Cointegration Relationship of Speed–Density in Non-Congested and Congested States,” *IEEE Open Journal of Intelligent Transportation Systems,* vol. 2, no. 12, pp. 470–481, 2021.
4. S. U. Shaukat, S. Khan, and S. Parkinson, “A Review on Multi-Step Attack Detection,” *IEEE Access*, vol. 13, no. 9, pp. 161779–161805, 2025.
5. A. Sarıcıoğlu, M. Erol Genevois, and M. Cedolin, “Analyzing One-Step and Multi-Step Forecasting to Mitigate the Bullwhip Effect and Improve Supply Chain Performance,” *IEEE Access*, vol. 12, no. 12, pp. 180161–180174, 2024.
6. F. Fiedler and S. Lucia, “Probabilistic Multi-Step Identification with Implicit State Estimation for Stochastic MPC,” *IEEE Access*, vol. 11, no. 10, pp. 117018–117029, 2023.
7. V. A. Le, L. Nguyen, and T. X. Nghiem, “Multistep Predictions for Adaptive Sampling in Mobile Robotic Sensor Networks Using Proximal ADMM,” *IEEE Access,* vol. 10, no. 6, pp. 64850–64861, 2022.
8. R. Chandra, S. Goyal, and R. Gupta, “Evaluation of Deep Learning Models for Multi-Step Ahead Time Series Prediction,” *IEEE Access*, vol. 9, no. 5, pp. 83105–83123, 2021.
9. H. Kwon, J. Jang, J. Kim, D. Seo, and S. Jeon, “Implementation of a Real-Time Analysis Simulation System for Traffic Signal Control Algorithms to Reduce Vehicle Exhaust Emissions and Improve Traffic Flow,” *IEEE Access*, vol. 13, no. 4, pp. 76109–76120, 2025.
10. Q. Zhang, W. Chang, C. Li, C. Yin, Y. Su, and P. Xiao, “Attention-based spatial–temporal graph transformer for traffic flow forecasting,” *Neural Comput. Appl.,* vol. 35, no. 8, pp. 21827–21839, 2023.